

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی شاهرود

دانشکده شیمی  
گروه شیمی فیزیک

عنوان:

پیش‌بینی ثابت‌های هنری بعضی از ترکیبات آلی با استفاده از روش‌های  
خطی و غیر خطی QSPR

دانشجو:

محمد علی فردوسی

اساتید راهنما:

دکتر حسین نیکوفرد

دکتر ناصر گودرزی

استاد مشاور:

دکتر زهرا کلانتر

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهار ۱۳۸۹



## دانشگاه صنعتی شاهرود

دانشکده شیمی

پایان نامه کارشناسی ارشد

**محمدعلی فردوسی**

رشته شیمی فیزیک

تحت عنوان

پیش بینی نایبهای هنری بعضی از ترکیبات آلی با استفاده از روشهای خطی و غیر خطی ارتباط کمی ساختار - خاصیت (QSPR)

در تاریخ ۱۳۸۹/۳/۲۲ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد رشته شیمی فیزیک مورد ارزیابی و با درجه عالی مورد پذیرش قرار گرفت.

امضاء	مرتبه علمی	نام و نام خانوادگی	اعضای هیأت داوران
	استادیار	حسین نیکوفرد	استاد راهنمای اول
	استادیار	ناصر گودرزی	استاد راهنمای دوم
	استادیار	زهرا کلانتر کهدمی	استاد مشاور
	استادیار	سیدعلینقی طاهری	نماینده شورای تحصیلات تکمیلی
	استادیار	زینب موسوی تکیه	استاد ممتحن
	دانشیار	قدمعلی باقریان دهقی	استاد ممتحن

تقدیم

به پیشگاه پدر و مادرم

که نگرانی‌ها و دلسوزی‌هایشان پایه گزار رشد فکری من در تمام مراحل

تحصیلم گردید و تنها سرمایه فقیقی من دعای بی منتشان است.

و به تمام دوستانم که اندیشیدن، اراده و تلاش را به من هدیه دادند.

## ابراز تواضع و قدردانی

اکنون که به یاری پروردگار این پایان نامه به اتمام رسیده است بر خود لازم می‌دانم از اساتید راهنما، آقایان دکتر حسین نیکوفرد و دکتر ناصر گودرزی که همواره حامی و پشتیبان من بوده‌اند و همچنین استاد مشاور سرکار خانم دکتر زهرا کلانتر خاضعانه قدردانی نمایم. از تمام اساتید گروه شیمی که لیاقت شاگردی در محضرشان را داشتم سپاسگزارم. و در نهایت از تمام دوستانم که با تکیه بر دوستی و محبتشان توانستم این پروژه را به پایان برسانم، به خصوص خانم‌ها شیرپور، یامی، دشت بزرگی و آقای بهزاد احمدی تشکر می‌کنم.

دانشجو تأیید می نماید که مطالب مندرج در این پایان نامه نتیجه تحقیقات خودش می باشد و در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده است.

کلیه حقوق مادی مترتب از نتایج مطالعات، آزمایشات و نوآوری ناشی از تحقیق موضوع این پایان نامه متعلق به دانشگاه صنعتی شاهرود می باشد .

خرداد ۱۳۸۹

## چکیده

امروزه ارتباط کمی ساختار - خاصیت (QSPR) برای مطالعه‌ی وابستگی خواص مختلف فیزیکوشیمیایی ترکیبات با ساختار مولکولی استفاده شده است. در بخش اول این پروژه، از رگرسیون خطی چندگانه (MLR) به عنوان یک روش خطی و از شبکه‌ی عصبی مصنوعی (ANN) به عنوان روشی غیر خطی برای پیش‌بینی ثابت هنری ۳۳۳ ترکیب آلی استفاده گردید. سری داده‌ها به سری‌های آموزش (۲۶۳ ترکیب)، آزمون (۴۰ ترکیب) و ارزیابی بیرونی (۳۰ ترکیب) تقسیم شدند. یک شبکه‌ی عصبی پیش‌خور سه لایه پس انتشار خطا با استفاده از چهار توصیف‌کننده‌ی مولکولی بدست آمده در مدل MLR، ایجاد گردید. مشخصات آماری بدست آمده به وسیله‌ی مدل خطی چندگانه ( $R^2=0/942$ ،  $MSE=0/109$  برای سری آزمون) توانایی پیش‌بینی آنرا نشان داد، درحالی‌که توانایی پیش‌بینی ANN تا اندازه‌ای بهتر می‌باشد ( $R^2=0/982$ ،  $MSE=0/039$  برای سری آزمون).

در بخش دوم این پروژه، روش‌های الگوریتم ژنتیک بر اساس رگرسیون خطی چندگانه (GA-MLR) و الگوریتم ژنتیک بر اساس شبکه‌ی عصبی مصنوعی (GA-ANN) برای ایجاد مدل‌های QSPR میان توصیف‌کننده‌ها و ضریب شکست ۱۴۴ پلیمر استفاده شدند. توصیف‌کننده‌های Mor15e، R2e، R2p و J3D به وسیله‌ی الگوریتم ژنتیک برای تشکیل مدل‌های خطی و غیر خطی به منظور پیش‌بینی ضریب شکست انتخاب شدند. مجذور میانگین مربعات خطای سری‌های آموزش، آزمون و ارزیابی برای مدل ANN به ترتیب، ۰/۰۱۸، ۰/۰۲۳، ۰/۰۱۸ و همچنین ضرایب تعیین ( $R^2$ )، ۰/۹۲۹، ۰/۸۸۱ و ۰/۹۳۱ هستند.

واژه‌های کلیدی: QSPR، ثابت هنری، شبکه‌ی عصبی مصنوعی، ضریب شکست، الگوریتم ژنتیک بر اساس رگرسیون خطی چندگانه

## فهرست مطالب

### فصل اول: مقدمه

- ۱-۱- قانون هنری ..... ۱
- ۲-۱- ضریب شکست ..... ۲
- ۳-۱- مروری بر کارهای انجام شده ..... ۴

### فصل دوم: کمومتریکس

- ۱-۲- کمومتریکس ..... ۷
- ۱-۱-۲- ارتباط کمی ساختار-خاصیت ..... ۸
- ۲-۱-۲- ارتباط کمی ساختار-فعالیت ..... ۹
- ۳-۱-۲- رگرسیون خطی چندگانه ..... ۱۱
- ۴-۱-۲- انتخاب سری داده‌ها ..... ۱۴
- ۵-۱-۲- انتخاب توصیف‌کننده‌ها ..... ۱۴
- ۱-۵-۱-۲- توصیف‌کننده‌های توپولوژیکی ..... ۱۵
- ۲-۵-۱-۲- توصیف‌کننده‌های هندسی ..... ۱۵
- ۳-۵-۱-۲- توصیف‌کننده‌های الکترونی ..... ۱۶
- ۴-۵-۱-۲- توصیف‌کننده‌های فیزیکوشیمیایی ..... ۱۶
- ۶-۱-۲- تجزیه و تحلیل آماری توصیف‌کننده‌ها ..... ۱۷
- ۷-۱-۲- آنالیز مدل‌های آماری و انتخاب مدل مناسب ..... ۱۷
- ۱-۷-۱-۲- روش ورود اجباری ..... ۱۸
- ۲-۷-۱-۲- روش انتخاب جلوبرنده ..... ۱۸
- ۳-۷-۱-۲- روش عقب برنده ..... ۱۸
- ۴-۷-۱-۲- روش مرحله ای ..... ۱۹
- ۵-۷-۱-۲- روش حذفی ..... ۱۹
- ۸-۱-۲- نرم افزارهای مورد نیاز ..... ۱۹
- ۱-۸-۱-۲- بسته‌ی نرم افزاری Hyperchem ..... ۱۹
- ۲-۸-۱-۲- نرم افزار Dragon ..... ۲۰
- ۳-۸-۱-۲- بسته‌ی نرم افزاری SPSS ..... ۲۰
- ۴-۸-۱-۲- بسته‌ی نرم افزاری MATLAB R2008a ..... ۲۱



## فصل سوم: الگوریتم ژنتیک

۲۲	۱-۳- مقدمه
۲۳	۲-۳- الگوریتم ژنتیک
۲۴	۱-۲-۳- الگوی طبیعی الگوریتم ژنتیک
۲۶	۲-۲-۳- تاریخچه و اساس الگوریتم ژنتیک
۲۸	۳-۲-۳- مبانی الگوریتم ژنتیک
۲۸	۴-۲-۳- مراحل کلی الگوریتم ژنتیک
۳۰	۳-۳- رمزبندی یک کروموزوم
۳۰	۱-۳-۳- رمزبندی دودویی (باینری)
۳۱	۲-۳-۳- رمزبندی تبدیل
۳۱	۳-۳-۳- رمز بندی مقداری
۳۲	۴-۳-۳- رمزبندی درختی
۳۳	۴-۳- عملگرهای GA
۳۳	۱-۴-۳- پیوند
۳۴	۱-۱-۴-۳- پیوند در رمزبندی دودویی
۳۵	۲-۱-۴-۳- پیوند در رمزبندی تبدیل
۳۵	۳-۱-۴-۳- پیوند در رمزبندی مقداری
۳۵	۴-۱-۴-۳- پیوند در رمز بندی درختی
۳۶	۲-۴-۳- جهش
۳۶	۱-۲-۴-۳- جهش در رمزبندی دودویی
۳۶	۲-۲-۴-۳- جهش در رمزبندی تبدیل
۳۷	۳-۲-۴-۳- جهش در رمزبندی مقداری
۳۷	۴-۲-۴-۳- جهش در رمزبندی درختی
۳۷	۵-۳- عملگرهای ژنتیک از دیدگاه نحوه اعمال بر کروموزومها
۳۹	۶-۳- پارامترهای GA
۳۹	۱-۶-۳- احتمال (نرخ) پیوند
۴۰	۲-۶-۳- احتمال (نرخ) جهش
۴۰	۳-۶-۳- اندازه جمعیت
۴۱	۴-۶-۳- شایستگی
۴۱	۱-۴-۶-۳- تابع شایستگی

- ۴۲.....انتخاب ۵-۶-۳
- ۴۳.....روش انتخاب چرخ گردان ۱-۵-۶-۳
- ۴۳.....اجرای الگوریتم ژنتیک ۷-۳
- ۴۵.....الگوریتم ژنتیک و QSPR ۸-۳

### فصل چهارم: شبکه عصبی مصنوعی

- ۴۶.....۱-۴- مقدمه‌های بر شبکه‌های عصبی
- ۴۸.....۲-۴- نرون بیولوژیکی
- ۴۹.....۳-۴- ساختار و عملکرد نرون مصنوعی
- ۵۲.....۴-۴- ویژگی‌های شبکه‌های عصبی
- ۵۲.....۱-۴-۴- قابلیت یادگیری
- ۵۲.....۲-۴-۴- پردازش موازی
- ۵۳.....۳-۴-۴- مقاوم بودن
- ۵۳.....۴-۴-۴- قابلیت تعمیم
- ۵۳.....۵-۴- یادگیری در شبکه‌های عصبی
- ۵۳.....۱-۵-۴- یادگیری با ناظر
- ۵۴.....۲-۵-۴- یادگیری بدون ناظر
- ۵۴.....۳-۵-۴- یادگیری تقویتی
- ۵۴.....۶-۴- معماری شبکه‌های عصبی مصنوعی
- ۵۶.....۷-۴- شبکه‌های پرسپترون چند لایه
- ۵۶.....۸-۴- الگوریتم‌های آموزشی شبکه‌های عصبی
- ۵۷.....۱-۸-۴- الگوریتم آموزشی پس انتشار خطا
- ۵۸.....۲-۸-۴- الگوریتم لونبرگ-مارکورت
- ۶۰.....۹-۴- ارتقای عمومیت

### فصل پنجم: مدل سازی QSPR ثابت هنری

- ۶۲.....۱-۵- مدل سازی QSPR ثابت هنری
- ۶۲.....۱-۱-۵- سری داده‌ها
- ۷۸.....۲-۱-۵- مراحل مدل سازی
- ۷۸.....۱-۲-۱-۵- رسم و بهینه‌سازی ساختار مولکول‌ها
- ۷۸.....۲-۲-۱-۵- محاسبه‌ی توصیف‌کننده‌های مولکولی

۷۹.....	انتخاب توصیف‌کننده‌های مهم ۳-۲-۱-۵
۷۹.....	مدل رگرسیون خطی چندگانه ۴-۲-۱-۵
۸۶.....	پیش‌بینی ثابت هنری با شبکه عصبی مصنوعی ۵-۲-۱-۵
۸۷.....	انتخاب تعداد لایه‌های شبکه عصبی ۱-۵-۲-۱-۵
۸۷.....	انتخاب تابع انتقال ۲-۵-۲-۱-۵
۸۸.....	انتخاب تعداد گره‌های لایه پنهان و تعداد دوره‌های آموزش ۳-۵-۲-۱-۵
۸۹.....	معماری شبکه‌ی عصبی مصنوعی بهینه شده ۴-۵-۲-۱-۵
۹۰.....	فرایند ایجاد تصادفی-Y ۵-۵-۲-۱-۵
۹۶.....	بررسی نتایج ۲-۵
۹۶.....	ارزیابی مدل‌های شبکه‌ی عصبی و رگرسیون خطی چندگانه ۱-۲-۵
۱۰۱.....	تجزیه و تحلیل توصیف‌کننده‌های مدل ۲-۲-۵

### فصل ششم: مدل‌سازی QSPR ضریب شکست

۱۰۴.....	مقدمه ۱-۶
۱۰۴.....	سری داده‌ها ۲-۶
۱۱۶.....	مدل‌سازی ضریب شکست ۳-۶
۱۱۶.....	رسم و بهینه‌سازی ساختار مولکول‌ها ۱-۳-۶
۱۱۷.....	محاسبه‌ی توصیف‌کننده‌های مولکولی ۲-۳-۶
۱۱۷.....	مدل الگوریتم ژنتیکی بر اساس رگرسیون خطی چندگانه ۳-۳-۶
۱۱۷.....	رمزبندی ۱-۳-۳-۶
۱۱۸.....	تابع شایستگی ۲-۳-۳-۶
۱۱۸.....	جمعیت ۳-۳-۳-۶
۱۱۹.....	روش انتخاب ۴-۳-۳-۶
۱۱۹.....	عملگر پیوند ۵-۳-۳-۶
۱۲۰.....	عملگر جهش ۶-۳-۳-۶
۱۲۳.....	مدل رگرسیون خطی چندگانه ۴-۳-۶
۱۲۵.....	مدل شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های الگوریتم ژنتیک ۵-۳-۶
۱۲۵.....	انتخاب تابع انتقال ۱-۵-۳-۶
۱۲۶.....	انتخاب تعداد گره‌های لایه پنهان و تعداد دوره‌های آموزش ۲-۵-۳-۶
۱۲۷.....	معماری شبکه‌ی عصبی مصنوعی بهینه شده ۳-۵-۳-۶
۱۲۸.....	فرایند ایجاد تصادفی-Y ۴-۵-۳-۶

۱۳۱	.....	۴-۶- بررسی نتایج
۱۳۱	.....	۴-۶-۱- ارزیابی مدل‌های GA-ANN و GA-MLR
۱۳۶	.....	۴-۶-۲- تجزیه و تحلیل توصیف‌کننده‌های مدل
۱۳۹	.....	آینده نگری
۱۴۰	.....	فهرست منابع

## فهرست اشکال

- شکل (۳-۱) - پیوند باعث جابجا شدن قطعه‌ای از DNA بین دو کروموزوم و ایجاد کروموزوم‌های نوترکیب می‌شود ..... ۲۶
- شکل (۳-۲) - پیوند یک نقطه‌ای ..... ۳۴
- شکل (۳-۳) - پیوند دو نقطه‌ای ..... ۳۴
- شکل (۳-۴) - پیوند یکنواخت ..... ۳۴
- شکل (۳-۵) - پیوند عددی ..... ۳۵
- شکل (۳-۶) - مثالی از پیوند درختی ..... ۳۵
- شکل (۳-۷) - فلوجارت یک الگوریتم ژنتیکی ساده ..... ۴۵
- شکل (۴-۱) - ساختار نرون طبیعی ..... ۴۸
- شکل (۴-۲) - ساختار یک نرون عصبی مصنوعی ..... ۴۹
- شکل (۴-۳) - شبکه‌های عصبی مصنوعی سه لایه ..... ۵۰
- شکل (۴-۴) - توابع تبدیل (الف) خطی، (ب) سیگموئیدی ..... ۵۱
- شکل (۵-۱) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل لگاریتم سیگموئید ..... ۸۸
- شکل (۵-۲) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل تانژانت سیگموئید ..... ۸۸
- شکل (۵-۳) - ساختار هندسی شبکه عصبی استفاده شده ..... ۹۰
- شکل (۵-۴) - نمودار مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  بر حسب مقادیر واقعی با روش MLR برای (الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۹۷
- شکل (۵-۵) - نمودار مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  بر حسب مقادیر واقعی با روش ANN برای (الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۹۸
- شکل (۵-۶) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  با روش MLR از مقادیر واقعی بر حسب مقادیر تجربی برای (الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۹۹
- شکل (۵-۷) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  با روش ANN از مقادیر واقعی بر حسب مقادیر تجربی برای (الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۱۰۰
- شکل (۶-۱) - تغییرات مقادیر خطا بر حسب تعداد جمعیت ..... ۱۱۹
- شکل (۶-۲) - تغییرات خطا بر حسب مقادیر عملگر ..... ۱۲۰

- شکل (۳-۶) - تغییرات خطا بر حسب مقادیر عملگر جهش ..... ۱۲۰
- شکل (۴-۶) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل لگاریتم سیگموئید ..... ۱۲۶
- شکل (۵-۶) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل تانژانت سیگموئید ..... ۱۲۶
- شکل (۶-۶) - ساختار هندسی شبکه عصبی استفاده شده ..... ۱۲۷
- شکل (۷-۶) - نمودار مقادیر پیش‌بینی شده‌ی ضریب شکست بر حسب مقادیر واقعی با روش GA-MLR (برای الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۱۳۲
- شکل (۸-۶) - نمودار مقادیر پیش‌بینی شده‌ی ضریب شکست بر حسب مقادیر واقعی با روش GA-ANN (برای الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۱۳۳
- شکل (۹-۶) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی ضریب شکست با روش GA-MLR از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۱۳۴
- شکل (۱۰-۶) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی ضریب شکست با روش GA-ANN از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی ..... ۱۳۵

## فهرست جداول

- جدول (۳-۱) - کروموزوم‌هایی با رمزبندی دودویی ..... ۳۱
- جدول (۳-۲) - کروموزوم‌هایی با رمزبندی تبدیل ..... ۳۱
- جدول (۳-۳) - کروموزوم‌هایی با رمزبندی مقداری ..... ۳۲
- جدول (۳-۴) - کروموزوم‌هایی با رمزبندی درختی ..... ۳۲
- جدول (۳-۵) - پیوند یک نقطه‌ای دو کروموزوم ..... ۳۳
- جدول (۳-۶) - مثالی از جهش در دو کروموزوم فرزند ..... ۳۶
- جدول (۴-۱) - توابع تبدیل استفاده شده ..... ۵۱
- جدول (۵-۱) - سری ترکیبات مورد بررسی ..... ۶۲
- جدول (۵-۲) - مشخصات مدل حاصل از رگرسیون مرحله‌ای ..... ۸۰
- جدول (۵-۳) - ماتریس همبستگی توصیف کننده‌های وارد شده در مدل منتخب ..... ۸۰
- جدول (۵-۴) - مقادیر توصیف کننده‌های وارد شده در مدل منتخب ..... ۸۱
- جدول (۵-۵) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان ..... ۸۷
- جدول (۵-۶) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل تانژانت سیگموئید و تعداد گره‌های مختلف در لایه پنهان ..... ۸۷
- جدول (۵-۷) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان و چرخه‌های مختلف ..... ۸۹
- جدول (۵-۸) - پارامترهای شبکه‌ی بهینه شده ..... ۹۰
- جدول (۵-۹) - نتایج تصادفی کردن -Y ..... ۹۰
- جدول (۵-۱۰) - مقادیر تجربی و پیش‌بینی شده‌ی ثابت‌های قانون هنری بوسیله‌ی روش‌های رگرسیون خطی چندگانه و شبکه‌ی عصبی مصنوعی ..... ۹۱
- جدول (۵-۱۱) - پارامترهای آماری مدل شبکه عصبی و روش رگرسیون خطی چندگانه ..... ۹۶
- جدول (۶-۱) - نام، ساختار و ضریب شکست سری داده‌ها ..... ۱۰۵
- جدول (۶-۲) - مقادیر پارامترهای الگوریتم ژنتیک ..... ۱۲۱
- جدول (۶-۳) - مقادیر توصیف کننده‌های انتخاب شده با الگوریتم ژنتیک ..... ۱۲۱
- جدول (۶-۴) - مشخصات مدل حاصل از روش MLR ..... ۱۲۴
- جدول (۶-۵) - ماتریس همبستگی توصیف کننده‌های انتخاب شده با الگوریتم ژنتیک ..... ۱۲۴

- جدول (۶-۶) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان ..... ۱۲۵
- جدول (۶-۷) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل تانژانت سیگموئید و تعداد گره‌های مختلف در لایه پنهان ..... ۱۲۵
- جدول (۶-۸) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئیدی و تعداد گره‌های مختلف در لایه پنهان و چرخه‌های مختلف ..... ۱۲۷
- جدول (۶-۹) - پارامترهای شبکه‌ی بهینه ..... ۱۲۸
- جدول (۶-۱۰) - نتایج تصادفی کردن Y- ..... ۱۲۸
- جدول (۶-۱۱) - مقادیر تجربی و پیش‌بینی شده‌ی ضریب شکست بوسیله‌ی روش‌های GA-MLR و GA-ANN ..... ۱۲۹
- جدول (۶-۱۲) - پارامترهای آماری مدل شبکه عصبی و GA-MLR ..... ۱۳۱



فصل اول

مقدمه

## ۱-۱- قانون هنری<sup>۱</sup>

توزیع گونه‌ها بین آب و هوا یکی از مهمترین فرآیندهایی است که انتقال مواد شیمیایی را در طبیعت تعیین می‌کند و پیش‌بینی آن نیازمند آگاهی از خواص فیزیکی و شیمیایی گونه‌ها می‌باشد. یکی از خواص ترمودینامیکی، حلالیت در آب است که در محاسبه‌ی کاهش پس‌ماندها و مطالعات آلاینده‌های منابع آبی نقش ضروری دارد. رفتار توزیع اجزا در سیستم مایع - هوا توسط قانون هنری توصیف می‌گردد، این قانون بیان‌گر این است که در دمای ثابت فشار جزئی بخار گونه‌ی حل شده در بالای محلول رقیق ایده آل با کسر مولی آن گونه در محلول متناسب است:

$$p_i = k_i x_i^l \quad \text{یا} \quad k_i = \frac{p_i}{x_i^l} \quad (1-1)$$

که  $p_i$  فشار جزئی بخار و  $x_i^l$  کسر مولی گونه‌ی  $i$  در محلول می‌باشد [۱].

ثابت تناسب  $k_i$  به عنوان ثابت قانون هنری برای ترکیبات بیان می‌شود که اطلاعات مفیدی را برای ارزیابی توزیع آلاینده‌ها در اختیار ما قرار می‌دهد و تفکیک تعادلی یک ماده شیمیایی را میان فاز مایع و گاز به‌طور کمی بیان می‌کند. یک ماده شیمیایی با مقدار کم این ثابت، تمایل به تجمع در فاز محلول را دارد درحالی‌که ترکیبی با مقدار ثابت بزرگ، بیشتر تمایل به حضور در فاز گازی خواهد داشت [۲]. آگاهی از حلالیت هیدروکربن‌ها در فاز محلول در عرصه‌های متنوعی از صنعت نفت، مهندسی شیمی و ژئوشیمی از قبیل محاسبات طراحی و انتخاب رژیم لوله‌های تاسیسات تولید و فرآوری، ذخیره گاز طبیعی، جداسازی اسید گاز در حفره‌های زیر زمینی و حذف آلاینده‌های فرار و نیمه فرار مهم می‌باشد. روش‌های زیادی برای تعیین ثابت قانون هنری وجود دارند که جزئیات آن در تعدادی از منابع بیان شده‌اند [۳]. فقط محلول‌های رقیق در فشارهای نسبتاً پایین، از قانون هنری به خوبی پیروی می‌کنند. گازهایی که انحلال‌پذیری آنها بسیار زیاد است، عموماً با حلال خود ترکیب می‌شوند، به عنوان مثال گاز هیدروژن کلرید وقتی در آب حل می‌شود، با آن ترکیب شده و

---

۱- Henry's law

هیدروکلریک اسید تولید می‌کند که این محلول‌ها از قانون هنری پیروی نمی‌کنند. به‌طور کلی تعیین ثابت قانون هنری به علت اندازه‌گیری‌های آزمایشگاهی پرهزینه و تبعات اقتصادی آن، فقدان اندازه‌گیری‌های صحیح در غلظت‌های پایین و بویژه برای ترکیبات با حلالیت پایین، آسان نمی‌باشد.

## ۱-۲- ضریب شکست<sup>۱</sup>

ضریب شکست یک محیط به صورت نسبت سرعت یک موج نوری (یا صوتی) در محیط مرجع به سرعت نور در آن محیط بیان می‌شود که در متون علمی غالباً از سرعت نور در خلا به عنوان مرجع استفاده شده است. این ضریب همیشه بزرگ‌تر از یک است زیرا با ورود نور به محیط ماده سرعت آن بر اثر برهم‌کنش با اتم‌های تشکیل دهنده‌ی ماده کاهش می‌یابد درحالی‌که این برهم‌کنش‌ها در محیط مرجع وجود ندارد. بنابراین ویژگی‌های ساختاری یک گونه بر ضریب شکست آن بسیار تاثیر گذار خواهد بود. این خاصیت به روش تجربی توسط دستگاهی بنام رفاکتومتر<sup>۲</sup> اندازه‌گیری می‌شود. ضریب شکست، یک خاصیت نوری اصلی و اساسی برای پلیمرها می‌باشد که به خواص مغناطیسی، الکترونی و نوری مربوط می‌شود [۴]. این ضریب با تغییر فرکانس موج ورودی و چگالی محیط تغییر می‌کند و برای مطالعه خواص فیزیکی، شیمیایی و مولکولی پلیمرها با تکنیک‌های نوری با اهمیت است. همچنین افزایش ضریب شکست در پراش نور برای تعیین جرم مولکولی، اندازه و شکل پلیمرها مورد استفاده قرار می‌گیرد [۵]. بنابراین ترکیبات پلیمری به آسانی می‌توانند شرایط مورد نیاز برای کاربردهای گوناگون اپتیکی را بدست آورند [۶-۸].

با استفاده از داده‌های تجربی برای ثابت قانون هنری ترکیبات و ضریب شکست پلیمرها، می‌توان روش‌های پیش‌بینی کننده‌ای مبتنی بر ابزار نظری و نرم افزاری ابداع نمود که در تایید و

---

۱- Refractive index (RI)

۲- Refractometer

تقویت داده‌های موجود بسیار کمک کرده و با بهره‌گیری از اطلاعات تجربی، در مورد خواص ترکیبات جدید اظهار نظر کنند.

در اکثر مواقع آموزش ریاضیات و آمار برای شیمیدان‌ها کافی نبوده و بنابراین یکی از اهداف اولیه‌ی روش‌های محاسباتی کمومتریکس<sup>۱</sup> اجرایی کردن روش‌های ریاضیاتی پیچیده برای یک شیمیدان است. علاوه بر روش‌های ریاضی- آماری مباحث کمومتریکس مربوط به استفاده از داده‌های شیمیایی، اسپکتروسکوپی و نیز روش‌های هوش مصنوعی<sup>۲</sup> می‌شود. به دلیل پیشرفت سریع دستگاه‌های هوشمند و خودکار آزمایشگاهی و همچنین امکان استفاده از کامپیوترهای قدرتمند و نرم افزارهای ساده، کمومتریکس به عنوان یک دانش مستقل گسترش روزافزونی داشته است. یکی از زمینه‌های مهم کاربرد کمومتریکس در مطالعاتی است که خواص مولکول‌ها را به ویژگی‌های ساختاری آنها نسبت می‌دهد. از نظر شیمیدان‌ها فعالیت‌ها و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. این نوع از مطالعات به بررسی کمی ارتباط ساختار- فعالیت<sup>۳</sup> و همچنین بررسی کمی ارتباط ساختار- خاصیت<sup>۴</sup> می‌پردازند. هدف از بررسی ارتباط کمی ساختار- خاصیت که در این پایان‌نامه بحث می‌گردد پیدا کردن رابطه‌ای میان خواص شیمیایی یک مولکول با پارامترهای ساختاری آن می‌باشد. روش‌های متنوعی برای برقراری این رابطه بکار می‌روند که از پرکاربردترین آنها می‌توان به روش‌های رگرسیون خطی چندگانه<sup>۵</sup>، رگرسیون اجزای اصلی<sup>۶</sup> و حداقل مربعات جزئی<sup>۷</sup> اشاره کرد. مولفه‌های هوش محاسباتی<sup>۸</sup> شامل شبکه‌ی عصبی مصنوعی<sup>۹</sup>، الگوریتم‌های ژنتیک<sup>۱۰</sup> و منطق فازی<sup>۱۱</sup> نیز در ایجاد رابطه‌ی غیرخطی میان ساختار و خاصیت ترکیبات استفاده می‌شوند. نتایج این مطالعات علاوه

---

۱- Chemometrics

۲- Artificial Intelligence (AI)

۳- Quantitative Structure -Activity Relationship (QSAR)

۴- Quantitative Structure -Property Relationship (QSPR)

۵- Multiple Linear Regression (MLR)

۶- Principal Component Regression (PCR)

۷- Partial Least Square (PLS)

۸- Computational Intelligence (CI)

۹- Artificial Neural Network (ANN)

۱۰- Genetic Algorithm (GA)

۱۱- Fuzzy logic

بر شفاف سازی نحوه ی ارتباط بین خواص مولکول ها و ویژگی های ساختاری آنها به پژوهشگران در پیش بینی رفتار مولکول های جدید براساس رفتار مولکول های مشابه کمک می کند.

### ۳-۱- مروری بر کارهای انجام شده

در سال ۲۰۰۸ مدرسی برای مدل سازی وابستگی ثابت قانون هنری متیل کتون ها در محلول آبی سولفات سدیم به دما و غلظت از شبکه عصبی مصنوعی استفاده نمود. با استفاده از روش پس انتشار ساختار شبکه به صورت ۱-۲-۲ بهینه شد و شاخص انحراف میانگین نسبی<sup>۱</sup> تمام سیستم های مطالعه شده کمتر از یک درصد بود. سری داده ها شامل دو سری مجزا بوده و قدرت پیش بینی روش های شبکه عصبی عادی و شبکه عصبی بدست آمده با شاخص انحراف میانگین نسبی برای سری داده ها مقایسه و نشان داده شد که با این شاخص می توان شبکه ای با معماری مناسب تر را برگزید [۹]. مطالعه ی رابطه ی کمی ساختار و ثابت های قانون هنری هیدروکربن های آلیفاتیک توسط پابلو<sup>۲</sup> و همکارانش در سال ۲۰۰۷ جهت مقایسه با مطالعات قبلی انجام گرفت. با استفاده از الگوریتم روش جایگزینی<sup>۳</sup> روی ۱۵۰ ترکیب، بهترین مدل شامل هفت توصیف کننده<sup>۴</sup> با ضریب همبستگی<sup>۵</sup> ۰/۹۹۶ و مجذور میانگین مربعات خطای<sup>۶</sup> ۰/۰۶۳ بدست آمد [۱۰].

در سال ۲۰۰۲ یو<sup>۷</sup> و همکارانش از روش های QSPR برای پیش بینی ثابت قانون هنری تعدادی از ترکیبات آلی استفاده کردند. سری داده ها به دو سری آموزش<sup>۸</sup> شامل ۳۵۲ ترکیب و سری آزمون<sup>۹</sup> با ۵۸ ترکیب تقسیم شدند. در این مطالعه از رگرسیون خطی چندگانه مدلی بدست آمد که شامل ده توصیف کننده با ضریب همبستگی<sup>۵</sup> ۰/۹۹۰، میانگین مربعات خطای<sup>۶</sup> ۰/۳۲۷ و آماره F برابر با ۱۹۰۴

۱- Average Relative Deviation (ARD)

۲- Pablo

۳- Replacement method

۴- Descriptor

۵- Correlation coefficient

۶- Root Mean Square Errors (RMSE)

۷- Xiaojun Yao

۸- Train set

۹- Test set

بود. برای ایجاد مدل غیر خطی توصیف‌کننده‌ها به شبکه تابع پایه شعاعی<sup>۱</sup> وارد شدند و شبکه‌ی بهینه شده با ساختار ۱-۱۷-۱۰ و میانگین مربعات خطای برابر ۰/۲۹۰۵ و ۰/۳۲۶۷ به ترتیب برای سری آموزش و آزمون، بدست آمد [۱۱].

تا کنون مطالعات کمی در مورد ضریب شکست پلیمرها در حوزه کمومتریکس صورت گرفته است. در سال ۲۰۰۸ جی یو<sup>۲</sup> و همکارانش از مدل‌های خطی و غیر خطی QSPR برای پیش‌بینی ضریب شکست ۱۲۰ پلیمر استفاده کردند. برای ایجاد رابطه‌ی خطی میان توصیف‌کننده‌های محاسبه شده و ضریب شکست پلیمرها از رگرسیون خطی چندگانه مرحله‌ای با ارزیابی تقاطعی<sup>۳</sup> بر روی سری آموزشی شامل ۱۰۰ پلیمر استفاده شد. بهترین مدل دارای پنج توصیف‌کننده با ضریب تعیین ۰/۹۳۴ و آماره‌ی F برابر با ۳۳۹/۴۶ و خطای استاندارد ۰/۰۱۶۴ بدست آمد. روش استفاده شده برای یافتن مدل غیر خطی، شبکه‌ی عصبی مصنوعی بوده که در این مطالعه از الگوریتم شبه نیوتن<sup>۴</sup> BFGS استفاده شده است. نتایج حاصل از این مدل دارای ضریب تعیین<sup>۵</sup> ۰/۹۷۱ و خطای استاندارد ۰/۰۱۱۸ می‌باشد [۱۲].

یو<sup>۶</sup> و همکارانش در سال ۲۰۰۷ برای پیش‌بینی ضریب شکست وینیل پلیمرها از نظریه‌ی تابع چگالی<sup>۷</sup> برای محاسبه‌ی توصیف‌کننده‌های شیمی کوانتومی مونومرهای این دسته از پلیمرها استفاده کردند. در این روش از نرم افزار Gaussian استفاده شد. بهترین مدل برای سری آموزشی با ۴۷ ترکیب شامل چهار توصیف‌کننده کوانتومی با ضریب همبستگی ۰/۹۶۳ و خطای استاندارد ۰/۰۲۰ می‌باشد. با روش ارزیابی تقاطعی،  $q^2$  برابر با ۰/۹۰۳ بدست آمد که از نظر آماری مقادیر  $q^2 > 0/5$  نتیجه‌ی خوب را بیان می‌کند [۱۳].

---

۱- Radial Basis Function (RBF)

۲- Jie Xu

۳- Cross-validation

۴- Broyden-Fletcher-Goldfarb-Shanno

۵- Coefficient of determination

۶- Xinliang Yu

۷- Density Functional Theory (DFT)

در سال ۲۰۰۲ گارسیا<sup>۱</sup> و همکارش ضریب شکست یک سری از پلیمرهای خطی را با بکار  
بردن اندیس‌های تئوری گراف<sup>۲</sup> بررسی کردند. در این روش با استفاده از اندیس‌های توپولوژیکی<sup>۳</sup>  
محاسبه شده برای مونومر پلیمرها، مدلی شامل ده توصیف‌کننده‌ی توپولوژیکی با ضریب همبستگی  
۰/۹۸۱ بدست آمد [۱۴].

---

۱- Garcia

۲- Graph theoretical indices

۳- Topological indices

فصل دوم

کمومتریکس



## ۲-۱- کمومتریکس

با توجه به گسترش سیستم‌ها و فرآیندهای شیمیایی و از طرفی هزینه‌های بالای آنها، استفاده از روش‌های محاسباتی برای بهبود بخشیدن به فرآیندهای اندازه‌گیری و استخراج اطلاعات شیمیایی بیشتر از داده‌های موجود بسیار مفید می‌باشد. تعدادی از گروه‌های تجزیه‌ای در دهه‌ی ۷۰ میلادی از روش‌های ریاضیاتی و آماری که امروزه به روش‌های کمومتریکس مشهور هستند، استفاده کرده‌اند.

کمومتریکس در سال ۱۹۷۴ توسط اسوانت ولد<sup>۱</sup> سوئدی معرفی شده است [۱۵]. با تاسیس انجمن بین‌المللی کمومتریکس در سال ۱۹۷۴، کمومتریکس به عنوان یک رشته علمی معرفی گردید. اولین مجلات کمومتریکس با عنوان *Journal of Chemometrics (Wiley)* و *Chemometrics and Intelligent Laboratory Systems (Elsevier)* از سال ۱۹۸۷ شروع به کار کردند. تعریف فعلی از کمومتریکس عبارت است از: شاخه‌ای از شیمی که از روش‌های ریاضی و آماری برای الف) انتخاب و طراحی مطلوب‌ترین روش‌ها و آزمایش‌ها، ب) فراهم آوردن بیشترین اطلاعات شیمیایی به وسیله‌ی تجزیه و تحلیل داده‌های شیمیایی و پ) شناخت سیستم‌های شیمیایی سود می‌برد [۱۶].

کاربردهای عمومی روش‌های کمومتریکس توسعه ارتباط کمی ساختار- فعالیت و یا ارتباط کمی ساختار- خاصیت یا ارزیابی داده‌های شیمیایی هستند. داده‌های فراوانی که به وسیله‌ی ظهور تجهیزات مدرن تجزیه‌ای تولید شدند مهمترین دلیل برای افزایش کاربرد روش‌های کمومتریکس نزد شیمیدانان گردید. با در دسترس قرار گرفتن رایانه‌های شخصی در اوایل دهه‌ی ۸۰ میلادی دوره‌ی نوینی برای کشف، پردازش و تفسیر داده‌های شیمیایی آغاز گردید. اولین مقاله در مبحث کمومتریکس در ایران مربوط به کار جلالی هروی و گرکانی نژاد از دانشگاه باهنر کرمان است که با عنوان پیش‌بینی اندیس بازداری کروماتوگرافی گازی تعدادی از مشتقات بنزن در سال ۱۹۹۳ به چاپ رسیده است [۱۷].

---

<sup>۱</sup>- Svant Wold

روند کلی در کمومتریکس عبارتست از:

- ۱- جمع آوری داده‌ها برای نمونه‌های شناخته شده
- ۲- ایجاد مدل ریاضی بر اساس تکنیک‌های چند متغیری یا شبکه‌های عصبی
- ۳- تفسیر پارامترهای مدل بر اساس قواعد یا اصول شیمی
- ۴- کاربرد مدل برای نمونه‌های جدید

## ۲-۱-۱- ارتباط کمی ساختار - خاصیت

در نیمه‌ی دوم قرن نوزدهم نظریه ساختاری<sup>۱</sup> در شیمی آلی توسعه یافت و منجر به این نتیجه شد که: "خواص شیمیایی، فیزیکی و بیولوژیکی تابعی از تغییرات ساختاری هستند". هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده‌ی مولکول انجام گیرد، به آن ارتباط کمی ساختار - خاصیت می‌گویند، که در واقع رابط بین خواص فیزیکوشیمیایی و ساختار شیمیایی ترکیبات است. بعضی از این خواص شامل ضریب شکست پلیمر، دمای خود اشتعالی هیدروکربن‌ها، کشش سطحی مایعات، ثابت قانون هنری ترکیبات آلی و غیره هستند. با توسعه روش‌های اندازه‌گیری خواص فیزیکوشیمیایی، داده‌های زیادی جمع‌آوری شدند که با استفاده از مطالعات کمومتریکس تلاش‌ها برای بسط مدل‌های کمی بر اساس وابستگی ساختاری این خواص صورت گرفت. بیش از چهل سال است که QSPR به عنوان یک روش اصلی و مهم در تحقیقات شیمیایی توسعه یافته است. رشته‌های گوناگونی از جمله شیمی فیزیک آلی، شیمی معدنی، اسپکتروسکوپی و شیمی محیط زیست از این روش جهت پیش‌بینی خواص مختلف استفاده کرده‌اند [۱۸]. با وجود کاربرد وسیع این روش، به‌طور کلی قواعد اصلی و روش کار در همه‌ی آنها یکسان یا مشابه است.

مدل‌های QSPR را می‌توان بر اساس پارامترهایی که به صورت تجربی بدست آمده و یا از طریق تئوری محاسبه شده‌اند، به زبان ریاضی بیان نمود. به‌طور کل اگر نتایج این نوع مطالعات همبستگی

---

<sup>۱</sup> - Structural theory

قابل قبولی را ارائه کند علاوه بر شفاف‌سازی نحوه‌ی ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمانی آنها، به پژوهش‌گران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند [۱۹]. از مزایای این روش‌ها وارد شدن تعداد کمی از توصیف‌کننده‌ها در مدل است که از لحاظ نظری مفهوم فیزیکی مشخصی را برای ترکیبات یک سری بیان می‌کنند.

## ۲-۱-۲- ارتباط کمی ساختار- فعالیت

شیمی مولکول‌های بیولوژیکی و طراحی داروها بر رابطه‌ی میان ساختار شیمیایی و خواص اندازه‌گیری شده‌ی ترکیبات از قبیل ویسکوزیته، اسیدیته، حلالیت، سمیت، اتصال آنزیمی و نفوذ غشایی متمرکز می‌شوند. برای هر سری از ترکیبات وقتی این خواص به فعالیت بیولوژیکی مرتبط باشند این روابط پیچیده می‌گردد [۲۰]. یکی از کاربردهای ویژه روش‌های کمومتریکس توسعه ارتباط کمی ساختار- فعالیت و ارزیابی داده‌های تجزیه‌ای است. مولکول‌های مشابه با تغییر کوچکی در ساختارشان می‌توانند فعالیت‌های شیمیایی کاملاً متفاوتی داشته باشند. این نوع ارتباط بین ساختار مولکولی و تغییر در فعالیت شیمیایی، در مرکز توجه مطالعات QSAR می‌باشد [۲۱]. در واقع QSAR سعی در پیدا کردن رابطه‌ی هماهنگ میان فعالیت شیمیایی و ویژگی‌های مولکولی، به منظور کاربرد این قواعد برای ارزیابی فعالیت ترکیبات جدید دارد [۲۲].

نظریه QSAR بر این اساس است که فعالیت یک ترکیب به ساختار آن که با خصوصیات الکترونی، آبگریزی و فضایی آن توصیف می‌شود، بستگی دارد. در شیمی دارویی اکثر خصوصیات یک دارو به برهم‌کنش با یک پذیرنده‌ی مشخص و شدت و میزان این برهم‌کنش به میل ترکیبی پذیرنده وابسته است، بنابراین تنها توصیف‌کننده‌های مولکولی که امکان اندازه‌گیری کمی برهم‌کنش را فراهم می‌کنند، در مطالعات QSAR قابل استفاده هستند.

هانش<sup>۱</sup> برای اولین بار در سال ۱۹۶۴ مفهوم رابطه‌ی کمی ساختار- فعالیت را مطرح کرد. او نشان داد که توصیف‌کننده‌های کمی<sup>۲</sup> مولکولی در ترکیبات دارویی با ویژگی‌های بیولوژیک این ترکیبات ارتباط تنگاتنگی دارند. او و دانشمندان پس از وی، از توصیف‌کننده‌های ساده‌ای برای توصیف ساختار ترکیبات استفاده کردند که توانایی توصیف ویژگی‌های الکترونی، هندسی، فضایی و آب‌گریزی ترکیبات دارویی را داشتند. هانش و فوجیتا<sup>۳</sup> از روش‌های رگرسیون خطی چندگانه در مطالعات QSAR استفاده کردند [۲۳].

برای انجام مطالعات QSPR و QSAR ابتدا باید مقادیر خاصیت و فعالیت شیمیایی بدست آیند و توصیف‌کننده‌های مناسب از طریق تجربی یا محاسبه به کمک کامپیوتر تولید شوند. مرحله‌ی بعد دستیابی به ارتباط ریاضی بین ساختار و خاصیت یا فعالیت است. به تکنیک‌هایی که برای مدل‌سازی و پیش‌بینی فعالیت یا خاصیت مولکولی مورد استفاده قرار می‌گیرند، روش‌های پارامتری<sup>۴</sup> می‌گویند. معمول‌ترین این تکنیک‌ها رگرسیون خطی چندگانه، آنالیز اجزای اصلی و حداقل مربعات جزئی می‌باشند. معمولاً انتخاب نوع روش محاسبه تحت تأثیر نوع توصیف‌کننده‌های مورد استفاده بوده و تکنیکی که معمولاً مورد استفاده قرار می‌گیرد، روش رگرسیون خطی چندگانه است. با این حال این روش تنها زمانی می‌تواند بکار رود که تعداد مولکول‌های مورد بررسی بیشتر از تعداد توصیف‌کننده‌ها بوده و متغیرها با هم همبستگی زیادی نداشته باشند.

برای برقراری رابطه بین ساختار و خاصیت ترکیبات شش مرحله دنبال می‌شود. این شش مرحله

عبارتند از:

۱- جمع‌آوری سری داده‌ها

۲- ترسیم ساختار هندسی ترکیبات

۳- محاسبه‌ی توصیف‌کننده‌ها

---

<sup>۱</sup>- Hansch

<sup>۲</sup>- Quantitative descriptors

<sup>۳</sup>- Fujita

<sup>۴</sup>- Parametric methods

۴- انتخاب بهترین توصیف‌کننده‌ها با روش‌های انتخاب متغیر برای ورود به مدل

۵- مدل‌سازی

۶- اعتبارسنجی مدل

### ۲-۱-۳- رگرسیون خطی چندگانه

برای اولین بار، فرانسیس گالتون<sup>۱</sup> ایده‌ی رگرسیون یا تنازل را مطرح و از آن در حل یک مسأله‌ی وراثتی استفاده کرد [۲۴]. از رگرسیون برای پیش‌بینی میزان رابطه‌ی یک متغیر مجهول (متغیر وابسته) با یک متغیر معلوم (متغیر مستقل) استفاده می‌شود.

متغیر مستقل متغیری است که مقدار آن توسط محقق مشخص می‌شود و ماهیت آن طوری است که می‌توان آن را به‌طور مستقل تغییر داد. از طرفی مقدار متغیر وابسته با آهنگ متغیر مستقل تغییر می‌کند به طوری که به ازای هر مقدار از متغیر مستقل، یک یا چند مقدار برای متغیر وابسته وجود دارد.

معادله‌ی این نوع رگرسیون به این صورت است:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e_i \quad (1-2)$$

که  $X_i$  ها متغیرهای مستقل (توصیف‌کننده‌های عددی) و مقادیر  $b_i$  ضرایب متغیرهای مستقل‌اند و  $e_i$  مقدار خطای رگرسیون است که باقیمانده نامیده می‌شود. برای بدست آوردن معادله‌ی رگرسیون خطی چندگانه لازم است مجموع مربعات انحراف‌های مقادیر پیش‌بینی شده برای متغیر وابسته از مقادیر واقعی آن به کمترین مقدار ممکن برسد. به عبارت دیگر باید مقدار  $\sum e_i^2$  در این معادله به کمترین حد ممکن کاهش یابد. مقدار  $\sum e_i^2$  طبق رابطه‌ی (۲-۲) تعریف می‌شود:

$$\sum e_i^2 = \sum (y_i - \hat{y})^2 \quad (2-2)$$

---

<sup>۱</sup> - Francis Galton

که در آن  $y_i$  مقدار مشاهده شده‌ی متغیر وابسته و  $\hat{y}$  مقدار محاسبه‌شده‌ی آن به وسیله‌ی معادله‌ی رگرسیون است. در این جا نیز همانند رگرسیون خطی ساده برای محاسبه ضرایب، از روش حداقل مجموع مربعات استفاده می‌شود. بی تردید روش حداقل مربعات<sup>۱</sup> معمول‌ترین ابزار برای بدست آوردن معادله‌ی خطی بین متغیرهای مستقل و وابسته در کمومتریکس است. اصولاً برای رسم یک خط دانستن مقادیر شیب و عرض از مبدا در معادله‌ی مربوط به آن کافی است. این مقادیر را می‌توان از دستگاه معادلات نرمال بدست آورد:

$$\begin{cases} \sum y_i = n\alpha + \beta \sum x_i \\ \sum x_i y_i = \alpha \sum x_i + \beta \sum x_i^2 \end{cases} \quad (3-2)$$

که  $n$  بیان‌گر تعداد مقادیر مشخص تجربی،  $\alpha$  شیب و  $\beta$  عرض از مبدا می‌باشد. با جایگزین کردن مقادیر تجربی متغیرهای مستقل و وابسته در معادلات مقادیر شیب و عرض از مبدا بدست می‌آیند.

خطی که با این روش بدست می‌آید دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است، هر چند همه‌ی مشاهدات تجربی بر روی این خط واقع نمی‌شوند. ویژگی چنین خطی این است که مجموع مربعات انحراف عمودی نقاط مشاهده شده ( $y_i$ ) از این خط، همواره کمتر از مجموع مربعات انحراف عمودی نقاط مشاهده شده ( $y_i$ ) از هر خط مفروض دیگر است.

پارامترهای آماری که در ادامه بیان شده‌اند را می‌توان از مدل حاصل از رگرسیون خطی چندگانه بدست آورد که دقت و صحت مدل را بیان می‌کنند.

آماره  $R$ : ضریب همبستگی بین مقدار تجربی و پیش‌بینی شده از متغیرهای مستقل را نشان می‌دهد. مقدار این آماره بین  $0-1$  متغیر است. مقدار بزرگ‌تر آن نشان دهنده‌ی این مسئله است که ارتباط خطی بیشتری میان متغیر وابسته و متغیرهای مستقل وجود دارد.

آماره  $R^2$ : مبین واریانس مشترک بین متغیرهای وابسته و مستقل است.  $R^2$  فصل مشترک دو زیر مجموعه را به صورت مقداری بیان می‌کند و به آن ضریب تعیین نیز می‌گویند، چون نسبتی از

---

<sup>۱</sup>- Least squares method

واریانس متغیر وابسته را که توسط متغیرهای مستقل به حساب آمده معین می‌کند [۲۶] که به این صورت محاسبه می‌شود:

$$R^2 = \frac{SSR}{SSR + SSE} \quad (۴-۲)$$

$$SSE = \sum (y_i - \hat{y})^2$$

$$SSR = \sum (\hat{y} - \bar{y})^2$$

که SSR بیان‌گر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است و SSE نشان‌گر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است. هر چه مقدار این شاخص بزرگ‌تر باشد، مدل صحیح‌تر خواهد بود.

آماره F: آزمون F یا آزمون فیشر، آزمون معنادار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است که مشابه آزمون‌های معنادار بودن آماری در تحلیل واریانس می‌باشد [۲۵] و از رابطه‌ی (۲-۵) بدست می‌آید:

$$F = \frac{(n-2)SSR}{SSE} \quad (۵-۲)$$

که n تعداد نمونه‌ها می‌باشد.

آماره t: آزمون معناداری هر متغیر وابسته در معادله‌ی رگرسیون می‌باشد. مقدار t از تقسیم مقدار ضریب هر متغیر مستقل بر خطای استاندارد آن بدست می‌آید. این مقدار برای کلیه متغیرهای موجود در معادله‌ی رگرسیون و خارج از آن محاسبه می‌شود [۲۶].

ضریب همبستگی پیرسون<sup>۱</sup>: معیاری از ارتباط خطی بین متغیرها می‌باشد. مقادیر ضریب همبستگی بین -۱ تا +۱ متغیر هستند که علامت آن جهت ارتباط خطی (مستقیم (+) یا عکس (-)) و قدر مطلق آن بزرگی این همبستگی را نشان می‌دهد [۲۷].

<sup>۱</sup> - Pearson correlation

## ۲-۱-۴- انتخاب سری داده‌ها

در مدل‌سازی یافتن مجموعه‌ای از مولکول‌ها که مقادیر تجربی خاصیت مورد بررسی برای آنها در شرایط یکسانی بدست آمده باشد، در اولویت قرار می‌گیرد. زمانی مدل‌سازی‌ها می‌تواند داده‌های مورد نظر را به خوبی پیش‌بینی کند که مولکول‌های مورد استفاده تقریباً مشابه یا گروهی از مشتقات یک ترکیب باشند. در یکی از روش‌های ارزیابی مدل‌های بدست آمده سری داده‌ها به سه مجموعه مجزا به صورت: سری آموزشی، سری آزمون و سری ارزیابی<sup>۱</sup> یا پیش‌بینی تقسیم می‌شوند. این دسته‌بندی به صورت تصادفی انجام می‌شود تا مولکول‌های مشابه به یکدیگر در یک مجموعه قرار نگیرند و هر مجموعه نماینده‌ی تمام داده‌ها باشد. مدل بدست آمده از سری آموزشی با سری آزمون بهینه شده و با سری پیش‌بینی، ارزیابی می‌گردد. البته هر چه سری داده‌ها بزرگتر باشد قدرت پیش‌بینی مدل و اعتبار آن بیشتر خواهد شد. اگر تعداد ترکیبات سری داده‌ها کم باشد، برای ارزیابی مدل از روش ارزیابی تقاطعی<sup>۲</sup> استفاده می‌شود.

## ۲-۱-۵- انتخاب توصیف‌کننده‌ها

با توسعه کمومتریکس و علوم رایانه به راحتی می‌توان ویژگی‌های ساختاری یا الکترونی مولکول را به طور کمی به وسیله پارامترهای عددی که به توصیف‌کننده‌های مولکولی مشهور شده‌اند، در زمان بسیار کوتاه بدست آورد. همانطور که گفته شد به مجموعه ابزارها و روش‌هایی که برای یافتن رابطه‌ای میان رفتار فیزیکوشیمیایی مولکول با پارامترهای ساختاری (مشخصات ساختاری) آن بکار می‌رود، روش‌های پارامتری می‌گوییم. در این روش‌ها بین توصیف‌کننده‌های مولکولی و فعالیت مورد نظر ارتباط منطقی برقرار می‌کنیم، که این ارتباط منطقی منجر به ارائه مدل پیش‌بینی‌کننده می‌شود. توصیف‌کننده‌های وارد شده در این مدل باید دارای پیچیدگی زیادی نبوده و از یکدیگر

---

<sup>۱</sup>- Validation set

<sup>۲</sup>- Cross-validation



مستقل باشند و از طرفی قادر به تفسیر ساختار ترکیبات باشند. گروه‌های مختلفی از توصیف‌کننده‌ها تاکنون معرفی شده‌اند که هر کدام جنبه‌های خاصی از ترکیبات را بررسی و توصیف می‌کنند. توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به دسته‌های متفاوتی تقسیم می‌شوند که تعدادی از مهم‌ترین آنها در اینجا آورده شده است.

### ۲-۱-۵-۱- توصیف‌کننده‌های توپولوژیکی<sup>۱</sup>

این توصیف‌کننده‌ها جزء ساده‌ترین نوع توصیف‌کننده‌ها می‌باشند و به ساختار فضایی مولکول ارتباطی نداشته و تنها به نوع اتم، نوع پیوندها و نحوه ارتباط اتم‌ها به یکدیگر وابسته است. بنابراین این پارامترها را بدون بهینه کردن ساختار مولکول می‌توان محاسبه نمود. تعداد اتم‌ها، تعداد اتم‌های هالوژن، تعداد پیوندهای ساده یا دوگانه و اندیس‌های ارتباطی مثال‌هایی از این نوع توصیف‌کننده‌ها هستند. توصیف‌کننده‌های توپولوژیکی به چهار دسته فرعی تقسیم می‌شوند که عبارتند از: اندیس‌های توپولوژی<sup>۲</sup>، توصیف‌کننده‌های جزئی<sup>۳</sup>، توصیف‌کننده‌های زیرساختاری<sup>۴</sup> و توصیف‌کننده‌های محیطی<sup>۵</sup> [۲۸-۳۱].

### ۲-۱-۵-۲- توصیف‌کننده‌های هندسی<sup>۶</sup>

توصیف‌کننده‌های هندسی نسبت به نوع توپولوژیک از اهمیت بیشتری برخوردارند. یکی از خصوصیات بارز این توصیف‌کننده‌ها، ارتباط آنها با ساختار سه بعدی مولکول‌ها می‌باشد. بنابراین قبل از محاسبه‌ی این توصیف‌کننده‌ها باید پایدارترین ساختار فضایی مولکول تعیین شود که بهینه کردن

---

<sup>۱</sup>- Topological descriptors

<sup>۲</sup>- Topological indices

<sup>۳</sup>- Partial descriptors

<sup>۴</sup>- Substructure descriptors

<sup>۵</sup>- Environmental descriptors

<sup>۶</sup>- Geometric descriptors

ساختار مولکول توسط روش‌های مکانیک کوانتومی انجام می‌شود. حجم و اندروالس، مساحت مولکول و درجه‌ی فاصله هندسی متوسط از این نوع توصیف‌کننده‌ها می‌باشند.

### ۲-۱-۵-۳- توصیف‌کننده‌های الکترونی<sup>۱</sup>

توصیف‌کننده‌های الکترونی بیان‌گر خواص الکترونی مولکول‌ها می‌باشند و اطلاعات مفیدی راجع به برهم‌کنش‌های الکترونی بین مولکولی و درون مولکولی ارائه می‌دهند. این نوع برهم‌کنش‌ها خواص فیزیکی و شیمیایی مولکول‌ها را شدیداً تحت تاثیر قرار می‌دهند. اهمیت این برهم‌کنش‌ها بویژه هنگامی که مولکول‌های عمل‌کننده قطبی باشند، بیشتر می‌شود. با توجه به اینکه مولکول‌ها از تعدادی الکترون، پروتون و نوترون تشکیل شده‌اند، بنابراین بسیاری از رفتارهای مولکول را می‌توان با توجه به تاثیرات متقابل الکتریکی درک و توجیه نمود. اندازه‌ی نیروهای حاصل از برهم‌کنش‌های الکترونی، تابعی از بار جزئی اتم‌ها، نحوه‌ی توزیع بار در مولکول، فاصله بین بارهای مخالف، گشتاور دوقطبی و قطبش‌پذیری مولکول‌ها می‌باشند. هر نوع توصیف‌کننده‌ای که بتواند به نحوی یکی از عوامل موثر در برهم‌کنش‌های الکترونی را بیان کند، در گروه توصیف‌کننده‌های الکترونی قرار می‌گیرد.

### ۲-۱-۵-۴- توصیف‌کننده‌های فیزیکوشیمیایی<sup>۲</sup>

این توصیف‌کننده‌ها از جمله مهم‌ترین توصیف‌کننده‌ها بوده و با عوامل موثر در فعل و انفعالات زیستی، همبستگی بالایی نشان می‌دهند. این دسته از توصیف‌کننده‌ها بیان‌گر بعضی از خواص فیزیکی و شیمیایی مولکول‌ها می‌باشند که به ساختار مولکول بستگی دارند. میزان حلالیت ترکیبات در آب، لگاریتم ضریب تقسیم آب-اکتانول و ویسکوزیته از این سری توصیف‌کننده‌ها می‌باشند.

---

<sup>۱</sup>- Electronic descriptors

<sup>۲</sup>- Physico-chemical descriptors

## ۲-۱-۶- تجزیه و تحلیل آماری توصیف‌کننده‌ها

تعداد توصیف‌کننده‌هایی که توسط نرم افزار محاسبه می‌شود بسیار زیاد است و تمام آنها برای توصیف همه‌ی مولکول‌ها مناسب نبوده و حتی ویژگی یا خاصیتی از مولکول‌ها را بیان نمی‌کنند، به همین دلیل توصیف‌کننده‌هایی که با متغیر وابسته همبستگی بیشتری دارند از بین سایر آنها انتخاب می‌شوند. بنابراین برای کاهش و حذف متغیرهای مستقلی که در مدل‌سازی تاثیر مثبتی ندارند موارد زیر را اعمال می‌کنیم:

- ❖ توصیف‌کننده‌هایی که بیش از ۹۰٪ مقادیر آنها یکسان باشد حذف می‌شوند.
- ❖ از بین توصیف‌کننده‌هایی که با یکدیگر همبستگی بالای ۰/۹ دارند توصیف‌کننده‌ای که با متغیر وابسته همبستگی کمتری دارد، حذف می‌شود.
- ❖ توصیف‌کننده‌هایی که محاسبه و تفسیر آنها مشکل است حذف می‌شوند.

## ۲-۱-۷- آنالیز مدل‌های آماری و انتخاب مدل مناسب

یک مدل بر اساس روش‌های آماری که از قواعد ریاضی بهره می‌برد تشکیل می‌شود و برای وارد کردن متغیرهای آن از ابزاری باید استفاده شود که با انتخاب متغیر<sup>۱</sup> مناسب رابطه‌ی بین متغیر وابسته و مستقل را به خوبی بیان کند. تعدادی از تکنیک‌های انتخاب متغیر عبارتند از: رگرسیون خطی چندگانه، الگوریتم ژنتیک<sup>۲</sup>، روش بایزین<sup>۳</sup> و غیره.

برای انتخاب متغیر در روش رگرسیون خطی چندگانه چندین روش رگرسیون مختلف وجود دارد که عبارتند از:

---

<sup>۱</sup>- Variable selection (VS)

<sup>۲</sup>- Genetic algorithm

<sup>۳</sup>- Bayesian

## ۲-۱-۷-۱- روش ورود اجباری<sup>۱</sup>:

روش ورود اجباری روشی برای انتخاب متغیرهاست که طی آن تمام توصیف‌کننده‌ها در یک مرحله وارد مدل می‌شوند. از آنجا که در این روش اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می‌شوند روش مطلوبی نیست.

## ۲-۱-۷-۲- روش انتخاب جلوبرنده<sup>۲</sup>:

تکنیکی مرحله‌ای برای انتخاب متغیرهاست که طی آن متغیرها یکی پس از دیگری وارد مدل می‌شوند. اولین متغیر که به منظور ورود به مدل انتخاب می‌شود دارای بالاترین همبستگی مثبت یا منفی با متغیر وابسته است. این روند تا زمانی که دیگر متغیری وجود نداشته باشد که دارای شرط ورود باشد، ادامه دارد.

## ۲-۱-۷-۳- روش عقب برنده<sup>۳</sup>:

در این تکنیک تمام توصیف‌کننده‌ها وارد معادله شده و سپس متغیر با کوچکترین همبستگی جزئی با متغیر وابسته حذف می‌شود. پس از اینکه اولین متغیر حذف شد، از بین متغیرهای باقیمانده در معادله متغیری که کوچکترین همبستگی را با متغیر وابسته دارد برای حذف در نظر گرفته می‌شود. این فرآیند زمانی که دیگر متغیر پیرو شرط فرآیند موجود نباشد، متوقف خواهد شد.

## ۲-۱-۷-۴- روش مرحله‌ای<sup>۴</sup>:

در این روش در هر مرحله برای متغیرهای وارد شده در مدل آماره‌ی F محاسبه می‌شود و متغیر با کوچکترین آماره‌ی F بدون معنا حذف شده و مدل مناسب بدست می‌آید.

---

<sup>۱</sup>- Enter

<sup>۲</sup>- Forward selection

<sup>۳</sup>- Backward

<sup>۴</sup>- Stepwise

## ۲-۱-۷-۵- روش حذفی<sup>۱</sup>:

در این تکنیک در هر مرحله توصیف‌کننده‌ها به صورت بلوکی حذف می‌شوند [۳۲].

## ۲-۱-۸- نرم افزارهای مورد نیاز

دانش کمومتریکس بسته‌های نرم افزاری متنوعی را برای انجام تمام مراحل مدل‌سازی بکار گرفته است. در ادامه به اختصار بسته‌های نرم افزاری استفاده شده در این پروژه معرفی می‌شوند.

### ۲-۱-۸-۱- بسته‌ی نرم افزاری Hyperchem

یکی از معروف‌ترین نرم افزارهای شیمی در زمینه‌ی محاسبات کوانتومی و دیگر خواص فیزیکی مولکول‌های شیمیایی است. این بسته، یک نرم افزار مدل‌سازی و شبیه‌سازی است که به ما اجازه می‌دهد محاسبات پیچیده‌ی شیمیایی را با رایانه انجام دهیم. ابتدا شکل مولکول به‌طور تقریبی بر روی صفحه نمایشگر رایانه رسم می‌گردد، سپس توسط روش‌های مکانیک کوانتومی، ساختار مولکول بهینه می‌شود. در این پروژه از روش‌های نیمه‌تجربی<sup>۲</sup> استفاده شد. روش‌های نیمه‌تجربی مختلفی مانند صرف‌نظر کردن کامل از انتگرال‌های همپوشانی جزئی<sup>۳</sup> (CNDO)، صرف‌نظر کردن از برخی انتگرال‌های همپوشانی در الکترون‌ها<sup>۴</sup> (INDO)، صرف‌نظر کردن دو اتمی از انتگرال همپوشانی<sup>۵</sup> (MNDO) ارائه شده است [۳۳]. روش AM1 همان روش MNDO اصلاح شده می‌باشد که روی مولکول‌های قطبی نتایج خوبی می‌دهد ولی پیوند هیدروژنی را نمی‌تواند در نظر بگیرد. در این روش هسته و لایه‌های داخلی را به شکل یک هسته‌ی مرکزی در نظر می‌گیریم و محاسبات روی الکترون‌های ظرفیت انجام می‌گردد. از این نرم افزار می‌توان اطلاعات فراوانی نظیر زوایای پیوندی،

---

<sup>۱</sup>- Remove

<sup>۲</sup>- Semi-empirical methods

<sup>۳</sup>- Complete neglected of differential overlap

<sup>۴</sup>- Intermediate neglected of differential overlap

<sup>۵</sup>- Modified neglected of diatomic overlap

طول پیوندها و زوایای پیچشی را بدست آورد. این بسته نرم افزاری، قابلیت ترسیم ساختار ترکیبات پیچیده را دارد و در کتابخانه‌ی آن ساختار نوکلئیک اسیدها و اسیدهای آمینه موجود است [۳۴].

## ۲-۱-۸-۲- نرم افزار Dragon

نرم افزاری است که با استفاده از یک فایل ورودی با پسوند (*\*.hin*) مربوط به نرم افزار Hyperchem، ۱۸ سری از توصیف‌کننده‌های مولکولی را محاسبه می‌کند. این بسته در سال ۲۰۰۱ توسط گروه کمومتریکس دانشگاه میلانو طراحی و عرضه شده است [۳۵] و قابلیت محاسبه ۱۴۸۱ توصیف‌کننده برای هر مولکول که کمتر از ۱۵۰ اتم داشته باشد را دارد. تعدادی از توصیف‌کننده‌ها از قبیل MDF<sup>۱</sup>، مجموع درجه‌ی والانس برای ترکیبات پلیمری<sup>۲</sup>، مقدار ویژه الکترونی<sup>۳</sup> و... نیز به صورت تجربی بدست آمده‌اند.

## ۲-۱-۸-۳- بسته‌ی نرم افزاری SPSS

این بسته‌ی نرم افزاری، به منظور تجزیه و تحلیل داده‌ها تدوین شده است که در عین سهولت و سادگی، یکی از قدرتمندترین ابزارها برای تحلیل آماری داده‌های اجتماعی و رفتاری بشمار می‌آید. یادگیری آسان، سادگی استفاده و قدرت بالا در انجام محاسبات پیچیده، SPSS را به یکی از رایج‌ترین بسته‌های نرم افزاری مورد استفاده در تجزیه و تحلیل آماری مبدل نموده است. در این پروژه از این نرم افزار، برای بدست آوردن بهترین مدل خطی، حذف متغیرهای با همبستگی بالا و ارزیابی و اعتبار مدل استفاده شده است [۳۲].

---

<sup>۱</sup>- Molecular Descriptors Family

<sup>۲</sup>- Sum of valence degrees (SVDe)

<sup>۳</sup>- Electronic Eigenvalue (EEVA)

## ۲-۱-۸-۴- بسته‌ی نرم افزاری MATLAB R2008a

نرم افزار MATLAB<sup>۱</sup> برای کسانی که با محاسبات عددی و بویژه جبرخطی سر و کار دارند تهیه شده است. هدف اولیه‌ی این نرم افزار قادر ساختن مهندسين و دانشمندان به حل مسائل شامل عملیات ماتریسی بدون نیاز به نوشتن برنامه در زبان‌های برنامه نویسی متداول همچون فورترن<sup>۲</sup> و سی<sup>۳</sup> بود. با گذشت زمان قابلیت‌های بسیار بیشتری به این نرم افزار افزوده شده است، به طوری که در حال حاضر Matlab به ابزاری پر قدرت برای ترسیم داده‌ها، برنامه نویسی و انجام محاسبات مهندسی و پژوهشی تبدیل شده است. Matlab شامل یک سری از برنامه‌های کاربردی افزوده شده برای راه حل‌های مخصوص است که جعبه ابزار<sup>۴</sup> نامیده می‌شود که به کاربران اجازه یادگیری و بکار بردن تکنولوژی‌های خاص را می‌دهد و محیط آن را برای حل مسائل ویژه توسعه می‌دهند. کاربرد جعبه ابزارها شامل پردازش سیگنال‌ها، کنترل سیستم، شبکه‌های عصبی، الگوریتم ژنتیک، منطق فازی، حرکت امواج و مسائل متنوع دیگر می‌شود [۳۶].

---

<sup>۱</sup>- MATrix LABoratory

<sup>۲</sup>- FORTRAN

<sup>۳</sup>- C

<sup>۴</sup>- Toolbox

فصل سوم

الگوریتم ژنتیک



الگوریتم‌های کلاسیک متنوع و فراوانی برای جستجوی فضای جواب‌ها و یافتن جواب بهینه طراحی و مورد استفاده قرار گرفته‌اند. برای حل یک مساله بهینه‌سازی خاص، انتخاب‌های متنوع و متفاوت زیادی در روش حل آن وجود دارد چرا که هیچ یک از روش‌های کلاسیک و سنتی موجود برای حل مسائل بهینه‌سازی، نسبت به نوع مسأله مقاوم نمی‌باشند. یعنی اینکه هر کدام از این روش‌ها برای یک نوع از مسائل خاص جواب خوبی پیدا می‌کنند و در سایر مسائل دچار مشکل می‌باشند و وقتی یک مسأله جدید مطرح می‌شود باید یک راه حل جدید برای آن ایجاد کنیم. مشکل دیگر روش‌های کلاسیک و سنتی این است که فقط در مورد مسائلی که کاملاً تعریف شده و با یک عبارت ریاضی توصیف شده‌اند می‌توانند بکار روند و بنابراین در تعداد زیادی از مسائل دنیای حقیقی که تعریف کامل مسأله و معادلات ریاضی آن را در دست نداریم نمی‌توان این روش‌ها را بکار برد.

اکثر روش‌های سنتی موجود، یافتن جواب دقیق را تضمین می‌کنند و از اینرو روش‌های دقیق<sup>۱</sup> نامیده می‌شوند اما هنگام مواجه شدن با مسائل دنیای حقیقی، استفاده از روش‌های تقریبی<sup>۲</sup> ارجحیت دارد. آن دسته از روش‌های سنتی که از نوع روش‌های تقریبی می‌باشند، از تکنیک جستجوی محلی استفاده می‌کنند که از آن جمله می‌توان روش‌های گرادیانی<sup>۳</sup> را نام برد. یک خطر عمده‌ی تخمین زدن، دچار شدن در نقطه بهینه محلی<sup>۴</sup> است. در واقع در تکنیک جستجوی محلی یک راه حل موجود، در تکرارهای متوالی با راه حل تغییر یافته خود، که یک راه حل جدید در همسایگی راه حل موجود می‌باشد مقایسه می‌شود و اگر راه حل تغییر یافته بهتر از راه حل موجود باشد، جایگزین آن می‌شود. اگر بازه‌ی همسایگی را بزرگ انتخاب کنیم امکان فرار از بهینه محلی بیشتر می‌شود ولی این کار باعث کند شدن الگوریتم می‌شود که غیرقابل قبول خواهد بود.

---

<sup>۱</sup>- Exact methods

<sup>۲</sup>- Approximate methods

<sup>۳</sup>- Gradient

<sup>۴</sup>- Local optimization

برای غلبه بر مشکلات روش‌های سنتی، روش‌های تجربی مدرن<sup>۱</sup> بوجود آمده‌اند که روش‌های تقریبی می‌باشند و تمرکز عمده آنها در فرار از نقطه بهینه محلی و یافتن بهینه کلی<sup>۲</sup> می‌باشد. مزیت این روش‌ها این است که این الگوریتم‌ها تنها برای یک نوع خاص مسائل جواب خوبی نمی‌دهند بلکه الگوریتم‌های همه منظوره می‌باشند. روش‌های مدرن به دو گروه کلی روش‌های مبتنی بر جمعیت<sup>۳</sup> و روش‌های مبتنی بر یک جواب واحد<sup>۴</sup> تقسیم می‌شوند. در روش‌های مبتنی بر جمعیت، یک جمعیت که هر یک از اعضای آن یک جواب بالقوه مساله می‌باشد، با همکاری یکدیگر سعی در یافتن نقطه‌ی بهینه دارند در حالی که در روش‌های مبتنی بر یک جواب واحد، الگوریتم از یک نقطه شروع می‌کند و این نقطه با مقایسه و در صورت لزوم جایگزین شدن با نقاط تغییر یافته خود، خود را به سمت نقطه بهینه می‌برد.

### ۳-۲- الگوریتم ژنتیک

علی‌رغم نام آن، باید الگوریتم ژنتیک را جزء روش‌های هوش مصنوعی غیرالگوریتمی به حساب آورد. برخلاف روش‌های معمول آماری که دارای الگوریتم مشخص‌اند و در یک مسأله خاص همواره به یک جواب مشخص منتهی می‌شوند، الگوریتم ژنتیک همواره به یک جواب نمی‌رسد. الگوریتم ژنتیک را می‌توان روشی برای جست و جو در میان راه حل‌های مختلف یک مسأله دانست. بطور کلی روش‌های جست و جو به سه دسته روش‌های مبتنی بر حساب دیفرانسیل<sup>۵</sup>، روش‌های شمارشی<sup>۶</sup> و روش‌های تصادفی<sup>۷</sup> تقسیم بندی می‌شوند. روش الگوریتم ژنتیکی از روش‌های تصادفی به حساب می‌آید [۳۷].

---

<sup>۱</sup>- Modern heuristics methods

<sup>۲</sup>- Global optimization

<sup>۳</sup>- Population based methods

<sup>۴</sup>- Single solution methods

<sup>۵</sup>- Differential calculus

<sup>۶</sup>- Numerical

<sup>۷</sup>- Stochastic

### ۳-۲-۱- الگوی طبیعی الگوریتم ژنتیک

آنچه الگوریتم ژنتیک از آن الگوبرداری شده است، نظریه‌ی تکامل موجودات زنده در طبیعت است که داروین<sup>۱</sup> آن را مطرح کرد. قبل از داروین افراد دیگری همچون لامارک<sup>۲</sup> هم تکامل را مطرح کرده بودند، اما وجه ممیز تئوری داروین انتخاب طبیعی<sup>۳</sup> است که تا قبل از داروین کسی آن را توضیح نداده بود. هرچند، سال‌ها پس از داروین که ماهیت ژن‌ها کشف شد، نظریه تکامل او دست خوش تغییراتی گشت، اما انتخاب طبیعی داروین همواره به صورت یک اصل غیر قابل تغییر در علم زیست‌شناسی باقی ماند.

صرف نظر از انواع موجودات زنده، خواص طبیعی آنها از دو منشأ ژنتیکی و محیطی تأثیر می‌پذیرد. طرح ژن‌های موجودات زنده از والد یا والدین آنها به ارث می‌رسد. برهم‌کنش این طرح با پارامترهای محیطی تعیین‌کننده خصوصیات جانداران می‌باشد. به‌طورکلی، هر ژن رمز لازم برای ساخت یک پروتئین را فراهم می‌کند. هر پروتئین نیز موجب ایجاد یک ویژگی خاص ساختاری یا عملکردی در جانداران می‌گردد. با در نظر گرفتن ماهیت وراثت در موجودات زنده، می‌توان ادعا کرد که محیط زیست هر موجود زنده تعیین‌کننده مفید یا مضر بودن هر کدام از ژن‌های آن است. هر ژنی که موجب زیاد شدن سازگاری<sup>۴</sup> جاندار با محیط اطراف شود مفید و هر ژنی که این سازگاری را کاهش دهد، مضر است. مفهوم انتخاب طبیعی این است که محیط باعث می‌شود ژن‌های مفید در یک جمعیت از موجودات حفظ شده و ژن‌های مضر به مرور زمان حذف گردد. پس از کشف ماهیت مولکولی ژن‌های جانداران، نحوه انجام انتخاب طبیعی به تعدادی از عملگرهای ژنتیکی نسبت داده شد. مهم‌ترین این عملگرها عبارتند از جهش<sup>۵</sup> و پیوند<sup>۶</sup> که هر کدام انواع مختلف دارند.

---

<sup>۱</sup>- Darwin

<sup>۲</sup>- Lamarck

<sup>۳</sup>- Natural selection

<sup>۴</sup>- Fitness

<sup>۵</sup>- Mutation

<sup>۶</sup>- Crossover

کروموزومها<sup>۱</sup> آرایه‌های بزرگی از ژن‌ها هستند که آنها را در طول نسل‌های جانداران حمل می‌کنند. اکثر جانداران پیشرفته (که به آنها دیپلوئید<sup>۲</sup> گفته می‌شود) دارای یک جفت از هر کروموزوم هستند که هر کدام از یک والد به ارث رسیده است. این کروموزوم‌ها را همولوگ<sup>۳</sup> می‌نامند. دو ژن متناظری که بر روی دو کروموزوم همولوگ قرار دارند، به صورت بالقوه توانایی ایجاد دو حالت مختلف یا یکسان برای یکی از صفات موجود را دارند. مثلاً یکی می‌تواند رنگ موی روشن و دیگری رنگ موی تیره را ایجاد کند. این حالات مختلف آلل‌های<sup>۴</sup> ژن نام دارند. بسته به نوع ژن، صفت نهایی ایجاد شده در فرد، برآیندی از اثر دو آلل موجود بر روی دو ژن متناظر آن است. البته در شکل‌گیری بعضی صفات، محیط نیز اثرگذار است. جهش به صورت تصادفی باعث تغییر، افزایش و یا حذف در قطعه‌ای از ماده ژنتیکی (DNA) موجود زنده می‌شود. این تغییر ساختاری موجب تغییراتی در عملکرد ژن موجود در آن قطعه می‌گردد. جهش باعث تولید آلل‌های جدید در یک جمعیت از موجودات زنده می‌گردد. اگر جهش در سلول‌های زایشی جاندار ایجاد شود، به نسل بعدی به ارث می‌رسد و اثر خود را نمایان می‌سازد. پیوند باعث جابجایی قطعاتی متناظر و با طول مساوی در دو کروموزوم همولوگ می‌شود. به این ترتیب، ترکیب جدیدی از آلل‌ها در کروموزوم‌ها ایجاد می‌شود که به این حالت نو ترکیبی<sup>۵</sup> گویند. همچنین بعضی آلل‌های جدید هم به همین طریق تولید می‌شوند. در مجموع، پیوند موجب ازدیاد تنوع در ویژگی‌های موجودات زنده می‌گردد. پیوند در سلول‌های زایشی رخ می‌دهد و کروموزوم‌های حاصل از آن به نسل بعد به ارث می‌رسد. شکل (۱-۳) نمونه‌هایی از پیوند را در کروموزوم‌های فرضی نشان می‌دهد.

---

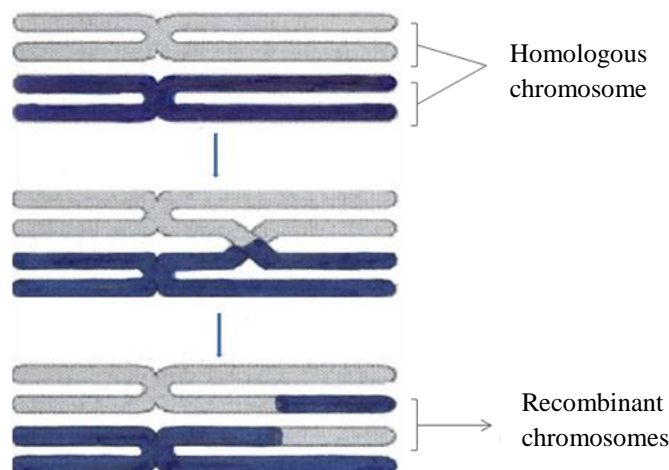
<sup>۱</sup>- Chromosome

<sup>۲</sup>- Diploid

<sup>۳</sup>- Homolog

<sup>۴</sup>- Alleles

<sup>۵</sup>- Recombination



شکل (۳-۱) - پیوند باعث جابجا شدن قطعه‌ای از DNA بین دو کروموزوم و ایجاد کروموزوم‌های نو ترکیب می‌شود.

وقتی یک ژن بوسیله عملگرهای جهش و پیوند تغییر می‌کند، باعث تغییر در سازگاری جاندار نسل بعد با محیط اطراف می‌شود. اگر سازگاری جاندار با محیط بیشتر شود، توانایی آن نیز برای زندگی بهتر و در نهایت تولید نسل افزایش می‌یابد. در این صورت، آلل جدید به نسل‌های بعدی به ارث می‌رسد و تعداد آن در جمعیت افزایش می‌یابد. از طرفی اگر ژن جدید سازگاری جاندار را با محیط اطراف کم کند، توانایی فرد را برای زندگی و تولید مثل کاهش می‌دهد و به این ترتیب ژن جدید کمتر تکثیر می‌شود و به تدریج تعداد آن در جمعیت به حداقل رسیده، و یا گاهی به‌طور کلی حذف می‌گردد. انتخاب ژن‌های سازگارتر با محیط، اساس نظریه انتخاب طبیعی داروین را شکل می‌دهد. می‌توان گفت روند تکامل گونه‌های زنده با رویکرد انتخاب طبیعی، توالی‌های ژنتیکی را در طول نسل‌های متوالی، برای سازگاری با محیط اطراف بهینه می‌کند [۳۸].

### ۳-۲-۲- تاریخچه و اساس الگوریتم ژنتیک

پیشرفت علم ژنتیک در قرن بیستم از نظر سرعت و قدرت عمل تنها با تکامل علمی مانند الکترونیک و کامپیوتر قابل رقابت بود. می‌توان ادعان داشت یکی از زمینه‌سازترین الگوریتم‌های بوجود آمده در قرن بیستم الگوریتم ژنتیک می‌باشد.

در دهه ۱۹۶۰ میلادی، دانشمندان ایده الگوبرداری از تکامل جانداران را در حل مسائل ریاضی مطرح کردند. در اواخر این دهه، جان هلند<sup>۱</sup> توانست این ایده را عملی کرده و در سال ۱۹۷۵ کتابی درباره الگوریتم ژنتیک منتشر کرد [۳۹]. سپس گلدبرگ<sup>۲</sup> در سال ۱۹۸۹ کتاب معروف خود را درباره بهینه‌یابی با الگوریتم‌های ژنتیک به رشته تحریر در آورد [۴۰]. در سال ۱۹۹۲ جان کوزا<sup>۳</sup> الگوریتم ژنتیک را برای تولید برنامه‌هایی برای انجام کارهای خاص بکار برد و روش خود را «برنامه نویسی ژنتیک»<sup>۴</sup> نامید.

کاربرد اساسی الگوریتم ژنتیک، بهینه‌سازی است. بر اساس این کاربرد اساسی می‌توان از این روش در موارد زیر استفاده کرد:

- ۱- برنامه ریزی خودکار
  - ۲- آموزش هوشمند ماشین‌های صنعتی و روش‌های دیگر هوش مصنوعی مانند شبکه‌ی عصبی مصنوعی.
  - ۳- بهینه‌سازی شاخص‌های اقتصادی
  - ۴- تحلیل سیستم ایمنی جانوران شبیه‌سازی شده
  - ۵- تحلیل پارامترهای مؤثر در اکوسیستم‌ها
  - ۶- بهینه‌سازی ساختار مولکول‌های پیچیده از جمله پروتئین‌ها
- هم‌چنین بسیاری کاربردهای دیگر برای الگوریتم ژنتیکی قابل تصور است.

انتخاب طبیعی اساس تفکر الگوریتم ژنتیک را شکل می‌دهد. برای مدل‌سازی روند تکامل و انتخاب طبیعی در الگوریتم ژنتیک، هر کدام از مفاهیم ژنتیکی که به آن اشاره کردیم را در یک مدل پیاده می‌کنیم. در مدل الگوریتم ژنتیک هر کروموزوم حاوی یک راه حل برای مسأله است و هر فرد نیز دارای یک کروموزوم است. هر کروموزوم رشته‌ای از ژن‌هاست و هر ژن جزئی از راه حل مسأله را

---

<sup>۱</sup>- John Holand

<sup>۲</sup>- Goldberg

<sup>۳</sup>- John Koza

<sup>۴</sup>- Genetic programming

بیان می‌کند. ماهیت ژن‌ها بستگی به مسأله مورد بحث دارد. مثلاً در یک مسأله بهینه‌سازی ساختار پروتئین، هر ژن نام یک آمینواسید است. در حل یک معادله، هر ژن یک رقم از جواب و در بهینه‌سازی پارامترهای دخیل در یک مدل دیگر مانند شبکه‌ی عصبی مصنوعی، هر ژن یک متغیر دو حالتی است که نشان دهنده‌ی دخالت یا عدم دخالت یک پارامتر در حل مسأله است. حالت‌های مختلفی که یک ژن می‌تواند به خود بگیرد را آلل‌های آن ژن می‌نامند [۴۱].

### ۳-۲-۳- مبانی الگوریتم ژنتیک

حل هر مسأله با الگوریتم ژنتیک یک فرآیند پویا یا تکاملی را شامل می‌شود. الگوریتم با یک مجموعه از جواب‌ها (که بوسیله‌ی کروموزوم‌ها نشان داده می‌شوند) به نام جمعیت<sup>۱</sup> آغاز می‌شود. جواب‌ها از یک جمعیت گرفته شده و برای ایجاد جمعیت جدیدی بکار می‌روند. هدف این است که جمعیت جدید از جمعیت پیشین بهتر باشد، از اینرو معیار انتخاب<sup>۲</sup> جواب‌هایی که برای ایجاد جواب‌های جدید (فرزندان<sup>۳</sup>) انتخاب می‌شوند شایستگی<sup>۴</sup> آنها خواهد بود و به عبارتی هر چه شایستگی یک جواب بیشتر باشد، شانس بیشتری برای تولید نسل دارد. این امر تا زمانی که شرایط خاصی (به عنوان مثال تعداد جمعیت‌ها یا بهبود بهترین جواب) حکمفرما گردد تکرار می‌شود.

### ۳-۲-۴- مراحل کلی الگوریتم ژنتیک

۱- آغاز: تولید جمعیت اتفاقی  $n$  کروموزومی (جواب‌هایی حدسی برای مسأله).

۲- شایستگی: محاسبه شایستگی  $f(x)$  متناظر با هر کروموزوم  $x$  در جمعیت.

---

<sup>۱</sup>- Population

<sup>۲</sup>- Selection

<sup>۳</sup>- Offspring

<sup>۴</sup>- Fitness

۳- جمعیت جدید: ایجاد یک جمعیت جدید بوسیله تکرار مراحل زیر تا هنگامی که جمعیت جدید کامل گردد:

الف- انتخاب: انتخاب دو کروموزوم والدین از جمعیت با توجه به شایستگی آنها (هر چه شایستگی بالاتر باشد شانس انتخاب بالاتر خواهد بود).

ب- پیوند: پیوند والدین با یک احتمال (نرخ) پیوند<sup>۱</sup> برای تشکیل فرزندان (اگر از روش پیوند استفاده نکنیم فرزندان دقیقاً شبیه والدین خواهند شد).

ج- جهش: تغییر فرزندان جدید با یک احتمال (نرخ) جهش<sup>۲</sup>، از طریق تغییر مقدار ژن‌ها در کروموزوم.

د- پذیرش: قرار دادن فرزندان جدید در یک جمعیت جدید.

۴- جایگزینی: بکار بردن جمعیت تازه ایجاد شده برای سیکل بعدی اجرای الگوریتم.

۵- آزمایش: اگر شرایط انتهایی بدست آید برنامه متوقف شده و بهترین جواب موجود در جمعیت را معرفی می‌کند.

۶- بازگشت: بازگشت به مرحله ۲.

همانطور که می‌دانیم این روند بسیار کُلی است. پارامترها و شرایط بسیاری هستند که می‌توانند به شیوه‌های متفاوتی در مسائل مختلف بکار روند.

اولین سوال این است که چگونه باید کروموزوم‌ها را ساخت و چه نوع رمزبندی<sup>۳</sup> برای این کار انتخاب گردد. سپس باید دو پارامتر اساسی GA یعنی پیوند و جهش را تعریف کرد.

سوال بعدی این است که چگونه والدین را برای پیوند انتخاب کنیم. این امر از راه‌های مختلفی امکان پذیر است، اما ایده اصلی، انتخاب بهترین والدین است به امید اینکه والدین بهتر فرزندان بهتری تولید کنند.

---

<sup>۱</sup>- Crossover probability/rate

<sup>۲</sup>- Mutation probability/rate

<sup>۳</sup>- Encoding



تولید جمعیت تنها به وسیله‌ی دو کروموزوم والدین ممکن است موجب از دست دادن بهترین کروموزوم از جمعیت قبل شود از اینرو در روش‌های انتخاب، اغلب نخبه‌گرایی<sup>۱</sup> محدود مورد استفاده قرار می‌گیرد (که در جای خود تشریح می‌گردد).

### ۳-۳- رمزبندی یک کروموزوم

در آغاز حل یک مساله با GA، رمزبندی کروموزوم‌ها اولین کاری است که باید انجام شود. رمزبندی بستگی کامل به خود مساله دارد. یک کروموزوم بایستی حاوی اطلاعاتی مطابق با جواب حاصل باشد. روشی که بیش از همه برای رمزبندی مورد استفاده قرار گرفته است روش رشته دودویی<sup>۲</sup> (باینری) می‌باشد. در این روش هر کروموزوم به وسیله‌ی یک رشته دودویی نشان داده می‌شود. هر بیت<sup>۳</sup> در رشته (کروموزوم) می‌تواند ارائه دهنده برخی مشخصات جواب باشد. امکان دیگر این است که کل رشته یک عدد را ارائه دهد.

البته راه‌های بسیاری برای رمزبندی وجود دارد ولی روش رمزبندی اساساً به ماهیت مساله بستگی دارد. در این بخش برخی روش‌های رمزبندی که تاکنون با موفقیت نسبی بکار رفته‌اند، معرفی خواهد شد.

### ۳-۳-۱- رمزبندی دودویی (باینری)

رمزبندی دودویی اساس رایج‌ترین نوع رمزبندی است زیرا اولین محققین GA این نوع از رمزبندی را بکار برده‌اند و از طرفی این روش نسبتاً ساده است.

در رمزبندی دودویی، هر کروموزوم یک رشته از بیت‌ها با مقدار صفر یا یک می‌باشد که مثالی از آن در جدول (۳-۱) آورده شده است.

---

<sup>۱</sup>- Elitism

<sup>۲</sup>- Binary string

<sup>۳</sup>- Bite

جدول (۳-۱) - کروموزوم‌هایی با رمزبندی دودویی

Chromosome A	101100101100101011100101
Chromosome B	111111100000110000011111

رمزبندی دودویی، کروموزوم‌های ممکن زیادی را با کمک تعداد محدودی ژن ارائه می‌دهد. اما از طرفی این رمزبندی گاهی برای برخی مسائل طبیعی مناسب نیست و اصلاحاتی در تعریف پیوند و جهش بایستی انجام گیرد.

### ۳-۳-۲- رمزبندی تبدیل<sup>۱</sup>

رمزبندی تبدیل را می‌توان در مسائل ترتیبی مانند مسأله‌ی فروشنده‌ی دوره گرد<sup>۲</sup> یا مسأله‌ی ترتیب کارها<sup>۳</sup> بکار برد. در این رمزبندی، هر کروموزوم یک رشته از اعداد است که یک مکان را در یک توالی ارائه می‌کند. جدول (۳-۲) نمونه‌ای از این رمزبندی را نشان می‌دهد.

جدول (۳-۲) - کروموزوم‌هایی با رمزبندی تبدیل

Chromosome A	1 5 3 2 6 4 7 9 8
Chromosome B	8 5 6 7 2 3 1 4 9

### ۳-۳-۳- رمزبندی مقداری<sup>۴</sup>

رمزبندی مقداری مستقیماً می‌تواند در مسائلی که تعدادی مقادیر پیچیده مانند اعداد حقیقی در آن بکار رفته، مورد استفاده قرار گیرد. در حالی که کاربرد رمزبندی دودویی برای این نوع از مسائل مشکل خواهد بود.

در رمزبندی مقداری، هر کروموزوم یک توالی از تعدادی مقدار می‌باشد که این مقادیر هر موردی مربوط به مسأله، مانند اعداد (حقیقی)، کاراکتر یا هر موضوع دیگر باشد.

<sup>۱</sup>- Permutation encoding

<sup>۲</sup>- Travelling salesman problem

<sup>۳</sup>- Task ordering problem

<sup>۴</sup>- Value encoding

جدول (۳-۳) - کروموزوم‌هایی با رمزبندی مقداری

Chromosome A	1.2324 5.3243 0.4556 2.3293 2.4545
Chromosome B	ABDJEIFJDHDIERJFDLDFLFEGT
Chromosome C	(back), (back), (right), (forward), (left)

رمزبندی مقداری برای برخی مسائل خاص مفید بوده و در این رمزبندی اغلب لازم است که پیوندها و جهش‌های خاصی را برای مسأله تعریف کنیم.

### ۳-۳-۴ - رمزبندی درختی<sup>۱</sup>

در رمزبندی درختی، هر کروموزوم یک درخت از تعدادی موضوع، مانند توابع یا دستورها در زبان برنامه نویسی می‌باشد.

جدول (۴-۳) - کروموزوم‌هایی با رمزبندی درختی

Chromosome A	Chromosome B
( + x ( / 5 y ) )	( do_until step wall )

رمزبندی درختی برای برنامه‌های تکامل تدریجی یا هر ساختار دیگری که بتواند به صورت درختی رمزبندی شود مفید است. از آنجا که برنامه‌های زبان برنامه نویسی LISP<sup>۲</sup> مستقیماً به فرم درختی ارائه شده و می‌توانند به سادگی به صورت درخت تجزیه شوند، اغلب LISP برای این رمزبندی بکار می‌رود. عملیات پیوند و جهش نیز به سادگی قابل انجام هستند.

<sup>۱</sup>- Tree encoding

<sup>۲</sup>- Lots of Irritating Superfluous Parentheses

### ۳-۴- عملگرهای GA

همانگونه که در طرح کلی الگوریتم ژنتیک دیدیم، پیوند و جهش از بخش‌های مهم الگوریتم ژنتیک هستند. عملکرد الگوریتم ژنتیک به‌طور اساسی تحت تاثیر این دو عملگر می‌باشد [۴۲-۴۴].

#### ۳-۴-۱- پیوند

پس از آنکه نوع رمزبندی مشخص شد، می‌توان مرحله‌ی پیوند را اجرا کرد. عملگر پیوند روی ژن‌های منتخب از کروموزوم‌های والدین عمل کرده و فرزندان جدید را تولید می‌کند. ساده‌ترین راه برای انجام این کار انتخاب اتفاقی تعدادی نقاط قطع و کپی کردن تمام چیزهای قبل از این نقطه از والد اول و کپی تمام چیزهای بعد از نقطه قطع از والد دیگر است. روش پیوند را می‌توان در جدول زیر روشن نمود (علامت / نشانگر نقطه‌ی قطع می‌باشد):

جدول (۳-۵) - پیوند یک نقطه‌ای دو کروموزوم

Chromosome 1	11011 / 00100110110
Chromosome 2	11011 / 11000011110
Offspring 1	11011 / 11000011110
Offspring 2	11011 / 00100110110

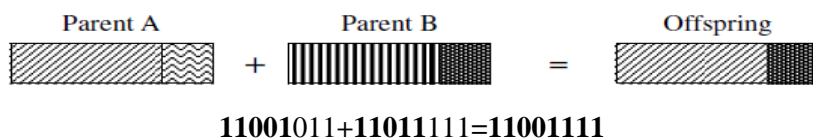
راه‌های دیگری نیز برای پیوند وجود دارد، که می‌توان نقاط قطع بیشتری را انتخاب کرد. پیوند می‌تواند کاملاً پیچیده باشد و این اساساً به روش رمزبندی کروموزوم‌ها بستگی دارد. گاهی اوقات ساختن یک پیوند خاص برای یک مسأله می‌تواند عملکرد الگوریتم ژنتیک را بهبود بخشد. در زیر چند نوع پیوند را برای انواع رمزبندی معرفی می‌کنیم.

---

<sup>۱</sup> - Operator

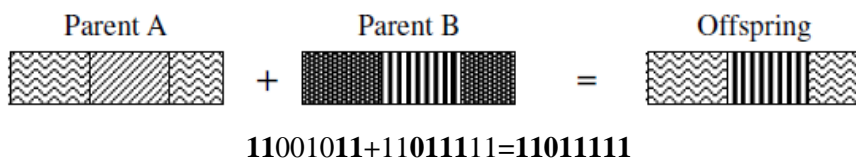
### ۳-۴-۱-۱- پیوند در رمزبندی دودویی

❖ **پیوند یک نقطه‌ای<sup>۱</sup>**: یک نقطه از کروموزوم برای پیوند انتخاب می‌گردد. رشته دودویی از ابتدای کروموزوم تا نقطه پیوند از یکی از والدین کپی می‌شود و بقیه کروموزوم از دیگری کپی می‌گردد.



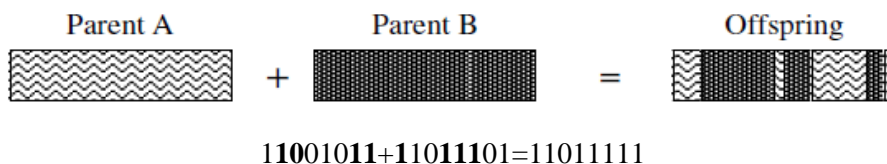
شکل (۲-۳) - پیوند یک نقطه‌ای

❖ **پیوند دو نقطه‌ای<sup>۲</sup>**: در این حالت دو نقطه از کروموزوم انتخاب می‌گردد و رشته دودویی از ابتدای کروموزوم تا نقطه پیوند اول از یکی از والدین، قسمت حد فاصل نقاط پیوند اول و دوم از والد دیگری و بقیه کروموزوم نیز از اولی کپی می‌شود.



شکل (۳-۳) - پیوند دو نقطه‌ای

❖ **پیوند یکنواخت<sup>۳</sup>**: بیت‌ها به صورت اتفاقی از والدین کپی می‌شوند.



شکل (۴-۳) - پیوند یکنواخت

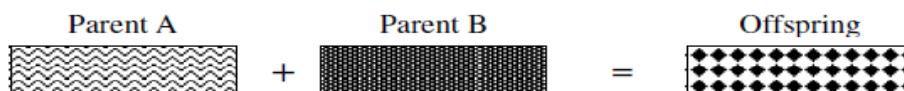
❖ **پیوند عددی<sup>۴</sup>**: چند عمل عددی برای تولید فرزندان جدید بکار می‌رود.

<sup>۱</sup>- Single point crossover

<sup>۲</sup>- Two point crossover

<sup>۳</sup>- Uniform crossover

<sup>۴</sup>- Arithmetic crossover



$$11001011 + 11011111 = 11001001 \text{ (AND)}$$

شکل (۳-۵) - پیوند عددی

### ۳-۴-۱-۲ - پیوند در رمزبندی تبدیل

❖ پیوند یک نقطه‌ای: یک نقطه پیوند انتخاب شده و تبدیل از والد اول تا نقطه پیوند کپی می‌شود. سپس والد دیگر مورد بررسی قرار گرفته و اگر عددی هنوز در فرزند حاصله وجود نداشته باشد، اضافه می‌گردد. البته راه‌های دیگری نیز برای تولید بقیه کروموزوم (از نقطه پیوند تا انتهای کروموزوم) وجود دارد، مانند:

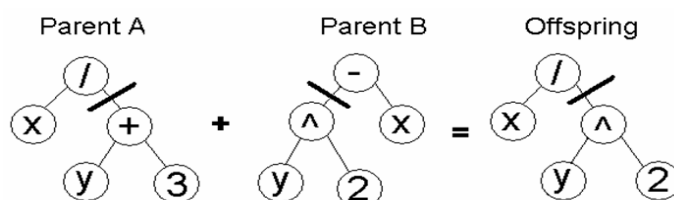
$$(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9) + (4\ 5\ 3\ 6\ 8\ 9\ 7\ 2\ 1) = (1\ 2\ 3\ 4\ 5\ 6\ 8\ 9\ 7)$$

### ۳-۴-۱-۳ - پیوند در رمزبندی مقداری

تمامی پیوندهای مربوط به رمزبندی دودویی را می‌توان استفاده نمود.

### ۳-۴-۱-۴ - پیوند در رمزبندی درختی

❖ پیوند درختی<sup>۱</sup>: یک نقطه پیوند در هر دو کروموزوم والدین انتخاب شده و سپس والدین از همان نقطه تقسیم شده و قسمت‌های زیر آن نقطه جایگزین شده و فرزندان جدید تولید می‌گردد.



شکل (۳-۶) - مثالی از پیوند درختی

<sup>۱</sup> - Tree crossover

### ۳-۴-۲- جهش

پس از انجام عملیات پیوند، عملیات جهش انجام می‌گیرد. جهش با هدف جلوگیری از سقوط تمامی جواب‌ها در جمعیت به یک بهینه‌ی محلی انجام می‌گیرد. عملکرد جهش، فرزندان نتیجه شده از عملیات پیوند را به‌طور اتفاقی تغییر می‌دهد. در رمزبندی دودویی می‌توان تعداد بسیار کمی از بیت‌هایی که به‌طور اتفاقی انتخاب می‌شوند را از یک به صفر و یا از صفر به یک تغییر دهیم. فرآیند جهش در مثال زیر نشان داده شده است:

جدول (۳-۶) - مثالی از جهش در دو کروموزوم فرزند

Original offspring 1	1101111000011110
Original offspring 2	1101100100110110
Mutated offspring 1	1100111000011110
Mutated offspring 2	1101101100110110

تکنیک جهش همانند پیوند اساساً به روش رمزبندی کروموزوم‌ها بستگی دارد. به عنوان مثال هنگامی که تبدیل‌ها را رمز بندی می‌کنیم، جهش می‌تواند به صورت تعویض دو ژن معرفی گردد. در زیر چند نوع جهش را برای انواع رمزبندی معرفی می‌نماییم.

### ۳-۴-۲-۱- جهش در رمزبندی دودویی

وارونه سازی بیت‌ها: با انتخاب تصادفی دو قسمت از کروموزوم‌ها زیر رشته‌ی بین این دو موقعیت وارونه می‌شوند.

### ۳-۴-۲-۲- جهش در رمزبندی تبدیل

تغییر رتبه: دو عدد انتخاب شده و تغییر می‌یابند، مانند:

$$(1\ 2\ 3\ 4\ 5\ 6\ 8\ 9\ 7) = (1\ 8\ 3\ 4\ 5\ 6\ 2\ 9\ 7)$$

### ۳-۴-۲-۳- جهش در رمزبندی مقداری

در این حالت، برای رمزبندی مقادیر حقیقی یک عدد کوچک به مقادیر انتخاب شده اضافه شده یا از آنها کسر می‌گردد.

$$(1.29 \ 5.68 \ 2.73 \ 4.22 \ 5.55) = > (1.29 \ 5.68 \ 2.86 \ 4.11 \ 5.55)$$

### ۳-۴-۲-۴- جهش در رمزبندی درختی

در این شرایط گره‌ها یا شاخه‌های انتخابی تغییر می‌یابند.

### ۳-۵- عملگرهای ژنتیک از دیدگاه نحوه اعمال بر کروموزوم‌ها

عملگرهای ژنتیک از دیدگاه نحوه اعمال بر کروموزوم‌ها به سه نوع زیر دسته بندی می‌شوند:

❖ عملگرهای رسمی<sup>۱</sup>

❖ عملگرهای حسابی<sup>۲</sup>

❖ عملگرهای جهت‌دار<sup>۳</sup>

عملگرهای رسمی برش، خود شامل دو نوع برش ساده<sup>۴</sup> و برش تصادفی<sup>۵</sup> می‌باشند. در برش ساده که نوع آشنای آن برش با یک نقطه پیوند<sup>۶</sup> می‌باشد، فرزندان با جابجا کردن قسمتی از ژن‌های والدین که با نقطه پیوند مشخص شده‌اند با یکدیگر تولید می‌شوند. انواع مختلف دیگری از برش‌های ساده، مانند برش با چند نقطه پیوند<sup>۷</sup>، برش یکنواخت و ... وجود دارند.

---

<sup>۱</sup>- Conventional operators

<sup>۲</sup>- Arithmetical operators

<sup>۳</sup>- Direction-Based operators

<sup>۴</sup>- Simple crossover

<sup>۵</sup>- Random crossover

<sup>۶</sup>- One-Cut point crossover

<sup>۷</sup>- Multi-Cut point crossover



در برش‌های تصادفی که یک نوع اصلی آن برش تخت<sup>۱</sup> نام دارد، هر ژن یک فرزند به صورت تصادفی در بازه‌ی تشکیل شده توسط دو ژن متناظر در والدین انتخاب می‌شود. عملگرهای رسمی جهش نیز انواع مختلفی دارند. آشناترین آن در مورد کدگذاری یا رمزبندی دودویی بکار می‌رود و تغییر حالت ژنی که برای جهش انتخاب شده است از یک به صفر و یا بالعکس صورت می‌گیرد. در مورد رمزبندی غیر دودویی می‌توان به انواع جهش یکنواخت<sup>۲</sup>، جهش مرزی<sup>۳</sup> و ... اشاره کرد. در جهش یکنواخت ژنی که برای جهش انتخاب شده است، به صورت تصادفی و با تابع توزیع احتمال یکنواخت به یک عدد تصادفی در یک بازه مشخص تبدیل می‌شود. عملگرهای حسابی فقط در مورد کدگذاری غیر دودویی کاربرد دارند. عملگرهای حسابی برش، عمدتاً از ترکیب خطی برداری بهره می‌برند. به عنوان مثال اگر کروموزوم‌های  $x_1, x_2$  را به صورت دو بردار در نظر بگیریم، فرزندان حاصل از ترکیب آنها به صورت زیر خواهد بود:

$$x'_1 = \lambda_1 x_1 + \lambda_2 x_2 \quad (۱-۳)$$

$$x'_2 = \lambda_1 x_2 + \lambda_2 x_1$$

با توجه به نحوه انتخاب ضرائب  $\lambda_1, \lambda_2$ ، انواع گوناگون برش پدید می‌آیند که برش خطی و برش محدب از آن گونه‌اند. نوع دیگری از این ترکیب‌های خطی، عملگر ترکیب میانی<sup>۴</sup> می‌باشد. در این ترکیب اگر  $P_1, P_2$  والدین باشند که به منظور تولید فرزند انتخاب شده‌اند و  $O_1, O_2$  فرزندان حاصل از امتزاج آنها باشند، داریم:

$$O_1 = P_1 + \alpha(P_2 - P_1) \quad (۲-۳)$$

$$O_2 = P_2 + \alpha(P_1 - P_2)$$

---

<sup>۱</sup>- Flat crossover

<sup>۲</sup>- Uniform mutation

<sup>۳</sup>- Boundary mutation

<sup>۴</sup>- Intermediate recombination

باید توجه داشت که فاکتور مقیاس<sup>۱</sup> که با  $\alpha$  نمایش داده شده است، برای هر ژن به طور جداگانه و به طور تصادفی در بازه‌ی  $[-0.25, 1/25]$  تولید می‌شود. این وضعیت موجب می‌شود تا محدوده ممکن برای تولید فرزند، در فضای اطراف والدین باشد و بنابراین جستجوی بهتری انجام شود.

عملگرهایی که تاکنون نام برده شدند، تضمینی برای تولید فرزندان بهتر از والدین خود نمی‌کنند. در صورتی که در عملگرهای جهت‌دار، دانش مسأله بهینه‌سازی به یاری عملگرها آمده و فرزندان بهتر از والدین معمولاً تولید می‌شوند. به عنوان مثال در یک برش جهت‌دار فرزند حاصله از  $x_1$  و  $x_2$  به صورت زیر خواهد بود:

$$x' = r(x_2 - x_1) + x_2 \quad (3-3)$$

که در آن  $r$  عددی تصادفی در فاصله صفر و یک بوده و به نحوی فرض می‌شود که از دیدگاه بهینه سازی کروموزوم  $x_1$  ارزش کمتری از کروموزوم  $x_2$  دارد [۴۵].

### ۳-۶- پارامترهای GA<sup>۲</sup>

دو پارامتر اساسی در GA شامل احتمال (نرخ) پیوند و احتمال (نرخ) جهش می‌باشند. در کنار این پارامترهای مهم یکسری عوامل دیگر نیز ایفای نقش می‌کنند که در ادامه به اختصار به این پارامترها خواهیم پرداخت [۴۲-۴۴].

### ۳-۶-۱- احتمال (نرخ) پیوند

اگر فرآیند پیوند نباشد، فرزندان دقیقاً همانند والدین خواهند بود و در صورت وجود پیوند، فرزندان از قسمت‌های مختلف کروموزوم والدین ساخته می‌شوند. اگر احتمال پیوند ۱۰۰٪ باشد در

<sup>۱</sup>- Scaling factor

<sup>۲</sup>- Parameters of GA

این صورت تمامی فرزندان به وسیله‌ی پیوند ساخته می‌شوند و اگر ۰٪ باشد تمام نسل جدید دقیقاً کپی کروموزوم‌های جمعیت پیشین خواهد بود. پیوند با این هدف ساخته شده است که کروموزوم‌های جدید بخش‌های خوب کروموزوم‌های قبلی را در برداشته و از اینرو کروموزوم‌های جدید بهتر باشند. به هر حال مناسب است که برخی بخش‌های جمعیت پیشین را برای حضور در نسل بعدی حفظ کنیم.

### ۳-۶-۲- احتمال (نرخ) جهش

اگر فرآیند جهش نباشد، فرزندان بلافاصله پس از پیوند (یا کپی مستقیم) و بدون هیچ تغییری تشکیل می‌شوند و اگر جهش داشته باشیم یک بخش (یا بیشتر) از یک کروموزوم تغییر می‌کند. اگر احتمال جهش ۱۰۰٪ باشد تمام کروموزوم تغییر کرده و اگر ۰٪ باشد هیچ چیز تغییر نمی‌کند. به‌طور کلی جهش از سقوط به نقاط بهینه محلی جلوگیری می‌کند. جهش نایستی زیاد اتفاق بیافتد زیرا در این صورت GA در اصل به شکل جستجوی اتفاقی در خواهد آمد.

### ۳-۶-۳- اندازه جمعیت<sup>۱</sup>

اندازه جمعیت یعنی اینکه چه تعداد کروموزوم در جمعیت (در یک نسل) وجود دارد. اگر تعداد کمی کروموزوم داشته باشیم، GA احتمال کمی برای اجرای پیوند دارد و تنها یک بخش کوچک از فضای جستجو مورد استفاده قرار می‌گیرد. از طرف دیگر اگر تعداد کروموزوم‌ها بسیار زیاد باشد، GA سرعت کمی خواهد داشت. تحقیقات نشان می‌دهد که از یک محدوده‌ی خاص به بعد (که به روش رمزبندی و خود مسأله بستگی دارد) بکار بردن جمعیت‌های زیاد سودمند نیست، زیرا مسأله را سریع‌تر از حالت جمعیت‌های با اندازه متوسط حل نمی‌کند [۴۲].

---

<sup>۱</sup> - Population size

### ۳-۶-۴- شایستگی

برای مدل کردن منطق انتخاب طبیعی در الگوریتم ژنتیکی از تابع شایستگی<sup>۱</sup> یا تابع ارزیابی<sup>۲</sup> بهره می‌بریم. در واقع، این تابع، که می‌توان ادعا کرد مهم‌ترین قسمت الگوریتم ژنتیکی است، نقش محیط را در تکامل داروینی بازی می‌کند. بسته به نوع مسأله، تابع شایستگی کروموزوم‌ها را به صورت کمی ارزیابی می‌کند، به طوری که انتخاب درست این تابع دارای بزرگ‌ترین تأثیر در صحت عملکرد الگوریتم ژنتیکی است. برای مثال، این تابع می‌تواند محاسبه کننده خطا در جواب یک مسأله یا انرژی مربوط به ساختار یک پروتئین باشد. در این مثال‌ها، کروموزوم‌هایی برای تولید مثل شایسته‌تر شناخته می‌شوند که خطای کمتری در جواب مسأله ایجاد کرده و یا پروتئینی با سطح انرژی ساختاری پایین‌تر را رمزبندی کنند. با توجه به نوع مسأله‌ای که با الگوریتم ژنتیکی بهینه می‌شوند، معمولاً محاسبه تابع شایستگی وقت گیرترین بخش الگوریتم است.

### ۳-۶-۴-۱- تابع شایستگی

برای تابع شایستگی تعاریف متنوعی ارائه شده است. راجیو و کریشنامورتی<sup>۳</sup> در مقاله‌ی خود [۴۶]، رابطه‌ی زیر را پیشنهاد نموده‌اند:

$$F = [\phi_{\max}(X) + \phi_{\min}(X)] - \phi(X) \quad (۴-۳)$$

در رابطه‌ی فوق F و  $\phi(X)$  به ترتیب مقادیر شایستگی و تابع هدف اصلاح شده هر کروموزوم و  $\phi_{\max}(X)$  و  $\phi_{\min}(X)$  نیز حداکثر و حداقل مقدار تابع هدف اصلاح شده در جمعیت جاری می‌باشد. طبق رابطه‌ی بالا کروموزومی با کمترین مقدار تابع هدف اصلاح شده، بیشترین مقدار شایستگی به آن

---

<sup>۱</sup>- Fitness function

<sup>۲</sup>- Evaluation function

<sup>۳</sup>- Rajeev and Krishnamoorthy

تعلق می‌گیرد. ساکاموتا و ادا<sup>۱</sup> در مقاله‌ی خود [۴۷]، تابع شایستگی مقیاس شده  $F^2$  را به صورت زیر تعریف کرده‌اند:

$$F = -A\phi(X) + b \quad (۵-۳)$$

$$\phi(X) = f(X) + r \sum_{q=1}^Q \max[0, g_q(X)]$$

$$b = \phi_{avg} (\alpha \phi_{avg} - \phi_{min}) / (\phi_{avg} - \phi_{min})$$

$$a = \phi_{avg} (\alpha - 1) / (\phi_{avg} - \phi_{min})$$

در روابط فوق  $\phi_{avg}$  و  $\phi_{min}$  به ترتیب مقدار متوسط و حداقل  $\phi(X)$  در کل جمعیت،  $\alpha$  پارامتری ثابت و برابر با  $1/5$  و ثابت  $r$  متناسب با انتخاب طراح است. تابع شایستگی مقیاس شده به واسطه‌ی نزدیک‌تر کردن محدوده‌ی پراکنش تابع‌های شایستگی جمعیت حول  $\phi_{avg}$  سبب همگرایی بهتر می‌شود.

کروموزوم‌هایی که شایستگی آنها کمتر از  $\phi_{avg}$  باشد به سمت بالا و کروموزوم‌هایی که شایستگی آنها بیشتر از  $\phi_{avg}$  باشد به سمت پایین مقیاس می‌شوند. به این ترتیب حاکمیت کروموزوم‌های غیر متعارف از بین رفته و همگرایی و کاوش بهتر در فضای طراحی انجام می‌پذیرد. ارباتور<sup>۳</sup> و همکارانش نیز الگوریتمی برای تعیین مقدار تابع شایستگی مقیاس شده هر کروموزوم ارائه نموده‌اند [۴۸].

### ۳-۶-۵- انتخاب

قبل از طرح کلی GA، کروموزوم‌ها برای اینکه نقش والدین را در فرآیند پیوند بازی کنند از میان جمعیت انتخاب می‌شوند. اما مسأله چگونگی انتخاب این کروموزوم‌ها می‌باشد. با توجه به نظریه تکامل تدریجی داروین، بهترین کروموزوم‌ها برای تولید فرزندان انتخاب می‌شوند و روش‌های بسیاری

<sup>۱</sup> - Sakamoto and Oda

<sup>۲</sup> - Scaled fitness function

<sup>۳</sup> - Erbatur

برای انتخاب بهترین کروموزوم‌ها وجود دارد. برای نمونه می‌توان روش‌هایی مانند روش انتخاب چرخ گردان<sup>۱</sup>، روش انتخاب دوره‌ای، روش انتخاب رتبه‌ای، روش انتخاب حالت پایا<sup>۲</sup> و ... را نام برد [۳۹،۴۲].

### ۳-۶-۵-۱- روش انتخاب چرخ گردان

مراحل عملیاتی این روش در مراحل زیر بیان می‌گردد:

- ۱- میزان شایستگی یا همان عدد تابع شایستگی برای تمام کروموزوم‌ها تعیین می‌شود.
- ۲- نسبت شایستگی هر کروموزوم به مجموع شایستگی کل جمعیت محاسبه می‌گردد.
- ۳- میزان تجمعی نسبت‌های شایستگی برای هر کروموزوم محاسبه می‌شود.
- ۴- به هر کروموزوم یک عدد تصادفی بین صفر و یک اختصاص داده می‌شود.
- ۵- اعداد تصادفی با مقدار تجمعی نسبت‌های شایستگی مقایسه شده و در صورت قرارگیری هر عدد تصادفی در قطاع مربوط به هر کروموزوم، کروموزوم مربوطه جهت مشارکت در تولید فرزندان نسل آتی انتخاب می‌گردد. کروموزوم با شایستگی بالا قطاع بزرگتری را در بین نسبت‌های تجمعی شایستگی به خود اختصاص می‌دهد و در نتیجه به همان نسبت شانس انتخاب شدن بیشتری خواهد داشت و بالعکس.

### ۳-۷- اجرای الگوریتم ژنتیک

شکل (۳-۷) مراحل اجرای یک الگوریتم ژنتیک ساده را نشان می‌دهد که به‌طور خلاصه توضیح

داده می‌شود:

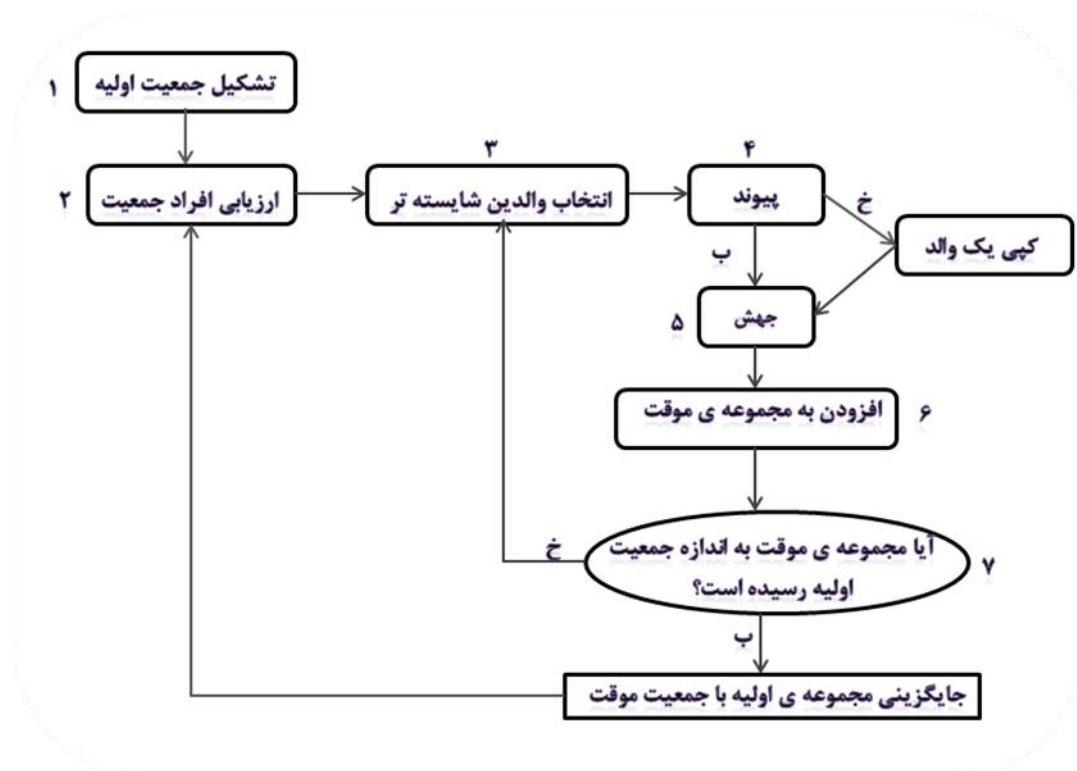
- ۱- در ابتدا، به صورت تصادفی جمعیتی از کروموزوم‌ها تشکیل می‌شوند. تعداد افراد جمعیت کاملاً انتخابی بوده و بستگی به نوع مسأله دارد.

---

<sup>۱</sup>- Roulette wheel selection

<sup>۲</sup>- Steady-state selection

- ۲- پس از تشکیل جمعیت، هر فرد با استفاده از تابع شایستگی ارزیابی می‌شود.
- ۳- با روش چرخه‌ی رولت افراد شایسته‌تر با احتمال بیشتری انتخاب می‌شوند. تعداد افراد انتخاب شده نیز امری انتخابی است و نخبه‌گرایی نام دارد. هرچه این تعداد بیشتر باشد، افراد شایسته جمعیت بهتر حفظ می‌شوند.
- ۴- افراد انتخاب شده با احتمال  $P_C$  باهم پیوند داده می‌شوند. برای این کار به صورت تصادفی یک عدد بین صفر و یک انتخاب می‌شود و اگر این عدد کمتر از  $P_C$  بود عمل تقاطع انجام شده، در غیر این صورت یکی از والدین به صورت تصادفی کپی می‌شود.
- ۵- عملگر جهش نیز با همان مکانیسم با احتمال  $P_M$  روی کروموزوم یا کروموزوم‌های ایجاد شده انجام می‌پذیرد.
- ۶- کروموزوم‌های ایجاد شده به یک مجموعه موقت از افراد می‌پیوندند.
- ۷- مراحل ۳ تا ۶ آن قدر تکرار می‌شوند تا تعداد افراد موجود در مجموعه موقت با تعداد افراد جمعیت ابتدایی برابر شود. وقتی این اتفاق افتاد، افراد مجموعه موقت جایگزین افراد جمعیت ابتدایی شده و به این ترتیب نسل جدیدی تشکیل می‌شود. سپس، الگوریتم دوباره از مرحله ۲ تکرار می‌شود.
- ۸- مراحل ۱ تا ۷ آن قدر تکرار می‌شوند تا تعداد نسل‌ها به عدد تعیین شده برسد. شایسته‌ترین فرد نسل آخر جواب مسأله است.



شکل (۳-۷) - فلوجارت یک الگوریتم ژنتیکی ساده

### ۳-۸ - الگوریتم ژنتیک و QSPR

نخستین کاربرد الگوریتم‌های ژنتیک در کمومتریکس بوسیله لاردی<sup>۱</sup> در سال ۱۹۹۲ گزارش شد که الگوریتم ژنتیک را برای انتخاب متغیر و رگرسیون خطی چندگانه را برای بهینه‌سازی شکل و اندازه خرپاها بکار برد [۴۹].

الگوریتم ژنتیک می‌تواند برای انتخاب متغیر و مدل‌سازی با روش‌هایی دیگر از قبیل رگرسیون خطی چندگانه، حداقل مربعات جزئی و شبکه‌ی عصبی و دیگر روش‌های مدل‌سازی بکار برده شود. که این روش ترکیب شده از تمام مزایای الگوریتم ژنتیک بهره می‌برد. GA-MLR برای مسائل غیر خطی حداقل مربعاتی که توابع آنها ترکیب‌های خطی از توابع غیر خطی هستند، طراحی شده است. الگوریتم ژنتیک پارامترهای غیر خطی را بهینه کرده و پارامترهای خطی نیز به وسیله رگرسیون خطی چندگانه در موازات یکدیگر محاسبه می‌شوند [۵۰].

<sup>۱</sup> - Leardi



فصل چہارم

شبکہ عصبی مصنوعی

## ۴-۱- مقدمه‌ای بر شبکه‌های عصبی

شبکه‌ی عصبی مصنوعی، شبکه‌ای محاسباتی است که به منظور پردازش اطلاعات، تلاش می‌کند سیستمی مانند سلول‌های عصبی زیستی را شبیه‌سازی کند. سیستم عصبی موجود زنده از تجمع و آرایش ویژه‌ی عناصر پردازش کننده‌ای به نام نرون<sup>۱</sup> تشکیل شده است که با مکانیسم فراگیری و آموزش و با تأثیرپذیری و تأثیرگذاری متقابل بر دستگاه‌های دیگر بدن ارتباط هدفمند جاندار را با محیط اطرافش ممکن می‌سازد. دریافت پیام‌های بیرونی، تحلیل و پاسخ‌دهی به آنها منوط به چینش صحیح و عملکرد سریع نرون‌ها می‌باشد که این شرایط از طریق تنظیم نحوه‌ی اتصالات میان آنها حاصل می‌شود. مدل‌های مصنوعی شبکه‌های عصبی طبیعی نیز بر اساس همین الگو بنا شده‌اند و هر کدام برای کاربردی ویژه مانند تشخیص الگو یا طبقه‌بندی اطلاعات از طریق فرآیند آموزش با مثال، صورت‌بندی شده‌اند. تحقیقات نشان می‌دهد که مغز اطلاعات را همانند الگوها ذخیره می‌کند که فرآیند ذخیره سازی اطلاعات به صورت الگو و تجزیه و تحلیل آن الگو اساس روش نوین محاسباتی را تشکیل می‌دهند. این حوزه از دانش محاسباتی از شبکه‌های بزرگی که به صورت موازی آرایش یافته و آموزش می‌بینند بهره می‌برد. شبکه‌های عصبی با توانایی چشمگیرشان در استنتاج مفاهیم از اطلاعات پیچیده یا مبهم، می‌توانند به منظور استخراج مدل‌ها و شناسایی نظام‌هایی که برای تشخیص توسط انسان‌ها یا تکنیک‌های کامپیوتری بسیار پیچیده هستند، بکار روند. شبکه‌های عصبی اطلاعات را همانند مغز انسان‌ها پردازش می‌کنند و به همین دلیل روش برخورد چنین مدل‌هایی با مسائل با روش‌های محاسباتی معمول در سیستم‌های ریاضی متداول که دانسته‌ها به صورت قواعد روشن و از پیش مشخص شده در سیستم وجود دارند، تفاوت دارد. در یک شبکه‌ی مصنوعی نیز می‌توان با تنظیم مقادیر هر اتصال تحت عنوان وزن اتصال، نحوه ارتباط بین اجزای آن را تعیین کرد. پس از تنظیم یا همان آموزش شبکه بر مبنای تطابق<sup>۲</sup> و همسنجی میان

---

<sup>۱</sup>- Neuron

<sup>۲</sup>- Adaptive

ورودی و هدف خروجی شبکه بر هدف منطبق می‌گردد. استفاده از زوج داده‌های ورودی و هدف بیانگر یادگیری با ناظر<sup>۱</sup> می‌باشد.

دوره‌ی ۱۹۴۰ تا ۱۹۶۰ را شاید بتوان به عنوان اولین دوره‌ی شکل‌گیری شبکه‌های عصبی مصنوعی در نظر گرفت. در حدود سال ۱۹۴۰ بطور همزمان اما جداگانه، از سوی نوروفیزیولوژیست‌ها سعی می‌کردند سیستم یادگیری و تجزیه و تحلیل مغز را کشف کنند و از سوی ریاضیدانان تلاش می‌کردند تا مدل ریاضی بسازند که قابلیت فراگیری و تجزیه و تحلیل عمومی مسائل را داشته باشد. در سال ۱۹۴۳ یک مدل خطی ساده توسط مک کلوج<sup>۲</sup> و پیتز<sup>۳</sup> مطرح شد [۵۱] و نخستین کاربرد عملی شبکه‌های عصبی برای شناسایی الگوها در اواخر دهه ۵۰، زمانیکه روزنبلات<sup>۴</sup> در سال ۱۹۵۸ شبکه پرسپترون<sup>۵</sup> را معرفی کرد، مطرح شد [۵۲]. سال ۱۹۶۹ آغاز افول موقت شبکه‌های عصبی بود، زیرا عدم توانایی شبکه‌های عصبی در حل مسائل غیر خطی آشکار شد. شبکه‌های عصبی مصنوعی آن زمان فقط قادر به حل مسائلی بودند که می‌توانستند پاسخ‌های آن مسئله را توسط یک خط در محور مختصات از هم جدا کنند. در سال ۱۹۸۲ هاپفیلد<sup>۶</sup> با معرفی شبکه‌های چند لایه و برگشتی، راه حلی برای حل مسائل غیر خطی ارائه کرد [۵۳] و سپس دیوید رامله‌هارت<sup>۳</sup> و جیمز مک‌کلند<sup>۴</sup> الگوریتم پس انتشار خطا<sup>۷</sup> را در سال ۱۹۸۶ مطرح کردند [۵۴]. امروزه با افزوده شدن قابلیت‌های بیشتر به شبکه‌های عصبی مصنوعی، این شبکه‌ها توانسته‌اند به عنوان ابزاری قدرتمند در مسائل بسیار متنوعی از قبیل تشخیص الگو، شناسایی گفتار، پردازش تصویر و حتی در کنترل یا مدل‌سازی سیستم‌هایی که ساختار داخلی ناشناخته یا پیچیده دارند، بکار روند.

---

<sup>۱</sup>- Supervised learning

<sup>۲</sup>- McClluch

<sup>۳</sup>- Pitts

<sup>۴</sup>- Rosenblat

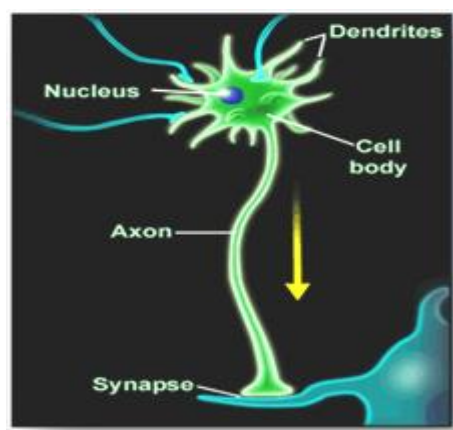
<sup>۵</sup>- Perceptrons

<sup>۶</sup>- Hopfield

<sup>۷</sup>- BackPropagation

## ۴-۲- نرون بیولوژیکی

هر نرون طبیعی دارای سه قسمت اصلی دندریت<sup>۱</sup>، جسم سلولی<sup>۲</sup> و اکسون<sup>۳</sup> است. محل تلاقی اکسون هر سلول به دندریت‌های سلول‌های دیگر را سیناپس می‌گویند که ارتباطات مابین نرون‌ها توسط سیناپس‌ها برقرار می‌شود. در حقیقت دندریت‌ها به عنوان ورودی نرون و اکسون به عنوان خروجی و فضای سیناپسی محل اتصال این دو می‌باشد. زمانی که سیگنال عصبی از اکسون به نرون‌ها و یا عناصر دیگر بدن مثل ماهیچه‌ها می‌رسد، باعث تحریک آنها می‌شود. نرون‌ها از هر یک از اتصالات ورودی خود یک ولتاژ کم دریافت می‌کنند و آنها را با هم جمع می‌زنند. اگر این حاصل جمع به یک مقدار آستانه برسد، نرون روی اکسون خود یک ولتاژ خروجی ارسال می‌کند. این ولتاژ به دندریت‌هایی که به این اکسون متصلند رسیده و باعث یکسری فعل و انفعال‌های شیمیایی در اتصالات سیناپسی می‌شود و می‌تواند باعث ارسال پیک‌های فعالیت الکتریکی بین نرون‌های دیگر گردد [۵۴]. شکل (۴-۱) نمای کلی یک نرون را نشان می‌دهد.



شکل (۴-۱) - ساختار نرون طبیعی

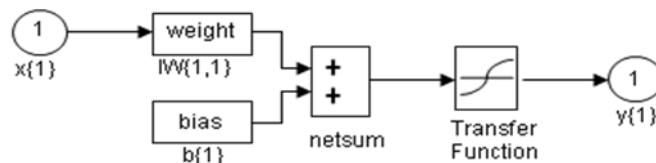
<sup>۱</sup>- Dendrite

<sup>۲</sup>- Cell body, Soma

<sup>۳</sup>- Axon

### ۳-۴- ساختار و عملکرد نرون مصنوعی

همانند نرون بیولوژیکی، یک نرون مصنوعی را می‌توان به صورت شبکه‌ای از عناصر محاسباتی به هم پیوسته به نام گره در نظر گرفت. هر گره، تبدیلی بر روی سیگنال‌های ورودی‌اش صورت داده و سیگنال خروجی تولید می‌کند که به نوبه‌ی خود به گره‌های دیگر فرستاده خواهد شد [۵۵]. عمل لایه‌ی ورودی، لایه‌ی میانی و لایه‌ی خروجی در یک شبکه عصبی به ترتیب شبیه عمل دندریت، جسم سلولی و اکسون در یک نرون بیولوژیکی می‌باشد. در بدنه‌ی یک نرون مصنوعی، ورودی‌ها در ضرایب مخصوصی که وزن<sup>۱</sup> ( $w$ ) نام دارد ضرب شده و وارد نرون می‌گردند و سپس مشابه آنچه در نرون‌های طبیعی دیدیم با هم جمع می‌شوند. این حاصل جمع که معادل پتانسیل عمل در یک نرون طبیعی می‌باشد وارد تابع محرک یا انتقال<sup>۲</sup> می‌گردد و خروجی نرون حاصل می‌شود [۵۶]. ساختار و عملکرد یک نرون مصنوعی ساده تک لایه برای یک ورودی در شکل (۲-۴) نمایش داده شده است.



شکل (۲-۴) - ساختار یک نرون عصبی مصنوعی

روابط حاکم بر گره‌ها به صورت زیر می‌باشند:

$$z_i = \sum_{j=1}^n (w_{ij} \cdot x_j + b) \quad (1-4)$$

$$y_i = f(z_i) \quad (2-4)$$

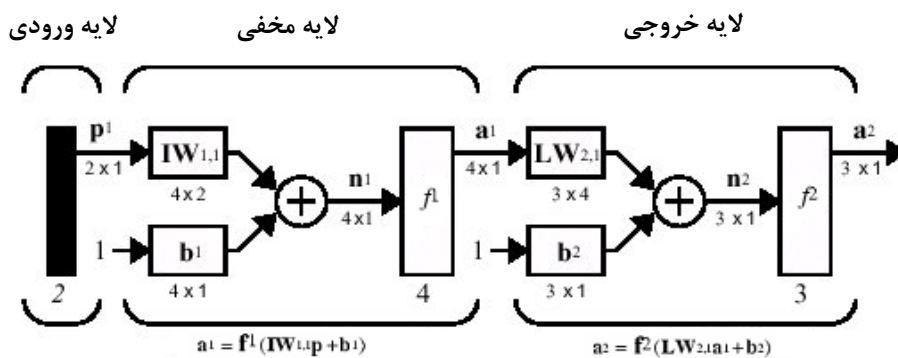
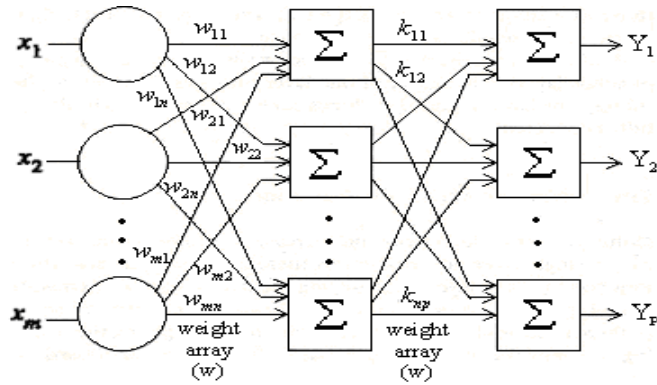
در روابط فوق،  $x_j$  خروجی‌های گره‌های لایه‌های قبل،  $w_{ij}$  وزن‌های متناظر با آنها،  $n$  تعداد ورودی‌ها،  $b$  یک مقدار تنظیم (بایاس<sup>۳</sup>)،  $z_i$  خروجی تابع جمع‌کننده و  $y_i$  خروجی گره می‌باشند.

<sup>۱</sup> - Weight

<sup>۲</sup> - Transfer function

<sup>۳</sup> - Bias

در شبکه‌های چندلایه، خروجی هر نرون وارد تک تک نرون‌های لایه‌ی بعدی می‌شود. شکل (۳-۴) یک شبکه‌ی سه لایه و نمایش مختصر شده‌ی آن را نشان می‌دهد که  $a_1$  خروجی لایه دوم که لایه مخفی<sup>۱</sup> نیز نامیده می‌شود وارد لایه سوم به عنوان لایه خروجی شده است.



شکل (۳-۴) - شبکه‌های عصبی مصنوعی سه لایه

بنابراین اجزای یک شبکه‌ی عصبی مصنوعی عبارتند از:

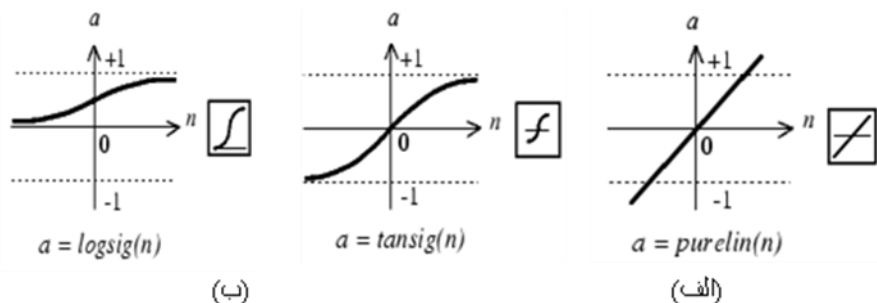
- ورودی‌ها: می‌توانند خروجی سایر لایه‌ها بوده و یا به حالت خام در اولین لایه به صورت داده‌های عددی و رقمی، متون ادبی، فنی، تصویر و یا شکل باشند.
- وزن‌ها: میزان تاثیر ورودی  $X_i$  بر خروجی  $y$ ، توسط وزن اندازه‌گیری می‌شود که با وارد شدن بایاس  $b_i$  مقدار وزن دار شده برای تطبیق بیشتر ورودی - خروجی نرون افزایش یا کاهش می‌یابد.
- تابع جمع‌کننده: در شبکه‌های تک نرونی، تابع جمع‌کننده خروجی مسأله را مشخص می‌کند و در شبکه‌های چند نرونی میزان سطح فعالیت نرون در لایه‌های درونی را مشخص می‌سازد.

<sup>۱</sup> - Hidden layers

● تابع تبدیل: وظیفه محدودسازی خروجی گره به یک بازه متناهی را بر عهده دارد. انواع متفاوتی از توابع تبدیل از قبیل پله‌ای، تکه‌ای-خطی و سیگموئیدی وجود دارند که بنا به ماهیت مسأله کاربرد پیدا می‌کنند. این تابع توسط طراح مسأله انتخاب شده و بر اساس انتخاب الگوریتم یادگیری، پارامترهای مسأله (وزن‌ها) تنظیم می‌گردد. در شبکه‌های امروزی، تابع تبدیل اکثراً از نوع لگاریتم سیگموئید<sup>۱</sup> و یا تانژانت سیگموئید<sup>۲</sup> است.

جدول (۱-۴) - توابع تبدیل استفاده شده [۵۹]

نام تابع	عملیات ریاضی
purelin	$x$
logsig	$\frac{1}{1 + e^{-x}}$
tansig	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$



شکل (۴-۴) - توابع تبدیل (الف) خطی، (ب) سیگموئیدی [۵۹]

● خروجی: منظور از خروجی، پاسخ مسأله است.

فرضیات مهم در شبکه‌های عصبی مصنوعی از این قرار است:

۱- داده پردازشی اطلاعات در اجزای ساده به نام نرون صورت می‌گیرد.

۲- اطلاعات بین نرون‌ها از طریق ارتباطات آنها رد و بدل می‌شود.

<sup>۱</sup>- Log-sigmoid

<sup>۲</sup>- Tan-sigmoid

۳- هر یک از این رابطه‌ها دارای وزن  $W$  مختص خود هستند که در مقدار اطلاعات رد و بدل شده با سایر نرون‌ها ضرب می‌شوند و به مرور زمان این وزن‌ها تنظیم می‌گردند. در واقع شبکه از محیط تاثیر پذیرفته و آموزش می‌بیند.

۴- هریک از نرون‌ها برای محاسبه خروجی خود، دارای یک تابع عملیاتی است که معمولاً تابعی غیرخطی است و روی ورودی‌ها اعمال می‌شود.

۵- زمانی که حاصل تابع عملیاتی هر نرون از یک آستانه حد بیشتر شود، خروجی ایجاد می‌شود.

#### ۴-۴- ویژگی‌های شبکه‌های عصبی

شبکه‌های عصبی مصنوعی با وجود آنکه با سیستم عصبی طبیعی قابل مقایسه نیستند، ویژگی‌هایی دارند که آنها را در هر جا که نیاز به یادگیری یک نگاشت خطی یا غیر خطی باشد ممتاز می‌نماید. از جمله‌ی این ویژگی‌ها می‌توان به موارد ذیل اشاره نمود:

#### ۴-۴-۱- قابلیت یادگیری

مهم‌ترین ویژگی شبکه‌های عصبی وجود قابلیت یادگیری در آنهاست، بدین ترتیب که در طول زمان که با تغییر محیط شبکه، شبکه شرایط جدیدی را تجربه می‌کند، قادر باشد با تغییر پارامترهایش، خود را با شرایط جدید وفق دهد.

#### ۴-۴-۲- پردازش موازی

به دلیل اینکه در شبکه‌های عصبی، وظیفه پردازش به پردازشگرهای کوچک‌تری (گره‌ها) محول می‌شود که در لایه‌های گوناگون به صورت موازی قرار گرفته‌اند، به هنگام پیاده‌سازی سخت افزاری، افزایش سرعت پردازش را شاهد خواهیم بود.



#### ۴-۴-۳- مقاوم بودن

آنچه که شبکه فرا می‌گیرد، در وزن‌های سیناپسی مستتر می‌باشد و رابطه‌ی یک به یک بین ورودی‌ها و وزن‌های سیناپسی وجود ندارد. پس هر وزن سیناپسی مربوط به تمام ورودی‌هاست ولی به هیچ‌یک از آنها به‌طور منفرد و مجزا مربوط نیست. به عبارتی هر نرون در شبکه از کل فعالیت سایر نرون‌ها متأثر می‌باشد. بر این اساس چنانچه بخشی از سلول‌های شبکه حذف شده یا عملکرد غلط داشته باشند باز هم احتمال رسیدن به پاسخ صحیح وجود دارد. اگر چه این احتمال برای تمام ورودی‌ها کاهش یافته ولی برای هیچ‌یک از آنها از بین نرفته است.

#### ۴-۴-۴- قابلیت تعمیم<sup>۱</sup>

پس از آنکه مثال‌های اولیه به شبکه آموزش داده شد، شبکه می‌تواند در مقابل یک ورودی آموزش داده نشده قرار گیرد و یک خروجی مناسب ارائه نماید. این خروجی بر اساس مکانیسم تعمیم که فرآیندی درونیابی است، بدست می‌آید [۵۷].

#### ۴-۵- یادگیری در شبکه‌های عصبی

همان‌طور که گفته شد مهم‌ترین ویژگی شبکه‌های عصبی قابلیت یادگیری آنهاست، چرا که بدین ترتیب شبکه قادر خواهد بود به هنگام اعمال شرایط جدید، با تنظیم پارامترهایش، خود را با آن شرایط وفق دهد. به‌طور کلی، سه روش برای یادگیری شبکه‌های عصبی وجود دارد:

#### ۴-۵-۱- یادگیری با ناظر

در یادگیری با ناظر، چنین فرض می‌شود که خروجی‌های مطلوب شبکه در دسترس می‌باشند، بدین ترتیب که در این یادگیری، پس از اعمال ورودی‌ها به شبکه، خروجی شبکه با خروجی مطلوب

---

<sup>۱</sup> - Generalization

مقایسه می‌گردد و پس از محاسبه‌ی خطای یادگیری، از آن جهت تنظیم پارامترهای شبکه به گونه‌ای استفاده می‌گردد که اگر مجدداً همین ورودی به شبکه اعمال گردد، خروجی شبکه به خروجی مطلوب، نزدیک‌تر باشد.

#### ۴-۵-۲- یادگیری بدون ناظر<sup>۱</sup>

در یادگیری بدون ناظر، تنها سیگنالی که از محیط به شبکه اعمال می‌گردد، بردار ورودی شبکه است، بدین ترتیب که دیگر ناظری وجود نخواهد داشت تا خروجی مطلوب را برای آموزش به شبکه اعمال نماید. شبکه‌هایی را که از این نوع یادگیری بهره می‌برند، اصطلاحاً شبکه‌های خود سازمان‌ده<sup>۲</sup> می‌گویند.

#### ۴-۵-۳- یادگیری تقویتی<sup>۳</sup>

یادگیری تقویتی را می‌توان حد فاصل یادگیری با ناظر و بدون ناظر دانست، بدین ترتیب که در این روش، عنصری به نام نقاد جایگزین ناظر می‌گردد و به جای آموزش سیستم، به نقد و بررسی عملکرد آن می‌پردازد و در نتیجه بجای اطلاعات کمی (مانند حالت یادگیری با ناظر)، اطلاعات کیفی ناشی از عملکرد سیستم را بیان می‌کند [۵۸].

#### ۴-۶- معماری شبکه‌های عصبی مصنوعی

شبکه‌های چندلایه بسیار قدرتمندتر از شبکه‌های تک‌لایه هستند. به تجربه نیز ثابت شده است که خطی بودن در لایه‌ی خروجی باعث می‌شود که شبکه از زمانی که تابع خروجی، سیگموئیدی باشد، بهتر به جواب برسد زیرا محدوده‌ی خطی بودن تابع خطی از تابع سیگموئیدی بیشتر است. همچنین

---

<sup>۱</sup>- Unsupervised

<sup>۲</sup>- Selforganized

<sup>۳</sup>- Reinforcement learning

با حضور تابع سیگموئیدی در لایه‌ی آخر، انحراف از محدوده‌ی خطی بودن در دو انتهای تابع مشاهده می‌شود [۵۷].

تعداد ورودی‌ها و خروجی‌های شبکه شامل پارامترهای آزاد کاربر نیستند و بر اساس مسأله خاصی که شبکه قرار است آن را حل کند بدست می‌آیند. اما تعداد لایه‌های مخفی، تعداد نرون‌های لایه‌های مخفی و سایر پارامترها (بسته به الگوریتم مورد استفاده) مربوط به درجات آزادی بوده و باید توسط کاربر بهینه شوند [۵۹].

تعداد نرون‌های هر لایه‌ی مخفی بین نصف تا یک و نیم برابر تعداد ورودی‌های سیستم در نظر گرفته می‌شود. اگر تعداد نرون‌های لایه‌ی مخفی زیاد انتخاب شود شبکه با تغییرات متعدد وزن‌ها و بایاس‌ها می‌کوشد بهترین تطبیق را با داده‌های سری آموزش بدست آورد در حالی که این موضوع باعث می‌شود که شبکه نتواند با داده‌های سری پیش‌بینی تطبیق ایجاد نماید و اگر تعداد نرون‌ها کم باشد قدرت تجزیه و تحلیل و دقت کم می‌شود [۶۰].

شبکه‌های عصبی دارای توپولوژی‌های مختلفی می‌باشند، اما در یک دسته‌بندی کلی می‌توان آنها را به دو دسته شبکه‌های پیشخور<sup>۱</sup> و شبکه‌های پسخور<sup>۲</sup> تقسیم کرد.

در شبکه‌های پیشخور، هیچگونه بازگشتی<sup>۳</sup> از یکی از گره‌های شبکه به گره‌های ما قبل آن یا گره‌های واقع در همان لایه وجود ندارد، در صورتی که چنین بازگشتی در شبکه‌های پسخور دیده می‌شود. شبکه‌های پیشخور معمولاً از یک یا چند لایه پنهان با تابع تبدیل سیگموئیدی و یک لایه خارجی با تابع تبدیل خطی تشکیل شده‌اند. از نمونه‌های بارز دسته‌ی اول می‌توان به شبکه‌ی پرسپترون چند لایه<sup>۴</sup> و آدالین<sup>۵</sup> و از نمونه‌های گروه دوم به کوهونن<sup>۶</sup> و هاپفیلد اشاره نمود [۵۷].

---

۱- Feed Forward

۲- Feedback

۳- Recurrent

۴- Multilayer Perceptron (MLP)

۵- Adaline

۶- Kohonen

#### ۴-۷- شبکه‌های پرسپترون چند لایه

مدل ریاضی یک سلول عصبی که از روی الگوی ریاضی آن ساخته شده است پرسپترون نامیده می‌شود. شبکه‌هایی که از یک نرون ساخته شده‌اند دارای محدودیت‌هایی می‌باشند. از جمله اینکه توانایی پیاده‌سازی توابع غیر خطی را ندارند. شبکه‌های پرسپترون چند لایه، در زمره کاربردی‌ترین شبکه‌های عصبی می‌باشند. این شبکه قادر است یک نگاشت غیر خطی را با دقت دلخواه انجام دهد و این همان چیزی است که در مسائل فنی-مهندسی به عنوان یک راه حل اصلی مطرح می‌باشد. این شبکه نماینده‌ی شبکه‌های پیشخور است و خروجی بدون هیچ بازگشتی مستقیماً از روی ورودی محاسبه می‌گردد. مدل عمومی شبکه‌های پرسپترون، شبکه‌ی پیشخور با روال آموزش پس انتشار است. شبکه‌های پرسپترون از یک لایه ورودی، تعدادی لایه پنهان و یک لایه خروجی تشکیل شده است [۶۰].

#### ۴-۸- الگوریتم‌های آموزشی شبکه‌های عصبی

در شبکه‌های عصبی برای بدست آوردن ارتباط بین متغیرهای ورودی و خروجی از فرآیند یادگیری با یک الگوی آموزشی استفاده می‌شود. قدرت شبکه‌های عصبی در همین آموزش پذیری آنهاست که آنها را قادر خواهد ساخت که به‌طور نسبی به تعمیم آموخته‌های خویش بپردازد چرا که آنچه یادگیری را از حفظ کردن متفاوت و بالاتر می‌سازد همین قدرت تعمیم آموخته‌ها و اطلاعات اولیه است [۵۷]. برای آموزش شبکه‌های عصبی از الگوریتم‌های یادگیری متفاوتی مانند الگوریتم پس انتشار خطا، الگوریتم پس انتشار گرادیان مزدوج<sup>۱</sup> و الگوریتم لوبنبرگ-مارکورت<sup>۲</sup> استفاده می‌شود. انتخاب هر الگوریتم بر سرعت یادگیری و دقت شبکه موثر است.

---

<sup>۱</sup>- Conjugate gradient

<sup>۲</sup>- Levenberg - Marquardt (LM)

#### ۴-۸-۱- الگوریتم آموزشی پس انتشار خطا

در شبکه‌های عصبی پیشرونده چند لایه که با انتخاب معماری مناسب و انتخاب الگوریتم آموزش مناسب مواجه هستیم، با استفاده از الگوریتم‌های مختلف تکرار که متداول‌ترین آنها الگوریتم پس انتشار خطا است، وزن‌های شبکه با هر تکرار تغییر کرده و برای کمینه‌کردن تابع کارایی<sup>۱</sup> تعدیل می‌شوند (اصطلاحاً شبکه آموزش داده می‌شود) [۶۱]. تابع کارایی استفاده شده در شبکه‌های عصبی پیشخور میانگین مربعات خطای شبکه است. معماری به معنی تعداد نرون‌ها در هر لایه می‌باشد که انتخاب تعداد نرون‌ها در هر لایه مبتنی بر مجموعه داده‌ها و خصایص آنهاست [۶۲]. الگوریتم پس انتشار خطا جهت آموزش از قانون یادگیری دلتا و گرادیان نزولی<sup>۲</sup> استفاده می‌کند. قانون یادگیری که وزن‌ها و بایاس‌های شبکه را با بردارهای مشتق خطای برگشت داده شده به عقب تنظیم می‌کند به قاعده دلتا<sup>۳</sup> مشهور است. الگوریتم پس انتشار خطا از دو مسیر اصلی تشکیل می‌شود: مسیر اول، به مسیر رفت موسوم می‌باشد که در این مسیر بردار ورودی به شبکه اعمال شده و تأثیراتش از طریق لایه‌های میانی به لایه‌ی خروجی انتشار می‌یابد. بردار خروجی تشکیل یافته در لایه خروجی پاسخ واقعی شبکه را تشکیل می‌دهد. در این مسیر پارامترهای شبکه ثابت و بدون تغییر در نظر گرفته می‌شوند. مسیر دوم، به مسیر برگشت موسوم می‌باشد که در این مسیر، برعکس مسیر رفت پارامترهای شبکه تغییر و تنظیم می‌گردند که این تنظیم مطابق با قانون اصلاح خطا انجام می‌گیرد و سیگنال خطا در لایه خروجی شبکه تشکیل می‌گردد. بردار خطا برابر اختلاف بین پاسخ مطلوب و پاسخ واقعی شبکه می‌باشد. مقدار خطا پس از محاسبه در مسیر برگشت از لایه خروجی و از طریق لایه‌های شبکه در کل شبکه توزیع می‌گردد [۵۷]. مراحل آموزش به کمک این الگوریتم عبارتند از [۶۳، ۶۴]: (الف) اختصاص ماتریس وزن تصادفی به هریک از اتصالات، (ب) انتخاب بردار ورودی و خروجی متناسب با آن، (پ) محاسبه خروجی نرون در هر لایه و در نتیجه محاسبه خروجی نرون‌ها در

<sup>۱</sup>- Performance function

<sup>۲</sup>- Gradient descent

<sup>۳</sup>- Delta rule

لايهی خروجی، (ت) به روزسازی وزن‌ها به روش انتشار خطای شبکه به لایه‌های قبل که خطای یاد شده ناشی از اختلاف بین خروجی واقعی و خروجی محاسبه شده است، (ث) ارزیابی عملکرد شبکه آموزش دیده به کمک برخی شاخص‌های تعریف شده مانند جذر میانگین مربعات خطا و سرانجام برگشت به قسمت پ یا پایان آموزش.

برای اصلاح وزن‌ها از ضریب سرعت آموزش  $\alpha$  و ضریب ممنتوم  $\mu$  استفاده می‌شود. رابطه‌ای که اصلاح وزن‌ها را با کمک هر دو ضریب ذکر شده انجام می‌دهد، به صورت زیر می‌باشد:

$$\Delta w_{ij}(n) = -\alpha \cdot \frac{\partial E}{\partial w_{ij}} + \mu \cdot \Delta w_{ij}(n-1) \quad (3-4)$$

که  $\Delta w_{ij}(n)$ ،  $\Delta w_{ij}(n-1)$  به ترتیب میزان تغییر وزن ما بین گره‌های  $i$  و  $j$  در خلال  $n$  و  $(n-1)$  امین عبور یا چرخه<sup>۱</sup> می‌باشد. هر چند در الگوریتم کاهش گرادیان، خطا با یک روند خطی (درجه‌ی اول) به کمترین مقدار می‌رسد، اما سرعت همگرایی آن کم می‌باشد و احتمال به دام افتادن در مینیمم‌های محلی وجود دارد در نتیجه یادگیری شبکه سخت و کند است. افزودن عبارت ممنتوم در پس انتشار تاحدی به رفع مشکل همگرایی کمک می‌کند.

#### ۴-۸-۲- الگوریتم لونبرگ-مارکورت

همانطور که گفته شد الگوریتم پس انتشار با وجود معمول بودنش دارای دو ضعف عمده و اساسی است. اول سرعت همگرایی پایین و دوم ساکن شدن و متوقف شدن پارامترهای شبکه به دلیل به دام افتادن آنها در نقاط مینیمم محلی قبل از پایان یادگیری کامل شبکه می‌باشد. بنابراین از روش آموزشی لونبرگ-مارکورت که از نوع روش‌های آموزش درجه‌ی دوم است، استفاده می‌شود. این روش نسبت به روش پس انتشار صدها برابر سریعتر است. ویژگی دیگر این روش کاهش شدید تعداد نرون‌های لایه مخفی در ساختار شبکه است. این روش هم از مشتق اول (گرادیان) و هم مشتق دوم

---

<sup>۱</sup> - Epoch

موسوم به هسین<sup>۱</sup> برای اصلاح وزن‌ها استفاده می‌کند. این روش در حقیقت ترکیبی از دو روش کاهش گرادیان و گوس-نیوتن<sup>۲</sup> است که در آن از سرعت روش اول به همراه دقت بالای روش دوم در تعیین پارامترهای بهینه توابع غیر خطی استفاده می‌شود [۶۵].

طبق روش نیوتن اصلاح پارامترها به صورت زیر انجام می‌شود:

$$x_{k+1} = x_k - [H + \mu I]^{-1} J^T e \quad (۴-۴)$$

$$H(x) = \nabla^2 F(x) \approx J^T J \quad (۵-۴)$$

که  $x$  وزن‌های شبکه عصبی،  $H$  ماتریس هسین،  $J$  ماتریس ژاکوبین<sup>۳</sup>،  $\mu$  نرخ یادگیری و  $e$  بردار خطای باقیمانده است که به صورت زیر تطبیق پیدا می‌کند.

- اگر خطا نسبت به تکرار قبلی الگوریتم افزایش یافت،  $\mu$  افزایش می‌یابد تا سهم مشتق اول (ژاکوبین) نسبت به مشتق دوم افزایش پیدا کند.

- اگر خطا نسبت به تکرار قبلی الگوریتم کاهش یافت،  $\mu$  کاهش می‌یابد تا سهم مشتق دوم (هسین) افزایش پیدا کند.

ماتریس هسین به صورت معادله زیر نوشته می‌شود:

$$H = \nabla^2 F(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} F(x) & \frac{\partial^2}{\partial x_1 \partial x_2} F(x) & \dots & \frac{\partial^2}{\partial x_1 \partial x_n} F(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} F(x) & \frac{\partial^2}{\partial x_2^2} F(x) & \dots & \frac{\partial^2}{\partial x_2 \partial x_n} F(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} F(x) & \frac{\partial^2}{\partial x_n \partial x_2} F(x) & \dots & \frac{\partial^2}{\partial x_n^2} F(x) \end{bmatrix}_{n \times n} \quad (۶-۴)$$

که  $F(x)$  تابع مجموع مربعات خطا است.

<sup>۱</sup>- Hessian  
<sup>۲</sup>- Gauss-Newton  
<sup>۳</sup>- Jacobian

مشتق اول تابع خطا برای زامین عنصر گرادیان این گونه است:

$$[\nabla F(x)]_j = \frac{\partial F(x)}{\partial x_j} = 2 \sum_{i=1}^N e_i(x) \cdot \frac{\partial e_i(x)}{\partial x_j} \quad (7-4)$$

$$\nabla F(x) = 2J^T(x)e(x)$$

$$J(x) = \begin{bmatrix} \frac{\partial e_1}{\partial x_1} & \frac{\partial e_1}{\partial x_2} & \dots & \frac{\partial e_1}{\partial x_n} \\ \frac{\partial e_2}{\partial x_1} & \frac{\partial e_2}{\partial x_2} & \dots & \frac{\partial e_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N}{\partial x_1} & \frac{\partial e_N}{\partial x_2} & \dots & \frac{\partial e_N}{\partial x_n} \end{bmatrix} \quad (8-4)$$

که  $J(X)$  ماتریس ژاکوبین است.

بررسی قابلیت تعمیم شبکه‌ی عصبی آموزش دیده، آخرین مرحله در توسعه‌ی مدل شبکه‌ی عصبی مصنوعی است. برای یافتن شبکه‌ای با معماری مناسب به کمک الگوریتم‌های آموزشی، از رسم منحنی مقایسه‌ای مقادیر پیش‌بینی شده توسط مدل و مقادیر واقعی و نیز معیار مجذور خطای میانگین مربعات استفاده می‌شود [۶۵].

#### ۴-۹- ارتقای عمومیت<sup>۱</sup>

یکی از مشکلاتی که هنگام آموزش شبکه عصبی ممکن است پیش بیاید، بیش‌آموزی<sup>۲</sup> یا بیش‌برازش<sup>۳</sup> شبکه است. اگر شبکه بیش از حد آموزش داده شود، شروع به خاطر سپردن داده‌ها می‌کند و از پیدا کردن ارتباط موجود در داده‌ها دور می‌شود. بدین صورت که در هنگام آموزش شبکه، خطا به مقدار قابل قبول می‌رسد ولی هنگام ارزیابی داده‌های جدید، خطای شبکه به مراتب از خطای داده‌های

<sup>۱</sup>- Improving Generalization

<sup>۲</sup>- Overtraining

<sup>۳</sup>- Overfitting



آموزشی بیشتر می‌باشد که گفته می‌شود شبکه تعمیم برای حالت‌های جدید را یاد نگرفته است. یک روش برای ارتقای عمومیت شبکه، بکارگیری شبکه‌هایی بزرگ برای ایجاد یک انطباق مناسب است. اگر ما از شبکه کوچکی استفاده کنیم، این شبکه توانایی لازم را برای انطباق داده‌ها نخواهد داشت. متأسفانه دانستن این مورد از ابتدا، که اندازه شبکه برای یک کاربرد خاص چه میزان می‌بایست بزرگ باشد، مشکل است. در جعبه ابزار شبکه‌های عصبی برای جلوگیری از بیش آموزش و ارتقای عمومیت روش‌های مختلفی از جمله توقف اولیه<sup>۱</sup>، تنظیم<sup>۲</sup> و... وجود دارد.

در توقف اولیه برای آموزش شبکه، داده‌ها به‌طور تصادفی به سه قسمت به صورت ۶۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای ارزیابی و ۲۰ درصد برای آزمون شبکه تقسیم می‌شوند. در حین آموزش شبکه، زمانی که خطای بین داده‌های آموزش و ارزیابی می‌خواهد زیاد شود، فرآیند آموزش قطع می‌شود.

روش تنظیم، تابع عملکرد را که عموماً میانگین مربعات خطاهای شبکه برای مجموعه آموزش می‌باشد تعدیل می‌کند:

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (9-4)$$

استفاده از این تابع عملکرد سبب می‌شود که شبکه، وزن‌ها و بایاس‌های کوچکتری داشته باشد و این پاسخ شبکه را هموارتر می‌کند و باعث می‌شود که پاسخ شبکه کمتر بیش آموزش داشته باشد. اگر تعداد پارامترهای شبکه خیلی کمتر از تعداد نقاط در سری آموزشی باشد یا به عبارتی اگر سری داده‌ی بزرگی در سری آموزشی داشته باشیم احتمال بیش آموزش کمتر خواهد شد [۶۶].

---

<sup>۱</sup>- Early stopping

<sup>۲</sup>- Regularization

## فصل پنجم

# مدل سازی QSPR ثابت هنری



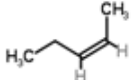
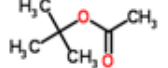

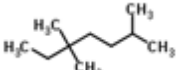
## ۵-۱- مدل سازی QSPR ثابت هنری

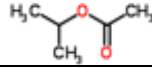
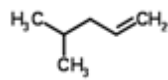
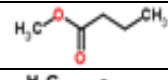
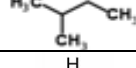

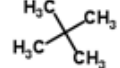

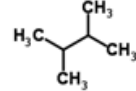
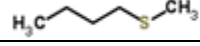
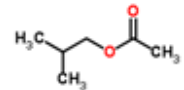
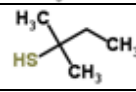
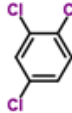
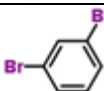
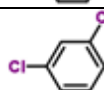
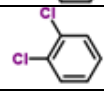
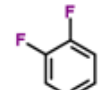
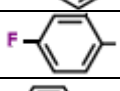
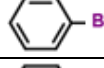
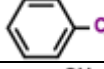

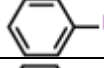
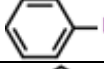

در این مرحله از پروژه، از رگرسیون خطی چندگانه و شبکه‌ی عصبی مصنوعی جهت مدل سازی و پیش‌بینی ثابت قانون هنری یک سری از ترکیبات آلی استفاده شده است. همچنین توانایی مدل‌های حاصله در پیش‌بینی این پارامتر با یکدیگر مورد مقایسه و ارزیابی قرار گرفته است.



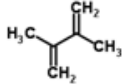

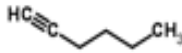
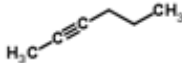
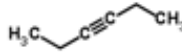

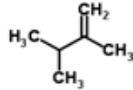
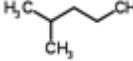
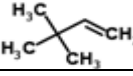

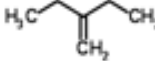
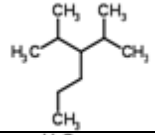
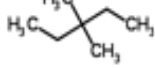
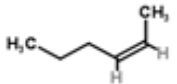
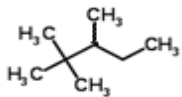
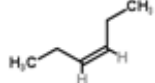
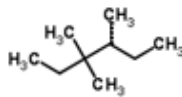
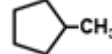
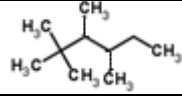
### ۵-۱-۱- سری داده‌ها

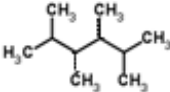

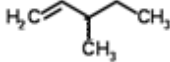
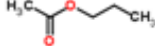
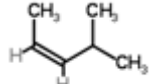
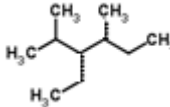
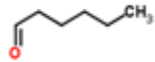
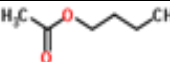
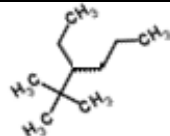
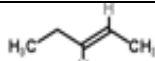
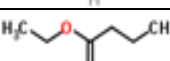

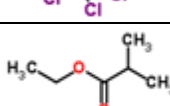
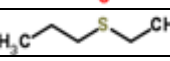
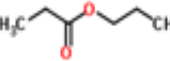
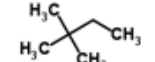
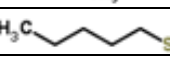
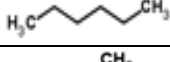
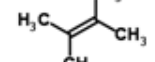
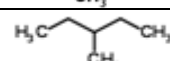

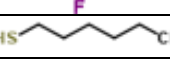
سری داده‌ها برای این تحقیق، ۳۳۳ ترکیب آلی شامل هیدروکربن، هالوکربن، الکل، اسید، کتون، آلدهید، استر، اتر، آمین، نیتریل، سولفید و تیو می‌باشد. مقادیر تجربی لگاریتم ثابت قانون هنری ( $\text{Log } K_H$ ) تمام ترکیبات در آب از مرجع [۶۷] جمع‌آوری شده است. مقادیر تجربی این ترکیبات در شرایط آزمایشگاهی یکسان بدست آمده است. ساختار و نحوه‌ی آرایش استخلاف‌ها در اطراف اسکلت اصلی مولکول، خواص هر ترکیب را تعیین می‌کند. لذا در این تحقیق سعی شده است ترکیبات متنوعی انتخاب گردند تا مدل ارائه شده بتواند ثابت قانون هنری دامنه‌ی وسیعی از ترکیبات را پیش‌بینی نماید. در جدول (۵-۱) نام، ساختار و مقادیر تجربی ثابت قانون هنری این ترکیبات آورده شده است.

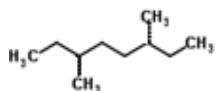


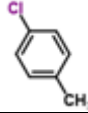

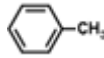
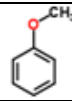


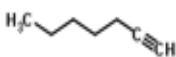
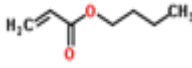
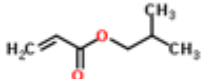

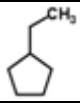
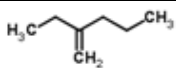
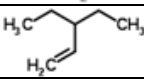

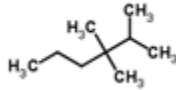
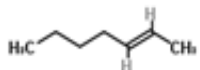
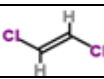
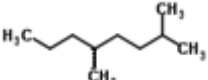
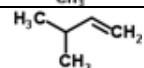
جدول (۵-۱) - سری ترکیبات مورد بررسی

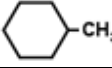
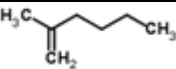
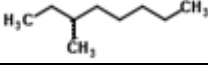
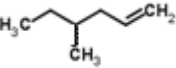
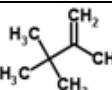
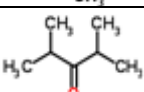

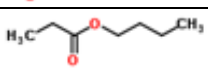
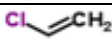
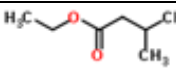
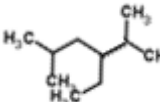
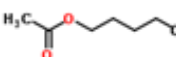
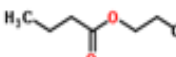
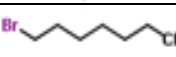
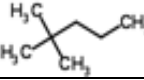
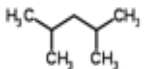

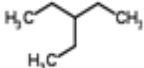
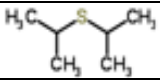
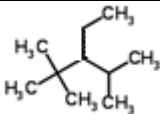
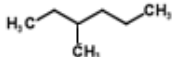
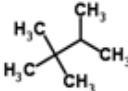

شماره	نام	ساختار	Log $K_H$
1	Tribromomethane		1/51
2	Methylchloride		2/66
3	Cis-2-pentene		4/10
4	Tert-butyl acetate		1/28
5	Valeraldehyde		1/26
6	2,5,5-Trimethylheptane		5/69

7	Isopropyl acetate		1/18
8	4-Methyl-1-pentene		4/54
9	Methyl n-butyrate		1/2۰
10	Isopentane		4/88
11	Methane		4/55
12	Neopentane		5/07
13	n-Pentane		4/85
14	2,3-Dimethylbutane		4/86
15	Butyl methyl sulfide		2/5۰
16	Isobutyl acetate		1/38
17	2-Methyl-2-butanethiol		2/74
18	1,2,4-Trichlorobenzene		2/22
19	m-Dibromobenzene		1/67
20	m-Dichlorobenzene		2/27
21	o-Dichlorobenzene		2/2۰
22	o-Difluorobenzene		2/59
23	p-Difluorobenzene		2/63
24	Bromobenzene		2/08
25	chlorobenzene		2/4۰
26	Methylmercaptan		2/03
27	Fluorobenzene		2/54
28	Iodobenzene		1/9۰
29	Benzene		2/49

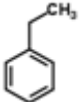
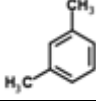
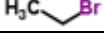


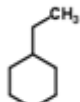
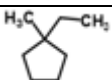

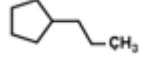
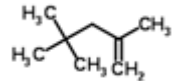
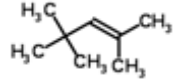
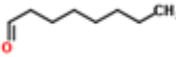

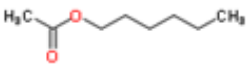
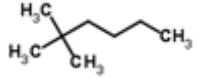
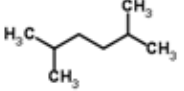
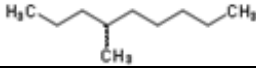
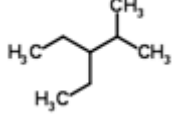

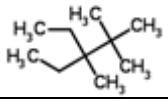
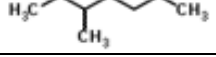
30	Chloropentafluoroethane		5/16
31	Cyclohexene		3/4·
32	2,3-Dimethyl-1,3-butadiene		3/83
33	1,5-Hexadiene		3/81
34	1-Hexyne		3/35
35	2-Hexyne		3/34
36	3-Hexyne		3/32
37	Cyclohexane		4/03
38	2,3-Dimethyl-1-butene		4/29
39	2-Methylpentane		4/98
40	3,3-Dimethyl-1-butene		4/28
41	1,2-Dichlorotetrafluoroethane		4/83
42	2-Ethyl-1-butene		4/3·
43	3-Isopropyl-2-methylhexane		5/86
44	3,3-Dimethylpentane		5/01
45	Cis-2-hexene		4/32
46	2,2,3-Trimethylpentane		5/36
47	Cis-3-hexene		4/31
48	3,3,4-Trimethylhexane		5/63
49	Methylcyclopentane		4/3·
50	2,2,3,4-Tetramethylhexane		5/8·

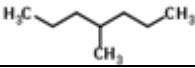

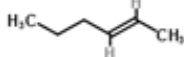
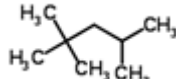
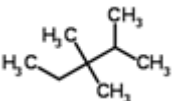
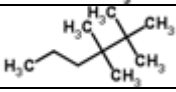
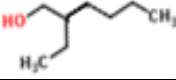

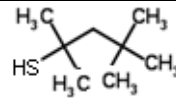
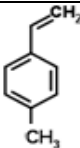
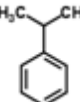
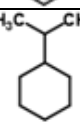

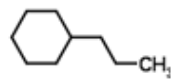

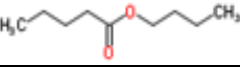
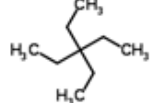
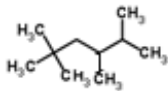
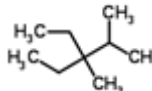
51	2,3,4,5-Tetramethylhexane		5/73
52	1,1,2,2-Tetrachlorodifluoroethane		3/84
53	3-Methyl-1-pentene		4/29
54	n-Propyl acetate		1/21
55	4-Methyl-cis-2-pentene		4/29
56	3-Ethyl-2,4-dimethylhexane		5/76
57	1-Hexanal		1/16
58	n-Butyl acetate		1/16
59	3-Ethyl-2,2-dimethylhexane		5/75
60	Trans-2-pentene		4/11
61	Ethyl n-butyrate		1/37
62	Hexachloroethane		3/15
63	Ethyl isobutyrate		1/36
64	Ethyl propyl sulfide		2/46
65	n-Propyl propionate		1/42
66	2,2-Dimethylbutane		4/93
67	n-Pentyl mercaptan		2/89
68	n-Hexane		4/86
69	2,3-Dimethyl-2-butene		4/33
70	3-Methylpentane		4/82
71	Tetrafluoroethylene		4/53
72	n-Hexyl mercaptan		3/09

73	3,6-Dimethyloctane		5/7.
74	n-Heptane		5/17
75	Dipropyl sulfide		2/66
76	p-Chlorotoluene		2/38
77	1,1-Dichloroethylene		3/1.
78	Toluene		2/55
79	Anisole		2/31
80	Dichlorofluoromethane		2/46
81	Cis-1,2-dichloroethylene		2/61
82	1-Heptyne		3/59
83	n-Butyl acrylate		1/5.
84	Isobutyl acrylate		1/58
85	1,1-Dimethylcyclopentane		4/38
86	Ethylcyclopentane		4/48
87	2-Ethyl-1-pentene		4/53
88	3-Ethyl-1-pentene		4/45
89	1-Heptene		4/35
90	2,3,3-Trimethylhexane		5/6.
91	Trans-2-heptene		4/55
92	Trans-1,2-dichloroethylene		2/57
93	2,5-Dimethyloctane		5/68
94	3-Methyl-1-butene		4/46

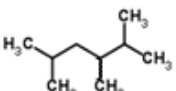


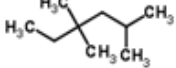
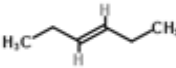
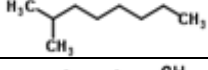
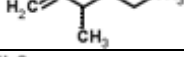
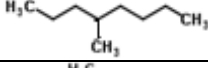
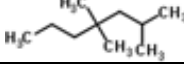
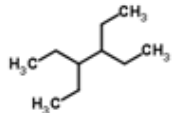
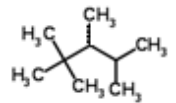
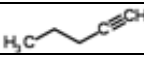
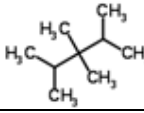
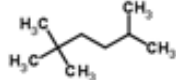
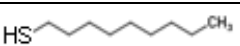
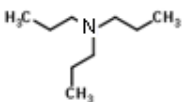
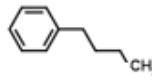
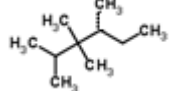
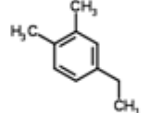
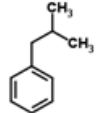
95	Methylcyclohexane		4/37
96	2-Methyl-1-hexene		4/5
97	3-Methyloctane		5/57
98	4-Methyl-1-hexene		4/47
99	2,3,3-Trimethyl-1-butene		4/45
100	2,4-dimethylpentan-3-one		1/29
101	Heptanal		1/29
102	n-Butyl propionate		1/59
103	Vinyl chloride		3/09
104	Ethyl isovalerate		1/55
105	3-Ethyl-2,5-dimethylhexane		5/71
106	n-Pentyl acetate		1/28
107	n-Propyl n-butyrate		1/54
108	1-Bromoheptane		3/4
109	2,2-Dimethylpentane		5/24
110	2,4-Dimethylpentane		5/21
111	1-Hexene		4/21
112	3-Ethylpentane		5/16
113	Diisopropyl sulfide		2/56
114	3-Ethyl-2,2,4-trimethylpentane		5/78
115	3-Methylhexane		5/23
116	2,2,3-Trimethylbutane		5/12
117	n-Heptyl mercaptan		3/22

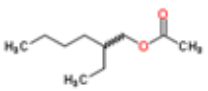
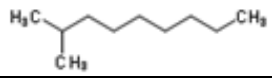
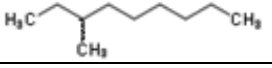
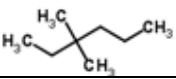
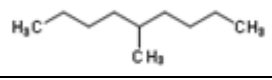
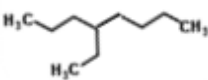
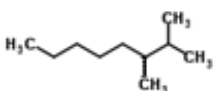
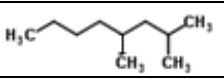
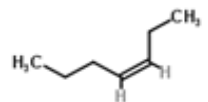
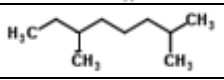
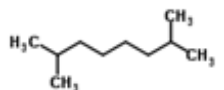
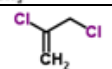
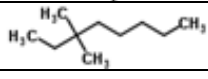
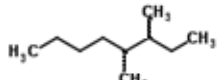
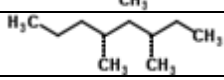

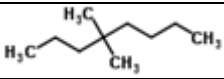
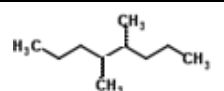
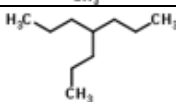
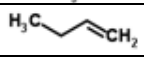
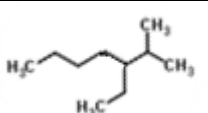


118	Ethylbenzene		2/66
119	m-Xylene		2/58
120	Bromoethane		2/61
121	p-Xylene		2/53
122	1,1-Dimethylcyclohexane		4/49
123	Ethylcyclohexane		4/56
124	1-Methyl-1-ethylcyclopentane		4/68
125	1-Octene		4/54
126	n-Propylcyclopentane		4/7
127	2,4,4-Trimethyl-1-pentene		4/66
128	2,4,4-Trimethyl-2-pentene		4/65
129	1-Octanal		1/47
130	Ethyl chloride		2/58
131	n-Hexyl acetate		1/44
132	2,2-Dimethylhexane		5/31
133	2,5-Dimethylhexane		5/31
134	4-Methylnonane		5/72
135	3-Ethyl-2-methylpentane		5/38
136	Trans-2-butene		4/11
137	3-Ethyl-2,2,3-trimethylpentane		5/92
138	3-Methylheptane		5/31

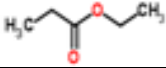
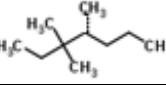
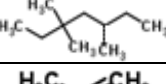


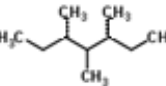
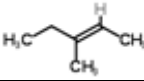
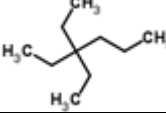
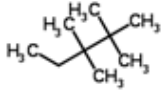
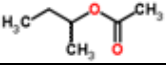
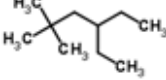
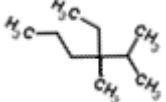
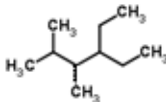
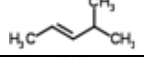
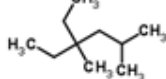
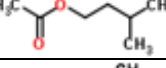
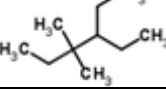
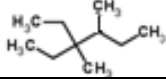
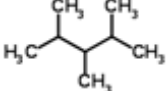
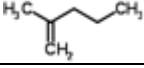

139	4-Methylheptane		5/36
140	Ethyl iodide		2/6
141	Trans-2-hexene		4/31
142	2,2,4-Trimethylpentane		5/28
143	2,3,3-Trimethylpentane		5/41
144	2,2,3,3-Tetramethylhexane		5/81
145	2-Ethyl-1-hexanol		1/11
146	n-Octyl mercaptan		3/26
147	Tert-octyl mercaptan		3/33
148	p-Methylstyrene		2/25
149	Ethane	$H_3C-CH_3$	4/43
150	Cumene		2/9
151	Isopropylcyclohexane		4/76
152	1-Nonene		4/65
153	n-Propylcyclohexane		4/76
154	1-Nonanal		1/56
155	n-Butyl valerate		1/55
156	3,3-Diethylpentane		5/69
157	2,2,4,5-Tetramethylhexane		5/7
158	3-Ethyl-2,3-dimethylpentane		5/67

159	2,4-Dimethyl-3-ethylpentane		5/59
160	2,2-Dimethylheptane		5/51
161	2,6-Dimethylheptane		5/52
162	2-Methyl-2-butene		4/2·
163	4-Ethylheptane		5/57
164	2,3-Dimethylheptane		5/58
165	2,3,5-Trimethylheptane		5/76
166	Sec-butyl mercaptan		2/57
167	3,4-Dimethylheptane		5/58
168	3,5-Dimethylheptane		5/53
169	Diethyl sulfide		2/09
170	3-Ethyl-2-methylhexane		5/56
171	2,2,4-Trimethylheptane		5/66
172	3-Ethyl-3-methylhexane		5/61
173	3-Ethyl-4-methylhexane		5/59
174	4-Ethyl-3-methylheptane		5/75
175	2,2,4-Trimethylhexane		5/5·
176	Cis-2-heptene		4/56
177	2,3,4-Trimethylhexane		5/6·

178	2,3,5-Trimethylhexane		5/52
179	Chloroform		2/36
180	Ethyl mercaptan		2/21
181	2,4,4-Trimethylhexane		5/56
182	Trans-3-hexene		4/31
183	2-Methyloctane		5/55
184	3-Methyl-1-hexene		4/45
185	4-Methyloctane		5/75
186	2,4,4-Trimethylheptane		5/70
187	3,4-Diethylhexane		5/77
188	2,2,3,4-Tetramethylpentane		5/59
189	1-Pentyne		3/14
190	2,3,3,4-Tetramethylpentane		5/66
191	2,2,5-Trimethylhexane		5/48
192	n-Nonyl mercaptan		3/17
193	Tripropylamine		1/33
194	n-Butylbenzene		2/86
195	2,3,3,4-Tetramethylhexane		5/84
196	4-Ethyl-o-xylene		2/64
197	Isobutylbenzene		2/64

198	2-Ethylhexyl acetate		1/64
199	2-Methylnonane		5/67
200	3-Methylnonane		5/71
201	3,3-Dimethylhexane		5/36
202	5-Methylnonane		5/69
203	4-Ethyloctane		5/71
204	2,3-Dimethyloctane		5/72
205	2,4-Dimethyloctane		5/66
206	Cis-3-heptene		4/54
207	2,6-Dimethyloctane		5/7.
208	2,7-Dimethyloctane		5/67
209	2,3-Dichloropropene		2/36
210	3,3-Dimethyloctane		5/73
211	3,4-Dimethyloctane		5/73
212	3,5-Dimethyloctane		5/7.
213	Butyl ethyl sulfide		2/67
214	4,4-Dimethyloctane		5/72
215	4,5-Dimethyloctane		5/73
216	4-Propylheptane		5/66
217	1-Butene		4/13
218	3-Ethyl-2-methylheptane		5/73


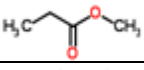

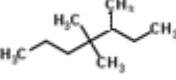
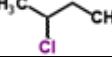

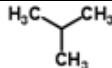

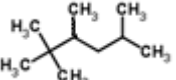
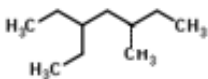
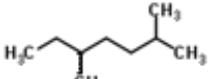
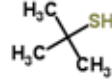
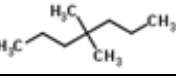
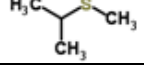


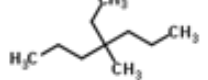
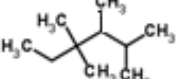

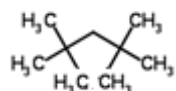
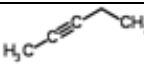
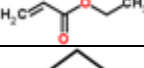

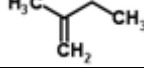
219	4-Ethyl-2-methylheptane		5/69
220	3-Bromo-1-propene		2/51
221	5-Ethyl-2-methylheptane		5/71
222	3-Ethyl-3-methylheptane		5/78
223	2,2,3-Trimethylhexane		5/56
224	Isobutyl mercaptan		2/58
225	3-Ethyl-4-methylheptane		5/75
226	Isoprene		3/71
227	2,2,3-Trimethylheptane		5/73
228	4-Ethyl-2-methylhexane		5/52
229	2,2,5-Trimethylheptane		5/68
230	2,2,6-Trimethylheptane		5/64
231	3-Chloropropene		2/71
232	2,3,3-Trimethylheptane		5/76
233	2,3,4-Trimethylheptane		5/74
234	2,4-Dimethylheptane		5/50
235	2,3,6-Trimethylheptane		5/69
236	n-Nonane		5/53
237	2,4,5-Trimethylheptane		5/71
238	Cis-2-butene		4/12

239	Ethyl propionate		1/09
240	3,3,4-Trimethylheptane		5/78
241	3,3,5-Trimethylheptane		5/74
242	Propylene		4/05
243	n-Butyl chloride		2/97
244	3,4,5-Trimethylheptane		5/77
245	Trans-3-methyl-2-pentene		4/32
246	3,3-Diethylhexane		5/81
247	2,2,3,3-Tetramethylpentane		5/65
248	Sec-butyl acetate		1/48
249	4-Ethyl-2,2-dimethylhexane		5/66
250	3-Ethyl-2,3-dimethylhexane		5/81
251	4-Ethyl-2,3-dimethylhexane		5/76
252	4-Methyl-trans-2-pentene		4/29
253	4-Ethyl-2,4-dimethylhexane		5/78
254	Isopentyl acetate		1/42
255	4-Ethyl-3,3-dimethylhexane		5/81
256	3-Ethyl-3,4-dimethylhexane		5/8
257	2,3,4-Trimethylpentane		4/99
258	2-Methyl-1-pentene		4/19
259	n-Butyl mercaptan		2/7

260	2,2,4,4-Tetramethylhexane		5/82
261	2,2-Dimethyl-3-ethylpentane		5/56
262	2,2,5,5-Tetramethylhexane		5/59
263	4-Ethyl-m-xylene		2/68
264	2,3,3,5-Tetramethylhexane		5/74
265	1,4-Pentadiene		3/82
266	2-Methyl-2-pentene		4/31
267	3,3,4,4-Tetramethylhexane		5/91
268	1-Tridecene		4/96
269	3,3-Diethyl-2-methylpentane		5/86
270	2-Methylheptane		5/36
271	2-Methylhexane		5/28
272	3-Ethyl-2,3,4-trimethylpentane		5/89
273	1-Bromopropane		2/71
274	2,2,3,3,4-Pentamethylpentane		5/91
275	2,2,3,4,4-Pentamethylpentane		5/88
276	n-Decyl mercaptan		2/92
277	n-Pentylbenzene		3/09
278	2-Ethylhexyl acrylate		1/38
279	1-Undecene		5/07
280	n-Undecane		5/03



281	Undecyl mercaptan		2/2۰
282	Bromochloromethane		1/95
283	2-Bromopropane		2/83
284	n-Hexylbenzene		3/15
285	1-Dodecene		5/03
286	n-Dodecyl mercaptan		2/59
287	n-Heptylbenzene		2/57
288	2,4-Dimethyl-3-isopropylpentane		5/78
289	n-Tetradecane		4/88
290	n-Propyl chloride		2/86
291	1-Pentadecene		4/18
292	Isopropyl iodide		2/8۰
293	n-Propyl iodide		2/7۰
294	Propane		4/58
295	n-Propyl mercaptan		2/4۰
296	Isopropyl mercaptan		2/37
297	Ethyl methyl sulfide		2/04
298	Dibromomethane		1/71
299	Hexachloro-1,3-butadiene		3/12
300	Furan		2/48
301	Thiophene		2/2۰
302	1,3-Butadiene		3/6
303	4-Isopropylheptane		5/72
304	2,4,6-Trimethylheptane		5/62
305	3-Methyl-3-ethylpentane		5/42
306	Isobutene		4/06

307	Chlorofluoromethane		2/54
308	Methyl propionate		1/02
309	1-Bromobutane		2/83
310	3,4,4-Trimethylheptane		5/78
311	sec-Butyl chloride		3/03
312	n-Butane		4/71
313	Isobutane		4/81
314	Dichloromethane		2/14
315	2,2,3,5-Tetramethylhexane		5/68
316	3-Ethyl-5-methylheptane		5/69
317	2,5-Dimethylheptane		5/53
318	Tert-butyl mercaptan		2/47
319	4,4-Dimethylheptane		5/56
320	Isopropyl methyl sulfide		2/2
321	Diiodomethane		1/28
322	Cyclopentene		3/55
323	4-Ethyl-4-methylheptane		5/76
324	2,3,4,4-Tetramethylhexane		5/82
325	Methylbromide		2/58
326	2,2,4,4-Tetramethylpentane		5/52
327	2-Pentyne		3/15
328	Ethyl acrylate		1/14
329	Cyclopentane		4/02
330	2-Methyl-1-butene		4/21

ادامه‌ی جدول (۵-۱)

331	3-Ethylheptane		5/58
332	Trans-3-heptene		4/53
333	1-Pentene		4/35

## ۵-۱-۲- مراحل مدل سازی

### ۵-۱-۲-۱- رسم و بهینه‌سازی ساختار مولکول‌ها

اولین مرحله جهت مدل‌سازی پس از انتخاب سری داده‌ها، رسم و بهینه‌سازی ساختار مولکول‌های مورد بررسی می‌باشد. بدین منظور ساختار شیمیایی تمام ترکیبات با استفاده از نرم‌افزار Hyperchem رسم و سپس با استفاده از روش نیمه تجربی AM1 بهینه گردید. لازم به ذکر است که بهینه‌سازی تا حد گرادیانی ۰/۰۰۱ کیلوکالری بر آنگستروم مول ادامه یافت.

### ۵-۲-۱-۲- محاسبه‌ی توصیف‌کننده‌های مولکولی

توصیف‌کننده‌های مولکولی مقادیر عددی هستند که ویژگی‌های مختلف ساختاری یا شکل مولکولی ترکیبات را توصیف می‌کنند. گروه‌های مختلفی از این توصیف‌کننده‌ها تاکنون معرفی شده‌اند که هر کدام جنبه‌های خاصی از ترکیبات را بررسی و توصیف می‌کنند. ساختارهای سه بعدی رسم شده از مرحله بالا به عنوان ورودی به نرم‌افزار Dragon وارد و بوسیله‌ی آن تعداد زیادی از این توصیف‌کننده‌ها برای ترکیبات محاسبه شد که برای هر ترکیب ۱۲۶۳ توصیف‌کننده‌ی مختلف بدست آمد. در هنگام محاسبه‌ی توصیف‌کننده‌ها با نرم‌افزار Dragon می‌توان توصیف‌کننده‌های با مقدار ثابت درکل ترکیبات را حذف نمود.

### ۵-۱-۲-۳- انتخاب توصیف‌کننده‌های مهم

همانطور که بیان شد تعداد توصیف‌کننده‌های محاسبه شده با استفاده از نرم افزار Dragon بسیار زیاد می‌باشد اما تمام این توصیف‌کننده‌ها با ثابت قانون هنری ارتباط ندارند. در ابتدا برای انتخاب توصیف‌کننده‌های مهم، تمام توصیف‌کننده‌هایی که برای حداقل ۹۰ درصد از ترکیبات مورد بررسی مقادیر یکسان و صفر داشتند از فهرست توصیف‌کننده‌ها حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیف‌کننده‌ای که ضریب همبستگی بزرگتر از  $0.9$  ( $r^2 > 0.9$ ) داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف شد زیرا متغیرهای وارد شده در مدل نباید با یکدیگر همبستگی داشته باشند. این کار با استفاده از نرم‌افزار SPSS برای هر گروه از توصیف‌کننده‌ها انجام گردید.

### ۵-۱-۲-۴- مدل رگرسیون خطی چندگانه

برای انتخاب توصیف‌کننده‌های مهم که بتوانند ارتباط ساختاری ترکیبات مورد بررسی را با ثابت قانون هنری آنها بیان کنند، از روش رگرسیون مرحله‌ای توسط نرم افزار SPSS استفاده گردید. در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل می‌شوند و ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته دارد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معناداری خود را از دست داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. به همین منظور، داده‌های ثابت قانون هنری به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیر مستقل در نظر گرفته شدند. با توجه به خصوصیات آماری مدل‌های مختلف بدست آمده، بهترین مدل که دارای بیشترین مقدار  $R, F$  و کمترین مقدار خطای استاندارد و شامل توصیف‌کننده‌های تاحد امکان قابل توجیه باشد، به عنوان مدل نهایی برای پیش‌بینی ثابت‌های قانون هنری ترکیبات انتخاب گردید. در این روش داده‌ها به‌طور تصادفی به سری آموزشی، آزمون و اعتبار تقسیم می‌شوند. سپس مدل‌های بدست آمده از سری آموزشی با سری آزمون

آزموده شده و بهترین مدل انتخاب می‌شود. در نهایت قدرت پیش‌بینی مدل برگزیده با سری اعتبار ارزیابی می‌گردد. توصیف‌کننده‌های انتخاب شده جهت ایجاد مدل MLR به همراه پارامترهای آماری مربوطه با استفاده از روش اجباری در جدول (۲-۵) آمده است.

جدول (۲-۵) - مشخصات مدل حاصل از رگرسیون مرحله‌ای

نماد توصیف‌کننده	مفهوم	ضریب رگرسیون	اثر متوسط
MLOGP	ضریب توزیع آب-اکتانول موری گوچی <sup>۱</sup>	۰/۹۲۹	<b>0/758</b>
GATS1e	خود ارتباط گری مرتبه اول/وزن دار شده با الکترونگاتیوی اتمی ساندرسون <sup>۲</sup>	۰/۶۳۴	<b>0/407</b>
XMOD	شاخص راندیک کای اصلاح شده <sup>۳</sup>	- ۰/۰۹۱	-0/517
Jhete	شاخص بالابان از نوع ماتریس فاصله وزن دار شده الکترونگاتیویته <sup>۴</sup>	۰/۱۶۱	0/1۲۰
		۰/۹۶۰	ضریب همبستگی
		۷۶۸/۶۳۰	مقدار آماره‌ی F
		۰/۴۳۷	خطای استاندارد
		۰/۱۸۸	میانگین مربعات خطا

معادله‌ی خطی حاصل از این مدل نیز به صورت رابطه‌ی (۱-۵) می‌باشد:

$$\text{Log } K_H = 0.972 + 0.929 \text{ MLOGP} + 0.634 \text{ GATS1e} - 0.091 \text{ XMOD} + 0.161 \text{ Jhete} \quad (1-5)$$

همچنین با توجه به ماتریس همبستگی در جدول (۳-۵)، توصیف‌کننده‌های وارد شده در این مدل دارای ضریب همبستگی برابر یا کمتر از ۰/۷ با یکدیگر می‌باشند.

جدول (۳-۵) - ماتریس همبستگی توصیف‌کننده‌های وارد شده در مدل منتخب

	MLOGP	GATS1e	XMOD	Jhete
MLOGP	۱/۰۰۰			
GATS1e	<b>0/700</b>	۱/۰۰۰		
XMOD	<b>0/551</b>	<b>0/02۱</b>	۱/۰۰۰	
Jhete	<b>0/44۸</b>	<b>0/321</b>	<b>0/332</b>	۱/۰۰۰

<sup>۱</sup>- Moriguchi octanol-water partition coefficient

<sup>۲</sup>- Geary autocorrelation-lag1/weighted by atomic sanderson electronegativities

<sup>۳</sup>- Modified Randic chi-1-index

<sup>۴</sup>- Blaban-type index from electronegativity weighted distance matrix

مقادیر توصیف‌کننده‌های وارد شده در مدل برای تمام ترکیبات در جدول (۴-۵) آورده شده

است.

جدول (۴-۵) - مقادیر توصیف‌کننده‌های وارد شده در مدل

شماره	MLOGP	GATS1e	Jhete	XMOD	شماره	MLOGP	GATS1e	Jhete	XMOD
1	2/419	1/854	2/724	35/507	33t	2/352	2/667	2/656	17/485
2	0/845	1/024	1/265	11/500	34	2/459	2/667	2/544	17/485
3t	2/081	3/000	2/622	14/485	35t	2/459	2/667	2/787	17/485
4	1/373	1/426	4/193	21/838	36	2/459	2/667	2/899	17/485
5	1/064	1/024	2/529	18/192	37	3/516	2/833	2/000	18/000
6	4/823	3/200	3/465	26/863	38	2/459	3/000	3/258	15/856
7t	0/996	1/441	3/742	20/149	39	3/516	3/333	2/627	16/620
8	2/459	3/000	2/814	16/620	40t	2/459	3/000	3/446	15/364
9	0/996	1/441	3/329	21/541	41	2/604	3/478	4/933	28/000
10t	3/138	3/400	2/540	13/620	42	2/459	3/000	2/988	16/848
11v	1/115	5/000	0/000	0/000	43	4/823	3/200	3/728	27/549
12	3/138	3/400	3/024	12/000	44	3/869	3/286	3/360	18/728
13	3/138	3/400	2/191	14/485	45	2/459	3/000	2/658	17/485
14	3/516	3/333	2/993	15/856	46	4/203	3/250	3/623	20/888
15t	2/226	2/073	2/400	23/521	47t	2/459	3/000	2/734	17/485
16	1/373	1/426	3/513	23/241	48	4/520	3/222	3/802	24/252
17	2/226	2/434	3/208	17/864	49v	3/516	2/833	2/184	17/363
18v	4/063	1/077	3/171	34/717	50	4/823	3/200	3/942	26/352
19	3/795	0/932	2/989	39/469	51	4/823	3/200	3/814	26/785
20t	3/478	0/966	3/034	29/077	52	2/957	3/308	4/798	32/000
21	3/478	0/966	3/077	29/178	53v	2/459	3/000	2/965	16/848
22t	3/144	1/010	3/157	24/559	54	0/996	1/441	3/317	21/106
23	3/144	1/010	3/072	24/458	55t	2/459	3/000	3/032	16/620
24	3/045	0/937	2/882	28/735	56	4/823	3/200	3/798	27/777
25	2/876	0/917	2/909	23/539	57	1/442	1/044	2/589	21/192
26	0/203	2/221	1/077	11/000	58	1/373	1/426	3/217	24/106
27v	2/702	0/927	2/955	21/229	59	4/823	3/200	3/809	27/116
28	3/210	1/477	2/826	33/931	60v	2/081	3/000	2/622	14/485
29v	2/255	1/833	2/769	18/000	61v	1/373	1/426	3/439	24/334
30t	2/419	3/697	5/004	26/000	62	3/291	4/000	4/670	36/000
31	2/459	2/500	2/255	18/000	63	1/373	1/426	3/783	23/570
32v	2/352	2/667	3/568	15/856	64	2/226	2/073	2/430	22/485

ادامه‌ی جدول (۴-۵)

شماره	MLOGP	GATS1e	Jhete	XMOD	شماره	MLOGP	GATS1e	Jhete	XMOD
65t	1/373	1/426	3/436	24/334	100t	1/795	1/065	3/704	21/898
66	3/516	3/333	3/168	15/364	101	1/795	1/065	2/639	24/192
67	2/226	2/434	2/359	21/021	102	1/727	1/420	3/355	27/334
68t	3/516	3/333	2/339	17/485	103	1/219	1/020	2/537	12/374

69	2/459	3/000	3/676	15/856	104	1/727	1/420	3/607	26/469
70	3/516	3/333	2/754	16/848	105	4/823	3/200	3/603	27/448
71	1/672	3/000	4/848	19/321	106v	1/727	1/420	3/146	27/106
72t	2/604	2/475	2/463	24/021	107	1/727	1/420	3/438	27/334
73	4/823	3/200	3/168	28/211	108	3/451	1/339	2/555	33/738
74	3/869	3/286	2/447	20/485	109	3/869	3/286	3/154	18/364
75t	2/604	2/144	2/533	25/485	110t	3/869	3/286	3/144	19/084
76v	3/210	0/944	2/927	25/902	111	2/459	3/000	2/488	17/485
77	1/672	1/248	3/359	16/743	112v	3/869	3/286	2/992	20/076
78	2/608	2/000	2/821	20/363	113	2/604	2/144	3/078	22/838
79	1/859	1/711	2/911	24/706	114	4/823	3/200	4/073	26/352
80	1/597	1/969	3/076	17/609	115t	3/869	3/286	2/832	19/848
81	1/672	1/248	2/994	19/263	116	3/869	3/286	3/144	19/084
82	2/812	2/714	2/599	20/485	117	2/957	2/511	2/542	27/021
83t	1/619	1/426	3/472	27/334	118	2/942	2/125	2/682	23/591
84	1/619	1/426	3/738	26/469	119	2/942	2/125	2/890	22/726
85	3/869	2/857	2/400	19/243	120t	1/597	1/033	1/763	18/738
86	3/869	2/857	2/140	20/591	121	2/942	2/125	2/859	22/726
87t	2/812	3/000	3/003	19/848	122	4/203	2/875	2/328	22/243
88	4/203	3/250	3/355	22/312	123t	4/203	2/875	2/125	23/591
89	2/812	3/000	2/559	20/485	124	4/203	2/875	2/400	22/607
90	4/520	3/222	3/702	24/024	125t	3/146	3/000	2/616	23/485
91	2/812	3/000	2/777	20/485	126	4/203	2/875	2/058	23/591
92	1/672	1/248	2/994	19/263	127	3/146	3/000	3/551	20/499
93	4/823	3/200	3/124	27/983	128	3/146	3/000	3/886	20/499
94	2/081	3/000	2/826	13/620	129	2/129	1/087	2/682	27/192
95	3/869	2/857	2/123	20/363	130t	1/364	1/000	1/825	12/374
96v	2/812	3/000	2/822	19/620	131	2/060	1/418	3/097	30/106
97	4/520	3/222	2/877	25/848	132	4/203	3/250	3/112	21/364
98	2/812	3/000	2/980	19/848	133	4/203	3/250	2/928	21/755
99t	2/812	3/000	3/791	17/660	134	4/823	3/200	2/968	28/848

ادامه‌ی جدول (۴-۵)

شماره	MLOGP	GATS1e	Jhete	XMOD	شماره	MLOGP	GATS1e	Jhete	XMOD
135	4/203	3/250	3/355	22/312	170t	4/520	3/222	3/410	25/312
136t	1/672	3/000	2/549	11/485	171	4/823	3/200	3/469	26/727
137	4/823	3/200	4/328	26/228	172	4/520	3/222	3/617	25/092
138	4/203	3/250	2/862	22/848	173	4/520	3/222	3/499	25/540
139t	4/203	3/250	2/920	22/848	174	4/823	3/200	3/564	28/540
140v	1/817	2/321	1/642	25/102	175	4/520	3/222	3/467	23/727
141	2/459	3/000	2/658	17/485	176	2/812	3/000	2/689	20/485
142	4/203	3/250	3/389	20/499	177	4/520	3/222	3/576	24/549
143	4/203	3/250	3/708	21/024	178	4/520	3/222	3/377	24/219
144	4/823	3/200	4/102	25/864	179	1/817	2/044	2/941	19/919
145	2/274	2/098	3/142	26/783	180t	0/845	2/274	1/694	12/021

146	3/911	2/595	2/697	36/021	181	4/520	3/222	3/577	23/863
147v	3/291	2/542	3/088	26/864	182	2/459	3/000	2/734	17/485
148	3/169	2/000	2/862	25/954	183	4/520	3/222	2/747	25/620
149	1/757	4/000	1/000	6/000	184	2/812	3/000	2/984	19/848
150t	3/259	2/222	2/720	25/827	185	4/520	3/222	2/955	25/848
151	4/520	2/889	2/228	25/827	186	4/823	3/200	3/626	26/863
152	3/463	3/000	2/664	26/485	187	4/823	3/200	3/698	28/769
153v	4/520	2/889	2/078	26/591	188	4/520	3/222	3/878	23/124
154	2/446	1/110	2/718	30/192	189	2/081	2/600	2/484	14/485
155	1/727	1/420	3/358	27/334	190t	4/520	3/222	4/014	23/321
156v	4/520	3/222	3/825	25/456	191v	4/520	3/222	3/281	23/499
157	4/823	3/200	3/684	25/963	192	3/608	2/570	2/655	33/021
158	4/520	3/222	3/919	24/388	193	1/940	2/419	3/352	23/689
159	4/520	3/222	3/678	24/549	194	3/562	2/300	2/346	29/591
160t	3/869	3/286	3/154	18/364	195	4/823	3/200	4/089	26/549
161	4/520	3/222	2/915	24/755	196	3/562	2/300	2/908	28/418
162v	2/081	3/000	3/143	13/620	197	3/562	2/300	2/496	28/726
163v	4/520	3/222	3/175	26/076	198	2/681	1/422	3/720	35/697
164	4/520	3/222	3/155	25/084	199	4/823	3/200	2/773	28/620
165	4/823	3/200	3/462	27/448	200	4/823	3/200	2/886	28/848
166	1/817	2/386	2/578	16/507	201	4/203	3/250	3/373	21/728
167	4/520	3/222	3/325	25/312	202	4/823	3/200	2/998	28/848
168	4/520	3/222	3/223	25/211	203	4/823	3/200	3/206	29/076
169	1/817	1/990	2/298	19/485	204	4/823	3/200	3/130	28/084

ادامه‌ی جدول (۴-۵)

شماره	MLOGP	GATS1e	Jhete	XMOD	شماره	MLOGP	GATS1e	Jhete	XMOD
205t	4/823	3/200	3/160	27/983	240	4/823	3/200	3/778	27/252
206	2/812	3/000	2/777	20/485	241	4/823	3/200	3/642	27/091
207	4/823	3/200	3/033	27/983	242	1/219	3/000	2/187	8/485
208	4/823	3/200	2/909	27/755	243	2/226	1/043	2/274	18/374
209	2/081	1/073	3/147	20/685	244	4/823	3/200	3/685	27/777
210v	4/823	3/200	3/277	27/728	245	2/459	3/000	3/219	16/848
211	4/823	3/200	3/309	28/312	246	4/823	3/200	3/875	28/456
212	4/823	3/200	3/269	28/211	247	4/520	3/222	4/145	22/864
213	2/604	2/144	2/522	25/485	248	1/373	1/426	3/765	23/377
214v	4/823	3/200	3/418	27/728	249	4/823	3/200	3/631	26/955
215	4/823	3/200	3/376	28/312	250	4/823	3/200	3/944	27/388
216	4/823	3/200	3/295	29/076	251	4/823	3/200	3/756	27/777
217	1/672	3/000	2/297	11/485	252	2/459	3/000	3/032	16/620
218	4/823	3/200	3/398	28/312	253	4/823	3/200	3/803	27/227
219	4/823	3/200	3/391	28/211	254v	1/727	1/420	3/363	26/241
220t	1/881	1/029	2/408	21/738	255	4/823	3/200	3/971	27/480
221	4/823	3/200	3/256	28/211	256	4/823	3/200	4/020	27/616
222	4/823	3/200	3/576	28/092	257	4/203	3/250	3/464	21/321



223	4/520	3/222	3/589	23/888	258	2/459	3/000	2/831	16/620
224	1/817	2/386	2/576	17/156	259	1/817	2/386	2/218	18/021
225	4/823	3/200	3/530	28/540	260v	4/823	3/200	3/888	25/607
226	1/974	2/600	3/205	13/620	261	4/520	3/222	3/793	24/116
227	4/823	3/200	3/518	26/888	262	4/823	3/200	3/563	25/243
228	4/520	3/222	3/307	25/211	263	3/562	2/300	2/945	28/418
229v	4/823	3/200	3/356	26/727	264	4/823	3/200	3/866	26/159
230t	4/823	3/200	3/205	26/499	265	1/974	2/600	2/653	14/485
231	1/672	1/000	2/460	15/374	266	2/459	3/000	3/137	16/620
232	4/823	3/200	3/633	27/024	267	4/823	3/200	4/282	26/228
233	4/823	3/200	3/583	27/549	268	4/611	3/000	2/792	38/485
234	4/520	3/222	3/151	24/983	269	4/823	3/200	4/154	27/752
235	4/823	3/200	3/301	27/219	270	4/203	3/250	2/716	22/620
236	4/520	3/222	2/595	26/485	271	3/869	3/286	2/678	19/620
237	4/823	3/200	3/503	27/448	272	4/823	3/200	4/229	26/684
238	1/672	3/000	2/549	11/485	273t	2/026	1/093	2/059	21/738
239	0/996	1/441	3/500	21/334	274	4/823	3/200	4/404	25/160

ادامه‌ی جدول (۴-۵)

شماره	MLOGP	GATS1e	Jhete	XMOD	شماره	MLOGP	GATS1e	Jhete	XMOD
275	4/823	3/200	4/231	24/928	305t	4/203	3/250	3/583	22/092
276	3/911	2/595	2/697	36/021	306	1/672	3/000	2/803	10/392
277	3/854	2/364	2/215	32/591	307	1/115	1/353	2/212	13/435
278	2/865	1/421	3/874	38/925	308	0/586	1/468	3/459	18/541
279	4/058	3/000	2/738	32/485	309	2/419	1/158	2/248	24/738
280v	5/115	3/182	2/691	32/485	310	4/823	3/200	3/823	27/252
281	4/203	2/617	2/733	39/021	311v	2/226	1/043	2/658	16/796
282	1/597	1/249	1/988	22/627	312	2/729	3/500	1/975	11/485
283	2/026	1/093	2/444	18/764	313v	2/729	3/500	2/324	10/392
284	4/135	2/417	2/110	35/591	314	1/364	1/313	2/067	16/263
285	4/339	3/000	2/767	35/485	315	4/823	3/200	3/735	26/023
286	4/484	2/637	2/763	42/021	316	4/823	3/200	3/412	28/439
287t	4/407	2/462	2/028	38/591	317	4/520	3/222	3/061	24/983
288	4/823	3/200	3/984	26/785	318	1/817	2/386	3/079	14/500
289	5/931	3/143	2/785	41/485	319	4/520	3/222	3/431	24/728
290t	1/817	1/016	2/097	15/374	320	1/817	1/990	2/649	19/197
291	5/130	3/000	2/834	44/485	321	2/226	1/317	1/652	41/719
292	2/226	2/472	2/333	23/960	322	2/081	2/400	2/324	15/000
293	2/226	2/472	1/981	28/102	323	4/823	3/200	3/690	28/092
294	2/276	3/667	1/633	8/485	324	4/823	3/200	4/034	26/488
295	1/364	2/332	2/014	15/021	325	1/115	1/000	1/172	20/500
296	1/364	2/332	2/381	13/279	326	4/520	3/222	3/746	22/243
297	1/364	1/892	2/082	17/521	327v	2/081	2/600	2/809	14/485
298	1/817	1/197	1/914	28/991	328	0/888	1/467	3/728	21/334
299v	3/659	2/500	5/218	45/837	329	3/138	2/800	2/083	15/000

300	0/156	1/704	2/975	16/000	330t	2/081	3/000	2/847	13/620
301	1/173	1/349	2/703	20/000	331	4/520	3/222	3/092	26/076
302	1/565	2/500	2/732	11/485	332	2/812	3/000	2/777	20/485
303	4/823	3/200	3/500	28/312	333	2/081	3/000	2/402	14/485
304	4/823	3/200	3/337	27/118					

t = سری آزمون

v = سری ارزیابی

### ۵-۱-۲-۵- پیش‌بینی ثابت هنری با شبکه عصبی مصنوعی

در این تحقیق برای پیش‌بینی ثابت هنری از مدل غیر خطی نیز استفاده شده است. برای ایجاد مدل غیرخطی که ارتباط مناسبی میان ساختار و ثابت هنری ترکیبات مورد نظر فراهم کند، از شبکه عصبی مصنوعی استفاده شد. شبکه‌ی انتخاب شده، یک شبکه عصبی چند لایه پیشخور با الگوی آموزش پس‌انتشار می‌باشد. برای آموزش موفق و در نتیجه قدرت پیش‌بینی شبکه باید عوامل موثر در آموزش شبکه عصبی مصنوعی بررسی و بهینه شود. الگوریتم‌های آموزشی مختلفی در جعبه ابزار نرم‌افزار مطلب وجود دارد که در این تحقیق الگوریتم آموزشی لونیبرگ - مارکورت (trainlm) مورد استفاده قرار گرفت. پارامترهای قابل بررسی در این الگوریتم آموزشی شامل تعداد دوره‌های چرخش (epochs)، مقدار اولیه‌ی ممنتوم ( $\mu$ ) در محاسبه‌ی تغییرات وزن، کمترین مقدار برای گرادیان خطا (mingrad)، کاهش دهنده مقدار  $\mu$  ( $\mu$ -dec)، مقدار ماکزیمم  $\mu$  ( $\mu$ -max)، افزایش دهنده مقدار  $\mu$  ( $\mu$ -inc) و سرعت یادگیری ( $lr$ ) می‌باشند.

پارامترهای موثر و بهینه شده عبارتند از:

۱- تعداد نرون‌ها در لایه پنهان

۲- نوع تابع انتقال

۲- تعداد دوره‌های آموزش

پارامترهای ممنوم و وزن نیازی به بهینه کردن ندارند زیرا مقدار آنها با اجرای شبکه توسط تابع یادگیری گرادیان نزولی ممنوم (learnngdm) تعیین می‌شود و به دلیل بالا بودن سرعت آموزش الگوریتم لوبنبرگ - مارکورت نیازی به بهینه کردن آنها در حین آموزش شبکه نیست. با توجه به اینکه خروجی‌های شبکه عصبی داده‌های کمی بوده و مقادیر لایه‌های داخلی بوسیله‌ی توابع لگاریتمی محدود شده‌اند، برای لایه‌ی خروجی تابع انتقال خطی انتخاب می‌گردد. همانطور که گفته شد سری داده‌ها به‌طور تصادفی به سری آموزش، آزمون و اعتبار تقسیم شدند. معیار در نظر گرفته شده در بهینه‌سازی به حداقل رساندن مجذور میانگین مربعات خطا می‌باشد.

### ۵-۱-۲-۱-۵- انتخاب تعداد لایه‌های شبکه عصبی

نتایج حاصل از مطالعات مختلف نشان می‌دهد که در بکارگیری شبکه عصبی در شیمی، انتخاب یک لایه پنهان نتایج رضایت‌بخشی را ایجاد می‌کند [۶۸]. لذا در این تحقیق شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی بکار گرفته شد.

### ۵-۱-۲-۲-۵- انتخاب تابع انتقال

برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان شبکه عصبی، از توابع لگاریتم سیگموئید و تانژانت سیگموئید به عنوان تابع فعالیت استفاده گردید. در روند بهینه‌سازی شبکه، هدف کاهش مقدار مجذور میانگین مربعات خطا می‌باشد. در تمام شبکه‌ها تعداد دوره‌های آموزش یکسان و برابر ۲۰ در نظر گرفته شد. با توجه به جداول (۵-۵) و (۶-۵) و نمودارهای (۱-۵) و (۲-۵) تابع تبدیل لگاریتم سیگموئید نسبت به تانژانت سیگموئید دارای کمترین مقدار مجذور میانگین مربعات خطا می‌باشد.

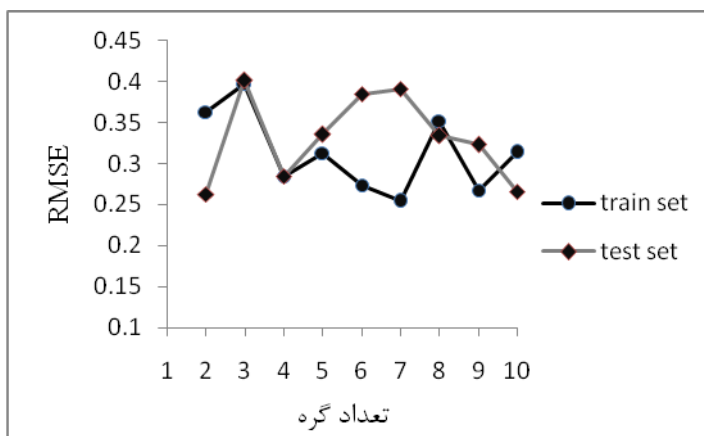
جدول (۵-۵) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان

تعداد گره‌ها
--------------

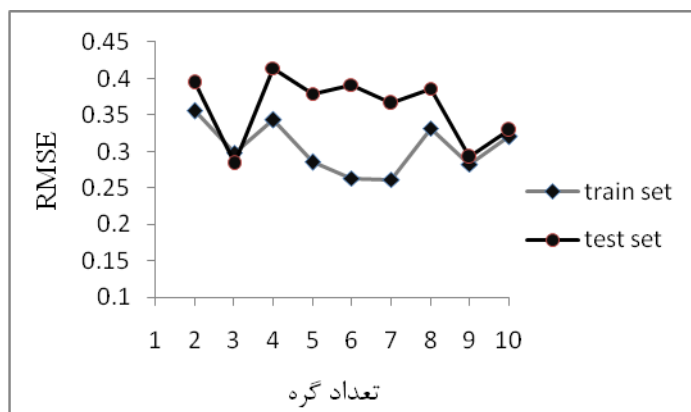
	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
سری آموزش	0/363 1	0/397 2	<b>0/284</b> 1	0/312 2	0/274 0	0/254 0	0/352 1	0/267 6	0/313 5
سری آزمون	0/262 6	0/401 2	<b>0/284</b> 6	0/335 1	0/384 6	0/391 1	0/334 2	0/323 0	0/265 8

جدول (۵-۶) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل تانژانت سیگموئید و تعداد گره‌های مختلف در لایه پنهان

		تعداد گره‌ها								
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
سری آموزش	ش	0/355 8	0/298 5	0/343 5	0/285 1	0/262 9	0/261 7	0/332 2	0/282 9	0/320 6
سری آزمون	ش	0/394 3	0/283 2	0/414 0	0/379 2	0/391 3	0/365 8	0/385 9	0/292 5	0/329 3



شکل (۵-۱) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل لگاریتم سیگموئید



شکل (۲-۵) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل تانژانت سیگموئید

### ۵-۱-۲-۳-۵- انتخاب تعداد گره‌های لایه پنهان و تعداد دوره‌های آموزش

در این مرحله از بهینه‌سازی، تعداد گره‌های لایه پنهان از ۲ تا ۸ در تعداد دوره‌های آموزش مختلف تغییر داده شدند. در نهایت از بین شبکه‌های آموزش دیده، با توجه به مجذور میانگین مربعات خطا برای سری‌های آموزش و آزمون، مدل نهایی دارای چهار نرون در لایه پنهان و نه دور آموزشی می‌باشد که نتایج حاصله نیز در جدول (۷-۵) آمده است. در تعداد گره‌های بالا با وجود خطای کمتر برای سری‌های آموزش و آزمون، افزایش خطا برای سری ارزیابی را داریم، بنابراین از تعداد گره‌های کمتر استفاده شد.

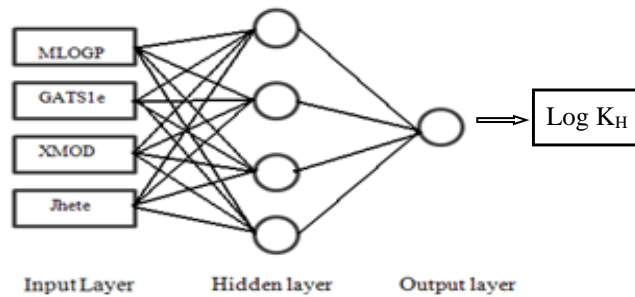
جدول (۷-۵) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان و چرخه‌های مختلف

		تعداد دوره‌های آموزشی							
سری آموزش		6	8	9	12	14	16	18	20
تعداد گره‌ها	2	1/529 9	0/875 4	0/6608	0/415 5	0/392 9	0/369 7	0/363 3	0/363 1
	3	0/520 1	0/457 3	0/4192	0/407 1	0/404 4	0/397 5	0/397 3	0/397 2
	4	0/302 3	0/298 5	<b>0/285</b> <b>3</b>	0/284 7	0/284 6	0/284 5	0/284 3	0/284 1
	5	0/422	0/352	0/3221	0/314	0/312	0/312	0/312	0/312

		6	9		7	9	5	3	2
	6	0/547 7	0/323 4	0/3057	0/293 2	0/285 9	0/279 2	0/276 6	0/274 0
	7	1/239 6	0/556 0	0/3625	0/321 0	0/302 2	0/291 4	0/262 5	0/254 0
	8	1/071 9	0/663 6	0/4725	0/402 5	0/398 4	0/395 9	0/380 6	0/352 1
	۹	0/386 7	0/308 4	0/2952	0/286 4	0/279 8	0/268 0	0/267 7	0/267 6
	۱۰	0/354 0	0/316 9	0/3647	0/355 0	0/332 7	0/321 7	0/316 7	0/313 5
سری آزمون		6	8	9	12	14	16	18	20
تعداد گروه‌ها	2	1/471 1	0/800 5	0/5694	0/418 7	0/395 8	0/276 6	0/263 7	0/262 6
	3	0/508 5	0/437 6	0/3941	0/391 1	0/397 6	0/400 8	0/400 5	0/401 2
	4	0/307 7	0/335 0	<b>0/277</b> <b>1</b>	0/284 0	0/284 9	0/285 0	0/282 2	0/284 6
	5	0/438 3	0/414 4	0/3061	0/315 8	0/326 7	0/331 8	0/334 0	0/335 1
	6	0/539 8	0/393 4	0/3690	0/366 2	0/380 8	0/383 6	0/387 0	0/384 6
	7	1/152 6	0/404 3	0/3339	0/326 9	0/329 1	0/359 4	0/374 3	0/391 1
	8	0/889 0	0/649 1	0/4704	0/417 7	0/417 0	0/413 1	0/386 3	0/334 2
	۹	0/307 6	0/279 1	0/2820	0/302 1	0/317 1	0/320 0	0/321 9	0/323 0
	۱۰	0/310 3	0/287 1	0/2793	0/289 7	0/266 2	0/266 1	0/266 1	0/265 8

#### ۵-۱-۲-۴- معماري شبکهي عصبي مصنوعي بهينه شده

با توجه به نتايج حاصل در بهينه‌سازي شبکهي، شبکهي عصبي پيشخور سه لايه و آموزش پس انتشار با الگوريتم آموزشي لونيبرگ - مارکورت و تابع انتقال لگاريتم سيگموئيد انتخاب گرديد. ساختار هندسي شبکهي بهينه شده در شکل (۵-۳) و پارامترهايي که تنظيم شدند و آنهايي که مقدار اوليه گرفتند، در جدول (۵-۸) آمده است.



شکل (۳-۵) - ساختار هندسی شبکه عصبی استفاده شده

جدول (۸-۵) - پارامترهای شبکه‌ی بهینه شده

۴	تعداد نرون‌های لایه‌ی پنهان
۹	تعداد دوره‌های آموزش
۰/۰۵	پارامتر ممتموم ( $\mu$ )
۰/۱	فاکتور کاهش برای پارامتر ممتموم
۱۰	فاکتور افزایش برای پارامتر ممتموم
$e^{-5}$	مقدار ماکزیمم برای پارامتر ممتموم

### ۵-۱-۲-۵-۵- فرایند ایجاد تصادفی Y<sup>۱</sup>

برای بررسی احتمال همبستگی تصادفی مدل بدست آمده، از فرآیند ایجاد تصادفی Y استفاده کردیم. در این فرآیند، ثابت هنری تجربی ترکیبات به صورت تصادفی جابجا و مدل‌های جدیدی ایجاد شدند. مقادیر پایین ضریب تعیین در این آزمون برای این مدل‌ها که در جدول (۹-۵) آمده است بیان‌گر تصادفی نبودن مدل اصلی می‌باشد.

جدول (۹-۵) - نتایج تصادفی کردن Y

مدل	۱	۲	۳	۴	۵	۶
$R^2$ برای سری آزمون	۰/۰۰	۰/۰۳	۰/۰۹	۰/۰۷	۰/۰۶	۰/۲۹
$R^2$ برای سری ارزیابی	۰/۰۴	۰/۰۱	۰/۱۱	۰/۰۱	۰/۳۰	۰/۰۹

<sup>۱</sup> - Y-Randomization

مقادیر تجربی و پیش‌بینی شده‌ی ثابت‌های هنری بوسیله‌ی روش‌های رگرسیون خطی چندگانه و شبکه‌ی عصبی مصنوعی در جدول (۵-۱۰) آمده است.

جدول (۵-۱۰) - مقادیر تجربی و پیش‌بینی شده‌ی ثابت‌های هنری بوسیله‌ی روش‌های رگرسیون خطی چندگانه و

شبکه‌ی عصبی مصنوعی

شماره	Log K <sub>H</sub>			شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN		Exp	MLR	ANN
1	1/510	1/602	1/564	35t	3/340	3/805	3/357
2	2/660	1/563	2/717	36	3/320	3/823	3/393
3t	4/100	3/911	4/361	37	4/030	4/718	4/223
4	1/280	1/839	1/529	38	4/290	4/240	4/490
5	1/260	1/361	1/553	39	4/980	5/262	4/859
6	5/690	5/595	5/700	40t	4/280	4/315	4/517
7t	1/180	1/580	1/527	41	4/830	3/842	4/107
8	4/540	4/099	4/411	42	4/300	4/106	4/454
9	1/200	1/387	1/526	43	5/860	5/575	5/700
10t	4/880	5/212	4/852	44	5/010	5/486	5/190
11v	4/550	5/178	4/960	45	4/320	3/995	4/395
12	5/070	5/438	4/893	46	5/360	5/620	5/592
13	4/850	5/077	4/831	47t	4/310	4/007	4/412
14	4/860	5/390	4/927	48	5/630	5/619	5/676
15t	2/500	2/600	2/430	49v	4/300	4/806	4/358
16	1/380	1/602	1/527	50	5/800	5/718	5/750
17	2/740	3/474	2/854	51	5/730	5/658	5/732
18v	2/220	2/781	2/471	52	3/840	3/677	4/085
19	1/670	1/978	2/470	53v	4/290	4/103	4/449
20t	2/270	2/658	2/471	54	1/210	1/424	1/526
21	2/200	2/656	2/471	55t	4/290	4/134	4/458
22t	2/590	2/807	2/471	56	5/760	5/565	5/696
23	2/630	2/802	2/471	57	1/160	1/462	1/602
24	2/080	2/244	2/470	58	1/160	1/476	1/526
25	2/400	2/551	2/471	59	5/750	5/627	5/723
26	2/030	1/741	2/442	60v	4/110	3/911	4/361
27v	2/540	2/614	2/471	61v	1/370	1/491	1/527
28	1/900	2/258	1/953	62	3/150	4/041	4/112
29v	2/490	3/037	2/801	63	1/360	1/616	1/527
30t	5/160	4/003	4/898	64	2/460	2/699	2/440
31	3/400	3/566	2/792	65t	1/420	1/490	1/527
32v	3/830	3/979	3/578	66	4/930	5/463	4/982
33t	3/810	3/684	3/331	67	2/890	3/050	2/752
34	3/350	3/766	3/282	68t	4/860	5/137	4/817

ادامه‌ی جدول (۵-۱۰)



شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
69	4/330	4/307	4/567
70	4/820	5/262	4/863
71	4/530	3/450	4/170
72t	3/090	3/171	2/874
73	5/700	5/424	5/598
74	5/170	5/179	4/863
75t	2/660	2/839	2/473
76v	2/380	2/667	2/471
77	3/100	2/334	2/305
78	2/550	3/264	2/768
79	2/310	2/004	1/537
80	2/460	2/597	2/375
81	2/610	2/046	1/854
82	3/590	3/859	3/498
83t	1/500	1/452	1/527
84	1/580	1/573	1/528
85	4/380	5/013	4/649
86	4/480	4/848	4/534
87t	4/530	4/164	4/482
88	4/450	5/447	5/336
89	4/350	4/034	4/400
90	5/600	5/624	5/680
91	4/550	4/069	4/447
92	2/570	2/046	1/854
93	5/680	5/438	5/610
94	4/460	4/023	4/389
95	4/370	4/866	4/537
96v	4/500	4/155	4/440
97	5/570	5/325	5/358
98	4/470	4/160	4/477
99t	4/450	4/490	4/589
100t	1/290	1/918	2/140
101	1/290	1/538	1/678
102	1/590	1/529	1/527
103	3/090	2/034	2/942

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
104	1/550	1/649	1/527
105	5/710	5/564	5/693
106v	1/280	1/517	1/526
107	1/540	1/543	1/527
108	3/400	2/368	2/447
109	5/240	5/486	5/223
110t	5/210	5/419	5/100
111	4/210	3/968	4/355
112v	5/160	5/304	4/958
113	2/560	3/168	2/514
114	5/780	5/739	5/759
115t	5/230	5/299	4/960
116	5/120	5/419	5/100
117	3/220	3/261	3/007
118	2/660	3/337	2/663
119	2/580	3/450	2/818
120t	2/610	1/689	2/199
121	2/530	3/445	2/814
122	4/490	5/050	4/746
123t	4/560	4/895	4/680
124	4/680	5/029	4/775
125t	4/540	4/081	4/441
126	4/700	4/884	4/656
127	4/660	4/503	4/569
128	4/650	4/557	4/624
129	1/470	1/596	1/766
130t	2/580	2/041	3/307
131	1/440	1/544	1/527
132	5/310	5/494	5/459
133	5/310	5/429	5/365
134	5/720	5/334	5/505
135	5/380	5/447	5/336
136t	4/110	3/793	4/320
137	5/920	5/791	5/777
138	5/310	5/319	5/157

ادامه‌ی جدول (۵-۱۰)

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
139t	5/360	5/328	5/167
140v	2/600	2/112	2/571
141	4/310	3/995	4/395
142	5/280	5/617	5/604

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
174	5/750	5/458	5/618
175	5/500	5/613	5/673
176	4/560	4/055	4/429
177	5/600	5/556	5/629

143	5/410	5/621	5/587
144	5/810	5/788	5/766
145	1/110	2/483	1/653
146	3/260	3/407	2/977
147v	3/330	3/694	3/204
148	2/250	3/283	2/980
149	4/430	4/755	4/950
150t	2/900	3/496	2/768
151	4/760	5/011	4/824
152	4/650	4/110	4/480
153v	4/760	4/917	4/780
154	1/560	1/638	1/842
155	1/550	1/530	1/527
156v	5/690	5/513	5/560
157	5/700	5/712	5/734
158	5/670	5/625	5/678
159	5/590	5/572	5/641
160t	5/510	5/486	5/223
161	5/520	5/430	5/522
162v	4/200	4/074	4/458
163v	5/570	5/352	5/369
164	5/580	5/439	5/518
165	5/760	5/541	5/679
166	2/570	3/086	2/672
167	5/580	5/446	5/512
168	5/530	5/439	5/511
169	2/090	2/518	2/411
170t	5/560	5/459	5/524
171	5/660	5/608	5/703
172	5/610	5/513	5/579
173	5/590	5/453	5/505

178	5/520	5/554	5/633
179	2/360	2/617	2/349
180t	2/210	2/378	2/485
181	5/560	5/618	5/677
182	4/310	4/007	4/412
183	5/550	5/325	5/373
184	4/450	4/161	4/478
185	5/750	5/337	5/371
186	5/700	5/621	5/714
187	5/770	5/459	5/611
188	5/590	5/734	5/732
189	3/140	3/635	3/034
190t	5/660	5/738	5/738
191v	5/480	5/604	5/664
192	3/170	3/376	3/280
193	1/330	2/692	2/154
194	2/860	3/424	2/684
195	5/840	5/724	5/757
196	2/640	3/621	2/888
197	2/640	3/527	2/778
198	1/640	1/715	1/531
199	5/670	5/323	5/502
200	5/710	5/321	5/492
201	5/360	5/503	5/441
202	5/690	5/339	5/509
203	5/710	5/352	5/511
204	5/720	5/430	5/604
205t	5/660	5/444	5/615
206	4/540	4/069	4/447
207	5/700	5/423	5/599
208	5/670	5/424	5/598

ادامه‌ی جدول (۵-۱۰)

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
209	2/360	2/210	2/455
210v	5/730	5/486	5/645
211	5/730	5/438	5/608
212	5/700	5/440	5/611
213	2/670	2/837	2/473
214v	5/720	5/508	5/661
215	5/730	5/449	5/616
216	5/660	5/366	5/524
217	4/130	3/752	4/255
218	5/730	5/452	5/618

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
244	5/770	5/547	5/685
245	4/320	4/144	4/500
246	5/810	5/516	5/659
247	5/650	5/801	5/758
248	1/480	1/630	1/527
249	5/660	5/613	5/712
250	5/810	5/624	5/726
251	5/760	5/558	5/692
252	4/290	4/134	4/458
253	5/780	5/616	5/719

219	5/690	5/460	5/626
220t	2/510	1/781	2/294
221	5/710	5/438	5/609
222	5/780	5/501	5/655
223	5/560	5/618	5/677
224	2/580	3/026	2/680
225	5/750	5/453	5/614
226	3/710	3/731	3/223
227	5/730	5/601	5/704
228	5/520	5/452	5/523
229v	5/680	5/590	5/692
230t	5/640	5/586	5/682
231	2/710	2/156	2/703
232	5/760	5/607	5/711
233	5/740	5/551	5/687
234	5/500	5/448	5/530
235	5/690	5/536	5/672
236	5/530	5/221	5/192
237	5/710	5/548	5/683
238	4/120	3/793	4/320
239	1/090	1/433	1/527
240	5/780	5/610	5/716
241	5/740	5/602	5/709
242	4/050	3/586	4/205
243	2/970	2/395	3/240

254v	1/420	1/630	1/527
255	5/810	5/620	5/724
256	5/800	5/616	5/723
257	4/990	5/555	5/517
258	4/190	4/102	4/415
259	2/700	2/890	2/643
260v	5/820	5/777	5/752
261	5/560	5/630	5/683
262	5/590	5/758	5/727
263	2/680	3/627	2/895
264	5/740	5/723	5/746
265	3/820	3/563	3/091
266	4/310	4/151	4/480
267	5/910	5/784	5/774
268	4/960	4/105	4/612
269	5/860	5/625	5/728
270	5/360	5/316	5/172
271	5/280	5/295	4/964
272	5/890	5/734	5/765
273t	2/710	1/900	2/363
274	5/910	5/901	5/787
275	5/880	5/894	5/777
276	2/920	3/407	2/977
277	3/090	3/442	2/740
278	1/380	1/616	1/531

ادامه‌ی جدول (۵-۱۰)

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
279	5/070	4/129	4/550
280v	5/030	5/218	5/471
281	2/200	3/425	2/722
282	1/950	1/508	1/534
283	2/830	2/233	2/476
284	3/150	3/447	2/815
285	5/030	4/121	4/582
286	2/590	3/430	2/849
287t	2/570	3/442	2/903
288	5/780	5/685	5/745
289	4/880	5/148	5/612
290t	2/860	2/243	3/318
291	4/180	4/048	4/307
292	2/800	2/802	2/875
293	2/700	2/369	2/697
294	4/580	4/902	4/914

شماره	Log K <sub>H</sub>		
	Exp	MLR	ANN
307	2/540	1/999	2/433
308	1/020	1/317	1/527
309	2/830	2/064	2/416
310	5/780	5/617	5/720
311v	3/030	2/601	3/292
312	4/710	4/999	4/869
313v	4/810	5/155	4/880
314	2/140	1/924	2/273
315	5/680	5/715	5/737
316	5/690	5/443	5/609
317	5/530	5/433	5/517
318	2/470	3/349	2/717
319	5/560	5/516	5/595
320	2/200	2/601	2/417
321	1/280	0/344	1/524
322	3/550	3/436	2/628

295	2/400	2/675	2/554	323	5/760	5/519	5/667
296	2/370	2/893	2/572	324	5/820	5/720	5/754
297	2/040	2/179	2/393	325	2/580	0/965	1/528
298	1/710	1/089	1/524	326	5/520	5/793	5/738
299 <sub>v</sub>	3/120	2/625	3/384	327 <sub>v</sub>	3/150	3/688	3/120
300	2/480	1/220	1/536	328	1/140	1/386	1/528
301	2/200	1/532	1/525	329	4/020	4/633	3/922
302	3/600	3/406	2/839	330 <sub>t</sub>	4/210	4/026	4/394
303	5/720	5/468	5/630	331	5/580	5/339	5/355
304	5/620	5/551	5/679	332	4/530	4/069	4/447
305	5/420	5/504	5/410	333	4/350	3/876	4/307
306	4/060	3/933	4/356				

## ۵-۲- بررسی نتایج

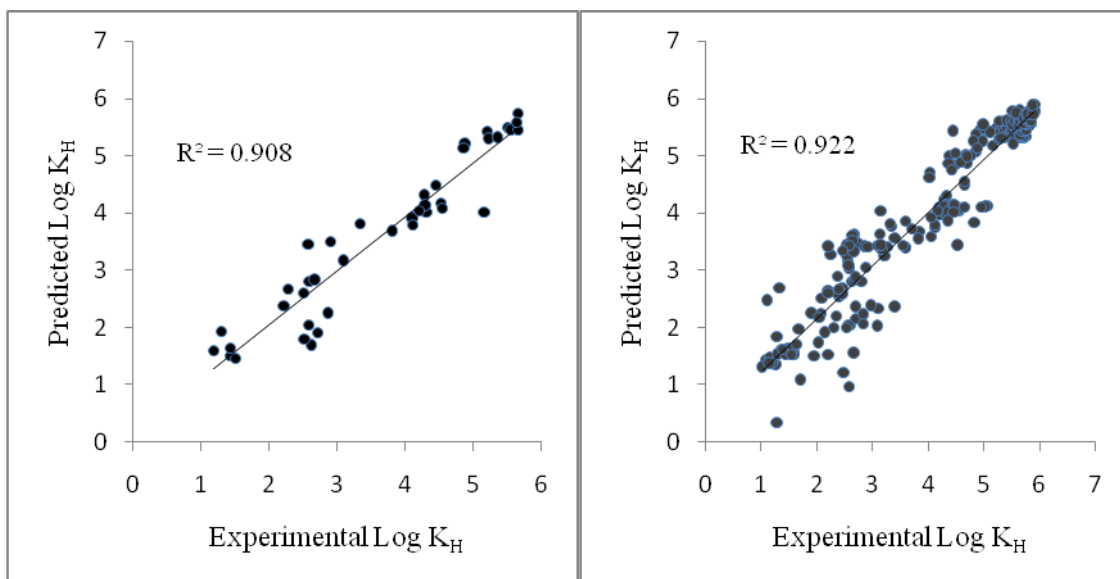
### ۵-۲-۱- ارزیابی مدل‌های شبکه‌ی عصبی و رگرسیون خطی چندگانه

برای بررسی میزان کارایی مدل شبکه‌ی عصبی ارائه شده و مقایسه‌ی آن با روش رگرسیون خطی چندگانه، پارامترهای آماری حاصل در جدول (۵-۱۱) آورده شده است. مقدار ضریب همبستگی و آماره‌ی F بالا نشان از توانایی خوب مدل‌های ANN و MLR در پیش‌بینی ثابت هنری دارد.

جدول (۵-۱۱)- پارامترهای آماری مدل شبکه عصبی و روش رگرسیون خطی چندگانه

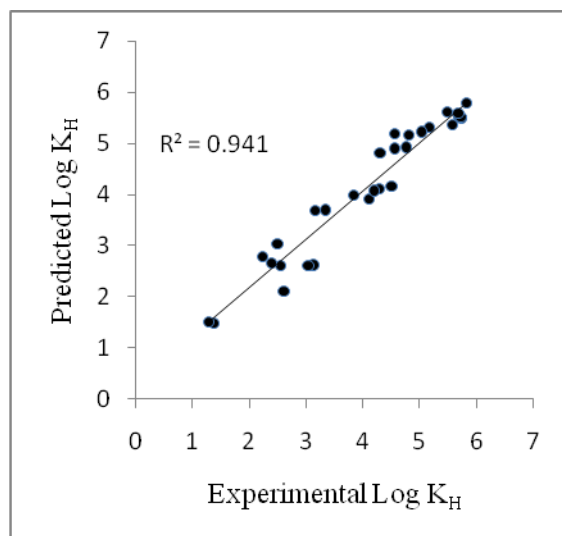
ANN	R	SE	F	MSE
سری آموزش	0/983	0/288	7412/747	0/082
سری آزمون	0/981	0/279	953/477	0/07۷
سری ارزیابی	0/991	0/186	1531/027	0/039
MLR	R	SE	F	MSE
سری آموزش	0/960	0/438	768/630	0/188
سری آزمون	0/953	0/431	377/892	0/180
سری ارزیابی	0/970	0/334	451/592	0/109

رابطه‌ی خطی ثابت هنری پیش‌بینی شده از دو روش بر حسب مقادیر واقعی نیز در شکل‌های (۴-۵) و (۵-۵) نشان داده شده است. با توجه به پارامترهای آماری بالاتر و نمودار خطی متراکم‌تر در این شکل‌ها، می‌توان بیان کرد که توانایی مدل حاصل از روش شبکه عصبی مصنوعی در پیش‌بینی ثابت هنری سری داده‌ها از مدل بدست آمده از روش رگرسیون خطی چندگانه بیشتر می‌باشد.



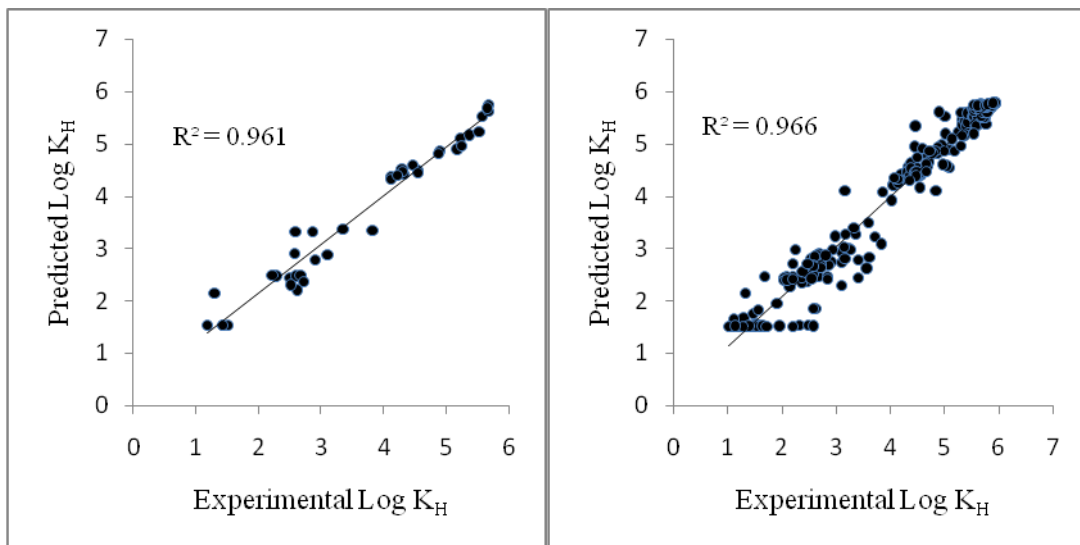
(ب)

(الف)



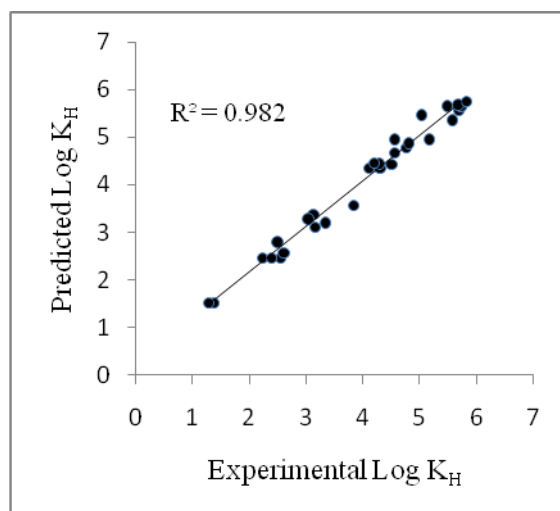
(ج)

شکل (۴-۵) - نمودار مقادیر پیش‌بینی شده  $\text{Log } K_H$  بر حسب مقادیر واقعی با روش MLR برای سری آموزش،  
 (ب) سری آزمون و (ج) سری ارزیابی



(ب)

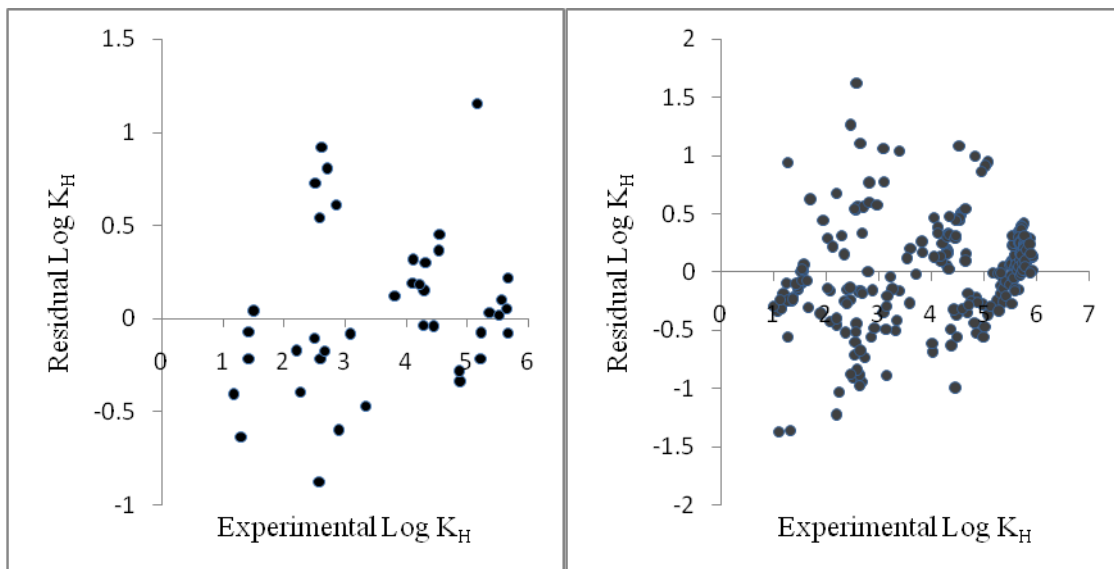
(الف)



(ج)

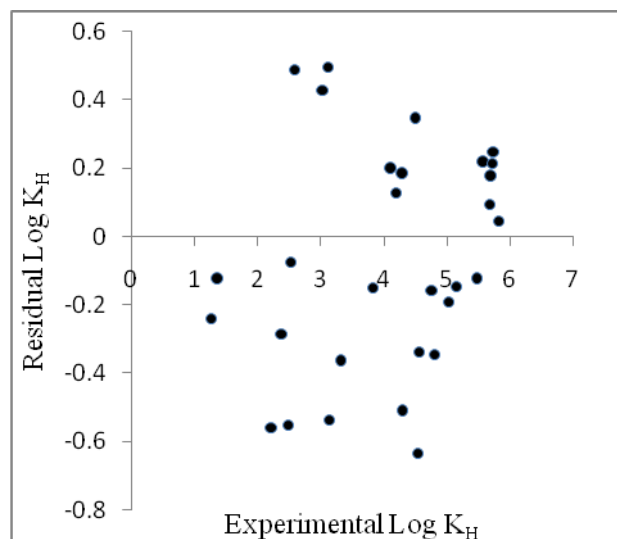
شکل (۵-۵) - نمودار مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  بر حسب مقادیر واقعی با روش ANN برای الف) سری آموزش، ب) سری آزمون و ج) ارزیابی

همچنین به منظور بررسی وجود خطای معین در این مدل‌ها، باقیمانده‌ی مقادیر محاسبه شده از  $\text{Log } K_H$  تجربی بر حسب مقادیر تجربی برای ترکیبات در شکل‌های (۵-۶) و (۵-۷) رسم شده است.



(ب)

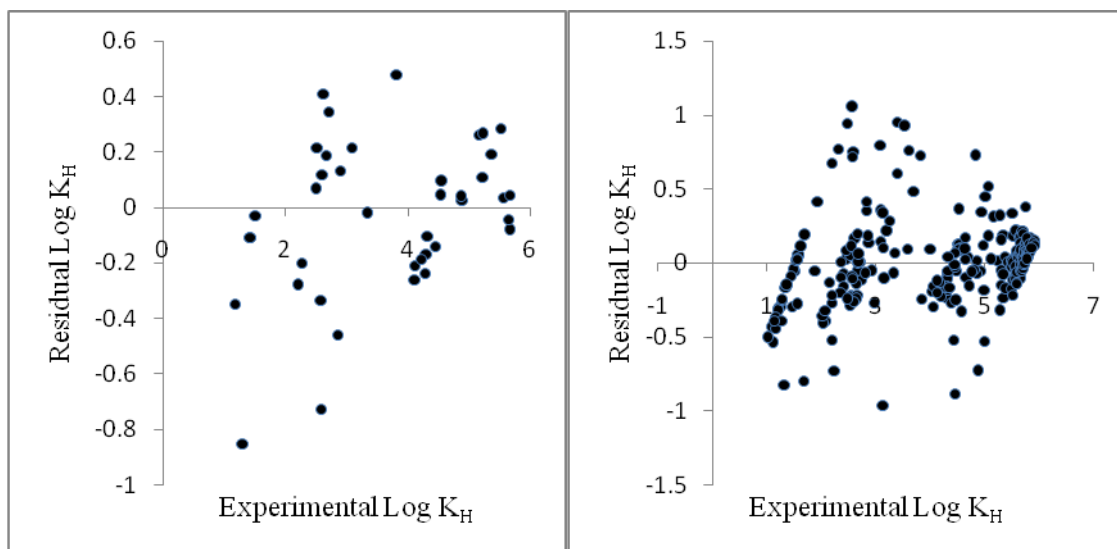
(الف)



(ج)

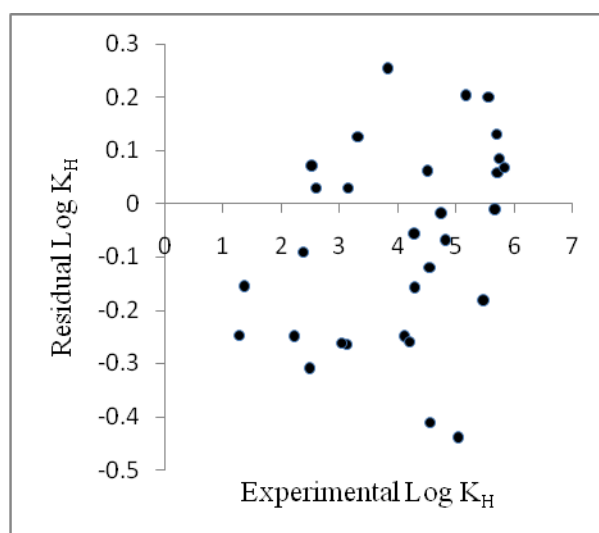
شکل (۵-۶) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی  $\text{Log } K_H$  با روش MLR از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، ب) سری آزمون و ج) سری ارزیابی





(ب)

(الف)



(ج)

شکل (۷-۵) - نمودار باقیمانده ی مقادیر پیش‌بینی شده ی  $\text{Log } K_H$  با روش ANN از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، ب) سری آزمون و ج) سری ارزیابی

پراکندگی مناسب و یکنواخت نقاط باقیمانده‌ها در دو طرف خط صفر نشان‌دهنده عدم وجود خطای معین قابل توجه در مقادیر پیش‌بینی شده ی روش مورد استفاده است. با توجه به پراکندگی نسبتاً یکنواخت‌تر در شکل (۷-۵) مشخص می‌گردد که خطای معین کمتری در مدل حاصل از روش شبکه عصبی مصنوعی اتفاق افتاده است.

## ۵-۲-۲- تجزیه و تحلیل توصیف‌کننده‌های مدل

توصیف‌کننده‌های گروه ارتباط دو بعدی<sup>۱</sup> براساس تابع خود ارتباطی  $AC_l$  تعریف می‌شوند:

$$AC_l = \int_a^b f(x).f(x+l).dx \quad (۲-۵)$$

که  $f(x)$  هر تابع از متغیر  $x$  و  $l$  اختلاف فاصله از  $x$  می‌باشد؛  $a$  و  $b$  فاصله‌ی مطالعه شده را بیان می‌کند.  $f(x)$  معمولاً یک تابع وابسته به زمان مانند یک سیگنال الکتریکی یا وابسته به فضا مانند دانسیته جمعیت در فضا می‌باشد. این گروه از توصیف‌کننده‌ها به بررسی اطلاعات مختلفی در مورد ترکیبات می‌پردازد. نوعی از این توصیف‌کننده‌ها گروه Geary Autocorrelation است که ضریب گری بدین صورت محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (۳-۵)$$

که  $w$  یک ویژگی اتم،  $\bar{w}$  میانگین مقدار آن روی مولکول،  $A$  تعداد اتم‌ها و  $d$  فاصله‌ی توپولوژیک است و  $\delta_{ij}$  نیز - که به تابع کرونیگر<sup>۲</sup> معروف است - در حالتی که  $d_{ij} = d$  باشد یک است، و گرنه صفر است و  $\Delta$  هم مجموعه‌ی  $\delta$  هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است [۳۰]. گروهی از این توصیف‌کننده‌ها نقش الکترون‌گاتیویته‌ی ساندerson را در ترکیبات نشان می‌دهند. وجود علامت مثبت در مدل برای توصیف‌کننده GATS1e از این گروه نشان می‌دهد که افزایش مقدار این توصیف‌کننده سبب افزایش  $\text{Log } K_H$  ترکیبات مورد بررسی می‌شود.

توصیف‌کننده‌های توپولوژیکی بر اساس نمایش گراف مولکولی می‌باشند. در این گراف‌ها هر نقطه نشان‌دهنده یک اتم بوده و خطوط بین نقاط نیز نشان‌دهنده پیوند شیمیایی بین اتم‌ها

<sup>۱</sup>- 2D Autocorrelation

<sup>۲</sup>- Kronecker

می‌باشند درحالی‌که از اتم‌های هیدروژن صرف‌نظر شده است. این توصیف‌کننده‌ها اطلاعاتی راجع به ساختمان، اندازه، شکل، تقارن، شاخه‌دار شدن، نحوه‌ی اتصال اتم‌ها و نوع اتم‌های موجود در یک مولکول در اختیار ما قرار می‌دهند. محاسبه‌ی این توصیف‌کننده‌ها به سادگی از روی ساختمان دو بعدی مولکول‌ها امکان‌پذیر می‌باشد [30]. از این گروه توصیف‌کننده‌های XMOD و Jhete در مدل وارد شدند. توصیف‌کننده‌ی XMOD، شاخص راندیک اصلاح شده بوده و به صورت زیر بیان می‌شود:

$${}^1\chi_{\text{mod}} = \frac{1}{2} \sum_{i=1}^A \sum_{j=1}^{\delta_i} \frac{Z_i}{(\delta_i \cdot \delta_j)^{1/2}} \quad (4-5)$$

که مجموع اول بر روی تمام اتم‌ها در گراف مولکولی اجرا شده، درحالی‌که مجموع دوم اولین اتم‌های همسایه‌ی مجاور را در بر می‌گیرد و  $\delta$  درجه‌ی راس‌ها یا به عبارتی تعداد پیوندهای هر اتم و  $Z_i$  عدد اتمی  $i$  امین اتم می‌باشد. تاثیر گذاری این توصیف‌کننده در مدل، ناشی از حضور اتم‌های با عدد اتمی بزرگ‌تر از کربن است که به عنوان شاخه‌ای به اسکلت هیدروکربنی افزوده شده و قطبیت بیشتر این اتم‌ها باعث کاهش ثابت هنری می‌شود. به همین دلیل علامت این توصیف‌کننده منفی می‌باشد.

توصیف‌کننده‌ی Jhete از نوع شاخص بالابان با ماتریس فاصله‌ی وزن دار شده با الکترونگاتیویته می‌باشد. اندیس بالابان  $J$  برای گراف  $G$  بدین صورت بیان می‌شود:

$$J = (M / (M - N + 2)) \sum_M (d_i d_j)^{-0.5} \quad (5-5)$$

که  $M$  تعداد پیوندها در  $G$  و  $d_i$  (برای  $i=1,2,\dots,N$  و  $N$  تعداد رئوس  $G$ ) مجموع فاصله است. شاخه-دار شدن و الکترونگاتیویته در ماتریس فاصله ویژگی‌های این توصیف‌کننده است. کم اثر بودن الکترونگاتیویته نسبت به شاخه‌دار شدن برای کل سری باعث مثبت بودن اما با پایین‌ترین اثر متوسط آن بر ثابت هنری شده است چرا که افزایش شاخه‌ها و کربن‌ها موجب افزایش این ثابت می‌شود.

از گروه توصیف‌کننده‌های خواص<sup>۱</sup> فیزیکوشیمیایی، لگاریتم ضریب توزیع آب - اکتانول (MLOGP) یا چربی دوستی<sup>۲</sup> در مدل ایجاد شده آمده است که به وسیله‌ی موری کوچی بیان شده است [۶۹]. این ضریب معرف خاصیت آبگریزی ترکیب می‌باشد. این توصیف‌کننده در مدل دارای علامت مثبت و بزرگ‌ترین اثر متوسط می‌باشد پس افزایش آن باعث افزایش ثابت هنری خواهد شد چرا که با افزایش خاصیت آبگریزی ترکیب فشار بخار آن در بالای سطح مایع افزایش یافته و این ثابت بزرگ‌تر می‌شود.

---

<sup>۱</sup>- Properties

<sup>۲</sup>- Lipophilicity

## فصل ششم

# مدل سازی QSPR ضریب شکست

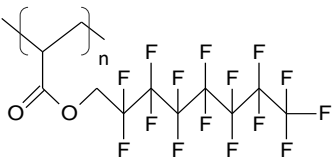
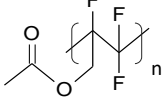
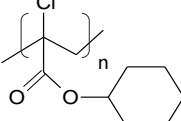
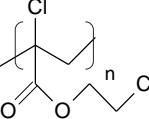
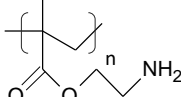
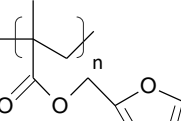
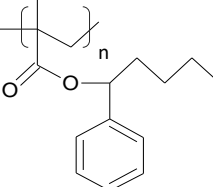
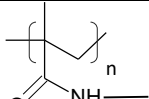
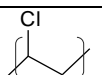
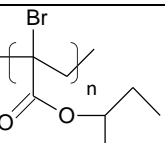
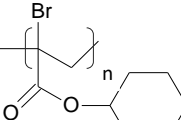
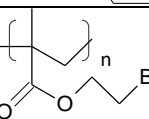
## ۶-۱- مقدمه

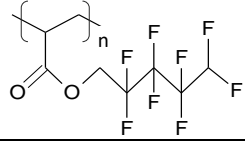
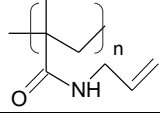
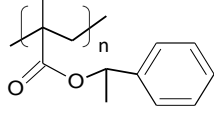
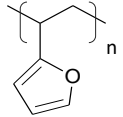
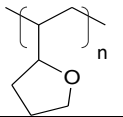
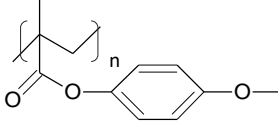
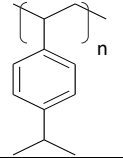
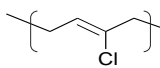
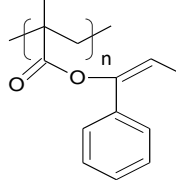
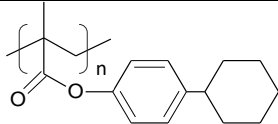
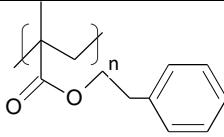
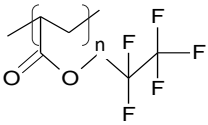
در فصل سوم مفاهیم و اصول الگوریتم ژنتیک مطرح گردید. قابل ذکر است که الگوریتم ژنتیک هیچ چیزی درباره‌ی مسأله‌ای که حل می‌کند، نمی‌داند و برای اجرای جستجوی موثر فقط به مقادیر تابع هدف نیاز دارد بنابراین شامل پیچیدگی‌های محاسباتی یا محاسبات اضافی نمی‌شود، اما از آنجا که روشی غیر جبری است پاسخ دقیق مسأله را نیافته و حتی ممکن است برای حل یک مسأله‌ی مشخص با هر بار بکارگیری، پاسخی متفاوت ارائه دهد، اگرچه تمامی این پاسخ‌ها می‌توانند پاسخ‌هایی باشند که دقت موردنظر را برآورده کنند. تاکنون هیچ روش استاندارد برای حل یک مسأله بخصوص در دسترس نبوده و با بکارگیری تجربیات موجود می‌توان الگوریتم ژنتیک را با احتیاجات خود منطبق کرد. بهینه کردن پارامتری‌های الگوریتم ژنتیک با حدس‌های اولیه‌ای که منجر به اصلاح بیشتر ساختار آن شود، آغاز شده و روش‌های مختلفی برای رسیدن به الگوریتم بهینه بکار گرفته می‌شود. در این مرحله از تحقیق، توصیف‌کننده‌های منتخب الگوریتم ژنتیک بر اساس رگرسیون خطی چندگانه، برای مدل‌سازی بوسیله‌ی شبکه‌ی عصبی مصنوعی در مطالعه‌ی رابطه‌ی ساختار - خاصیت برای پیش‌بینی ضریب شکست یک سری از ترکیبات پلیمری استفاده شده است.

## ۶-۲- سری داده‌ها

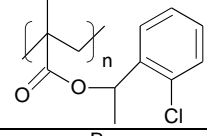
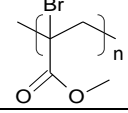
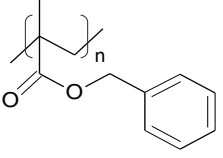
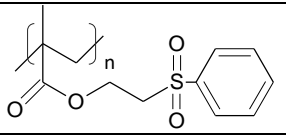
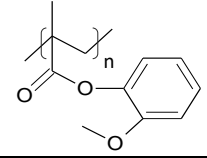
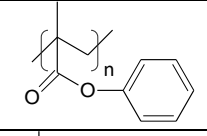
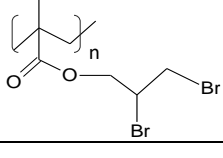
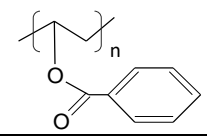
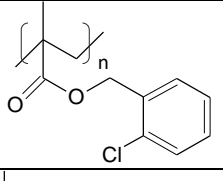
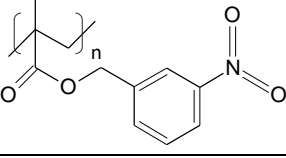
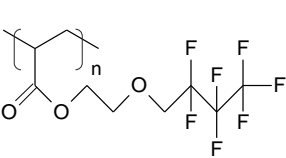
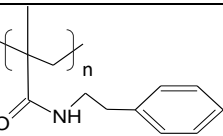
داده‌های تجربی گزارش شده از ضریب شکست در طول موج ۵۸۹۳ آنگستروم برای ۱۴۴ پلیمر جمع آوری شدند [۷۰]. در جدول (۶-۱) نام، ساختار مونومری و ضریب شکست این ترکیبات آورده شده است.

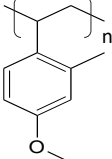
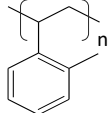
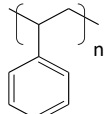
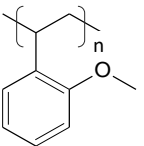
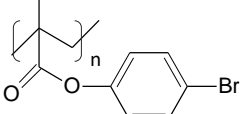
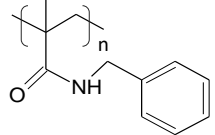
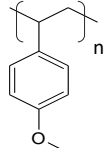
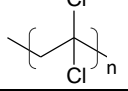
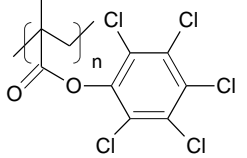
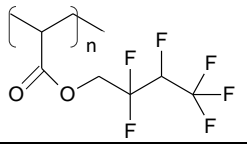
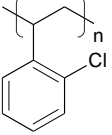
جدول (۶-۱) - نام، ساختار و ضریب شکست سری داده‌ها

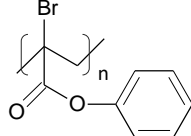
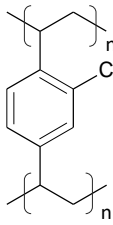
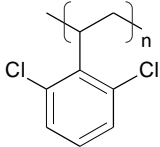
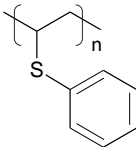
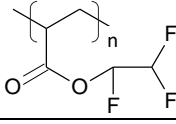
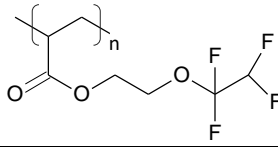
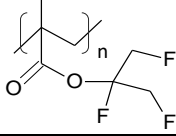
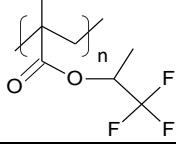
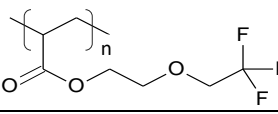
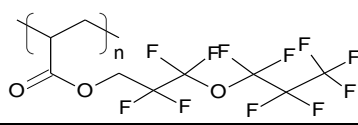
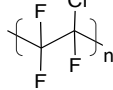
شماره	نام	ساختار	ضریب شکست
1	Poly(pentadecafluorooctyl acrylate)		1/339
2	Poly(trifluorovinyl acetate)		1/375
3	Poly(cyclohexyl α-chloroacrylate)		1/532
4	Poly(2-chloroethyl α-chloroacrylate)		1/533
5	Poly(2-aminoethyl methacrylate)		1/537
6	Poly(furfuryl methacrylate)		1/538
7	Poly(1-phenyl-n-amyl methacrylate)		1/540
8	Poly(N-methyl-methacrylamide)		1/540
9	Poly(vinyl chloride)		1/540
10	Poly(sec-butyl α-bromoacrylate)		1/542
11	Poly(cyclohexyl α-bromoacrylate)		1/542
12	Poly(2-bromoethyl methacrylate)		1/543

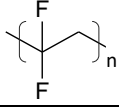
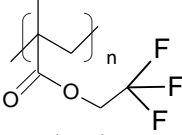
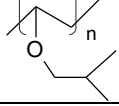
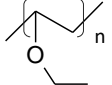
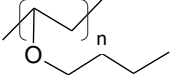
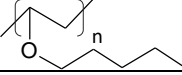
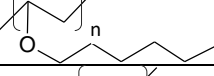

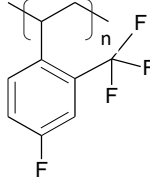
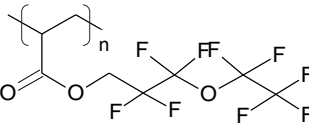
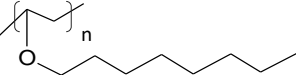

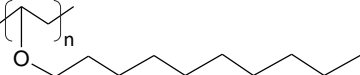
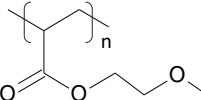
13	Poly(octafluoropentyl acrylate)		1/380
14	Poly(N-allyl methacrylamide)		1/548
15	Poly(1-phenylethyl methacrylate)		1/549
16	Poly(vinylfuran)		1/550
17	Poly(2-vinyltetrahydrofuran)		1/550
18	Poly(p-methoxybenzyl methacrylate)		1/552
19	Poly(p-isopropylstyrene)		1/554
20	Poly(chloroprene)		1/556
21	Poly(1-phenylallyl methacrylate)		1/557
22	Poly(p-cyclohexylphenyl methacrylate)		1/558
23	Poly(2-phenylethyl methacrylate)		1/559
24	Poly(pentafluoropropyl acrylate)		1/385

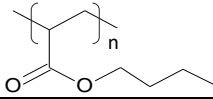
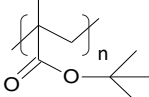
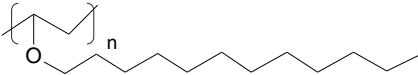
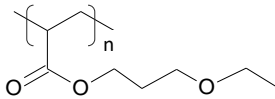
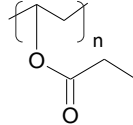
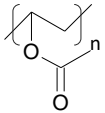
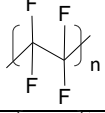
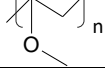
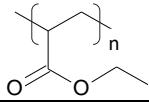
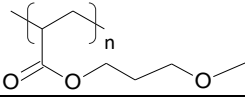
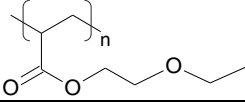
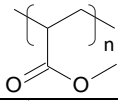
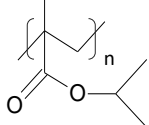
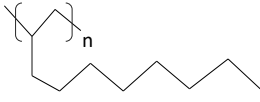



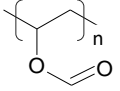
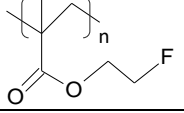
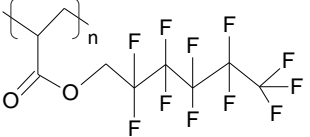
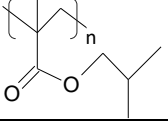
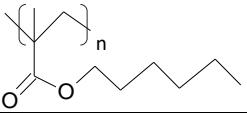
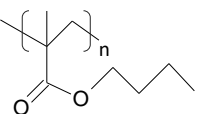
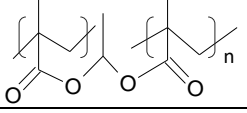
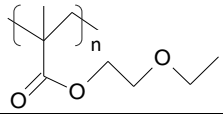
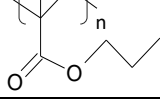
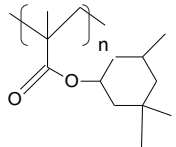
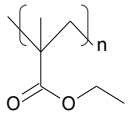
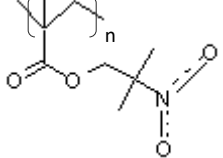
25	Poly(1-(o-chlorophenyl)ethyl methacrylate)		1/562
26	Poly(methyl α-bromoacrylate)		1/567
27	Poly(benzyl methacrylate)		1/568
28	Poly(2-(phenylsulfonyl)ethyl methacrylate)		1/568
29	Poly(o-methoxyphenyl methacrylate)		1/571
30	Poly(phenyl methacrylate)		1/571
31	Poly(2,3-dibromopropyl methacrylate)		1/574
32	Poly(vinyl benzoate)		1/578
33	Poly(o-chlorobenzyl methacrylate)		1/582
34	Poly(m-nitrobenzyl methacrylate)		1/585
35	Poly(2-heptafluorobutoxy)ethyl acrylate)		1/390
36	Poly(N-(2-phenylethyl) methacrylamide)		1/586

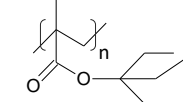
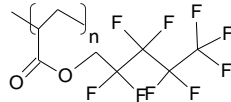
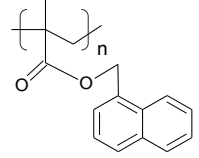
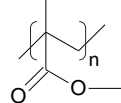
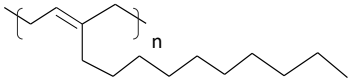
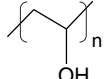
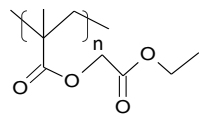
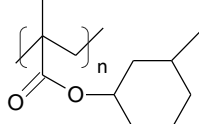
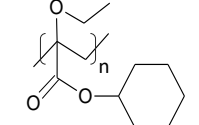
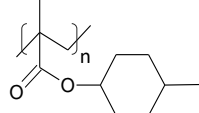
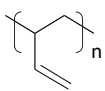
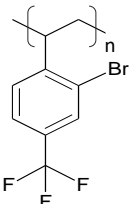
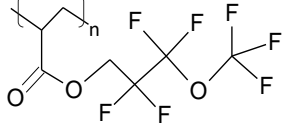
37	Poly(4-methoxy-2-methylstyrene)		1/587
38	Poly(o-methylstyrene)		1/587
39	Poly(styrene)		1/591
40	Poly(o-methoxystyrene)		1/593
41	Poly(p-bromophenyl methacrylate)		1/596
42	Poly(N-benzyl methacrylamide)		1/597
43	Poly(p-methoxystyrene)		1/597
44	Poly(vinylidene chloride)		1/600
45	Poly(pentachlorophenyl methacrylate)		1/608
46	Poly(2,2,3,4,4,4-hexafluorobutyl acrylate)		1/392
47	Poly(o-chlorostyrene)		1/610

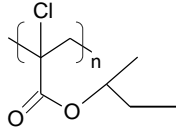
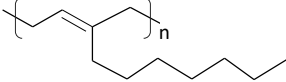
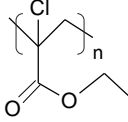
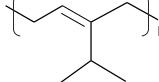
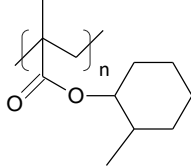

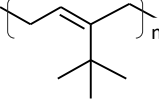
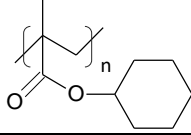
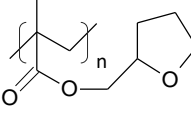
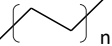
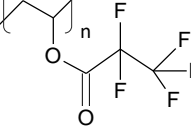
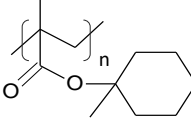
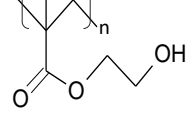
48	Poly(phenyl $\alpha$ -bromoacrylate)		1/612
49	Poly(p-divinylbenzene)		1/615
50	Poly(2,6-dichlorostyrene)		1/625
51	Poly(vinyl phenyl sulfide)		1/657
52	Poly(trifluoroethyl acrylate)		1/407
53	Poly(2-(1,1,2,2-tetrafluoroethoxy)ethyl acrylate)		1/412
54	Poly(trifluoroisopropyl methacrylate)		1/418
55	Poly(2,2,2-trifluoro-1-methylethyl methacrylate)		1/419
56	Poly(2-(trifluoroethoxy)ethyl acrylate)		1/419
57	Poly(tetrafluoro-3-(heptafluoropropoxy) propyl acrylate)		1/346
58	Poly(trifluorochloroethy lene)		1/420

59	Poly(vinylidene fluoride)		1/420
60	Poly(trifluoroethyl methacrylate)		1/437
61	Poly(vinyl isobutyl ether)		1/451
62	Poly (vinyl ethyl ether)		1/454
63	Poly(vinyl n-butyl ether)		1/456
64	Poly(vinyl pentyl ether)		1/458
65	Poly(vinyl hexyl ether)		1/459
66	Poly(4-methyl-1-pentene)		1/459
67	Poly(4-fluoro-2-trifluoromethylstyrene)		1/460
68	Poly(tetrafluoro-3-(pentafluoroethoxy)propyl acrylate)		1/348
69	Poly(vinyl octyl ether)		1/461
70	Poly(vinyl-2-ethylhexyl ether)		1/463
71	Poly(vinyl decyl ether)		1/463
72	Poly(2-methoxyethyl acrylate)		1/463

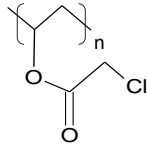
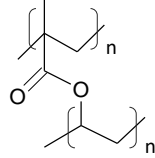
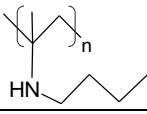
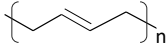
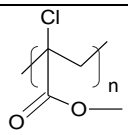
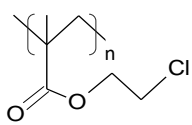
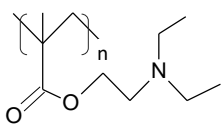
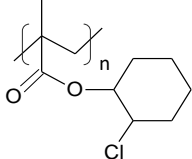
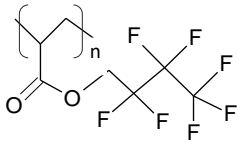
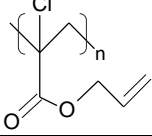
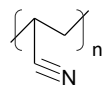
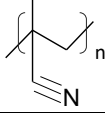
73	Poly(butyl acrylate)		1/463
74	Poly(tert-butyl methacrylate)		1/464
75	Poly(vinyl dodecyl ether)		1/464
76	Poly(3-ethoxypropyl acrylate)		1/465
77	Poly(vinyl propionate)		1/467
78	Poly(vinyl acetate)		1/467
79	Poly(tetrafluoroethylene)		1/356
80	Poly(vinyl methyl ether)		1/467
81	Poly(ethyl acrylate)		1/469
82	Poly(3-methoxypropyl acrylate)		1/471
83	Poly(2-ethoxyethyl acrylate)		1/471
84	Poly(methyl acrylate)		1/472
85	Poly(isopropyl methacrylate)		1/473
86	Poly(1-decene)		1/473

87	Poly(propylene)		1/503
88	Poly(vinyl formate)		1/476
89	Poly(2-fluoroethyl methacrylate)		1/477
90	Poly(undecafluorohexyl acrylate)		1/356
91	Poly(isobutyl methacrylate)		1/477
92	Poly(n-hexyl methacrylate)		1/481
93	Poly(n-butyl methacrylate)		1/483
94	Poly(ethylidene dimethacrylate)		1/483
95	Poly(2-ethoxyethyl methacrylate)		1/483
96	Poly(n-propyl methacrylate)		1/484
97	Poly(3,3,5-trimethylcyclohexyl methacrylate)		1/485
98	Poly(ethyl methacrylate)		1/485
99	Poly(2-nitro-2-methylpropyl methacrylate)		1/487

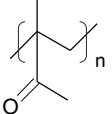
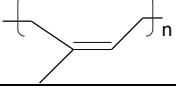
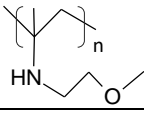
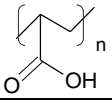
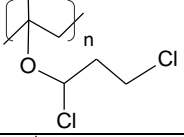
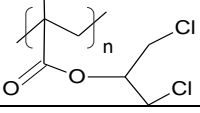
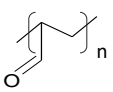
100	Poly(triethylcarbinyl methacrylate)		1/489
101	Poly(nonafluoropentyl acrylate)		1/360
102	Poly( $\alpha$ -naphthyl carbinyl methacrylate)		1/630
103	Poly(methyl methacrylate)		1/489
104	Poly(2-decyl-1,3-butadiene)		1/490
105	Poly(vinyl alcohol)		1/500
106	Poly(ethyl glycolate methacrylate)		1/490
107	Poly(3-methylcyclohexyl methacrylate)		1/495
108	Poly(cyclohexyl $\alpha$ -ethoxyacrylate)		1/497
109	Poly(4-methylcyclohexyl methacrylate)		1/498
110	Poly(1,2-butadiene)		1/500
111	Poly(2-bromo-4-trifluoromethylstyrene)		1/500
112	Poly(tetrafluoro-3-(trifluoromethoxy)propyl acrylate)		1/360

113	Poly(sec-butyl $\alpha$ -chloroacrylate)		1/500
114	Poly (2-heptyl-1,3-butadiene)		1/500
115	Poly(ethyl $\alpha$ -chloroacrylate)		1/502
116	Poly(2-isopropyl-1,3-butadiene)		1/503
117	Poly(2-methylcyclohexyl methacrylate)		1/503
118	Poly(isobutene)		1/505
119	Poly(2-tert-butyl-1,3-butadiene)		1/506
120	Poly(cyclohexyl methacrylate)		1/507
121	Poly(tetrahydrofurfuryl methacrylate)		1/510
122	Poly(ethylene)		1/510
123	Poly(pentafluorovinyl propionate)		1/364
124	Poly(1-methylcyclohexyl methacrylate)		1/511
125	Poly(2-hydroxyethyl methacrylate)		1/512



126	Poly(vinyl chloroacetate)		1/512
127	Poly(vinyl methacrylate)		1/513
128	Poly(N-butyl methacrylamide)		1/514
129	Poly(1,3-butadiene)		1/515
130	Poly(methyl α-chloroacrylate)		1/517
131	Poly(2-chloroethyl methacrylate)		1/517
132	Poly(2-diethylaminoethyl methacrylate)		1/517
133	Poly(2-chlorocyclohexyl methacrylate)		1/518
134	Poly(heptafluorobutyl acrylate)		1/367
135	Poly (allyl methacrylate)		1/520
136	Poly(acrylonitrile)		1/520
137	Poly(methacrylonitrile)		1/520

ادامه‌ی جدول (۱-۶)

138	Poly(methyl isopropenyl ketone)		1/520
139	Poly(isoprene)		1/521
140	Poly((N-2-methoxyethyl)methacrylamide)		1/525
141	Poly(acrylic acid)		1/527
142	Poly(1,3-dichloropropyl methacrylate)		1/527
143	Poly(2-chloro-1-(chloromethyl) ethyl methacrylate)		1/527
144	Poly(acrolein)		1/529

### ۳-۶- مدلسازی ضریب شکست

#### ۳-۶-۱- رسم و بهینه‌سازی ساختار مولکول‌ها

پس از جمع‌آوری داده‌ها، ساختار مونومرها به شکل بسته شده با هیدروژن رسم و بهینه‌سازی شدند. بدین منظور ساختار شیمیایی تمام ترکیبات با استفاده از نرم‌افزار Hyperchem رسم و سپس ساختارها با استفاده از روش نیمه تجربی AM1 بهینه گردید. بهینه‌سازی تا زمانی ادامه یافت که حد گرادینانی به ۰/۰۰۱ کیلوکالری بر آنگستروم مول برسد.

### ۶-۳-۲- محاسبه‌ی توصیف‌کننده‌های مولکولی

ساختارهای سه بعدی رسم شده برای ۱۴۴ ترکیب موجود در مجموعه‌ی داده‌ها از مرحله‌ی بالا وارد نرم‌افزار Dragon شده و تعداد ۱۲۶۱ توصیف‌کننده برای ترکیبات مورد مطالعه محاسبه گردید. لازم به ذکر است توصیف‌کننده‌هایی که برای تمام ترکیبات مقدار ثابتی داشته و همچنین آنهایی که دارای همبستگی بیش از ۹۵٪ با یکدیگر بودند، حذف شدند به طوری که تنها ۶۰۵ توصیف‌کننده برای مدل‌سازی باقی ماندند.

### ۶-۳-۳- مدل الگوریتم ژنتیکی بر اساس رگرسیون خطی چندگانه

همانطور که پیش‌تر توضیح داده شد، هدف از اجرای الگوریتم حاضر شناسایی ترکیبی از توصیف‌کننده‌ها است به گونه‌ای که برای پیش‌بینی ضریب شکست پلیمرها بوسیله‌ی شبکه‌ی عصبی مصنوعی مورد استفاده قرار گیرند. در فرآیند اجرای این الگوریتم، ابتدا به صورت اتفاقی یک جمعیت اولیه از توصیف‌کننده‌ها تولید می‌شود، برازندگی یا شایستگی تک تک اعضای هر نسل یعنی توصیف‌کننده‌ها، محاسبه و با توجه به شایستگی‌ها، نسل‌های بعدی با اعمال سه عملگر انتخاب، پیوند و جهش تولید یا باز ترکیب می‌شوند. این روند جستجو برای جمعیت‌های مختلف تا حصول ملاک خاتمه ادامه می‌یابد. در ادامه به مراحل انتخاب و بهینه‌سازی با این روش خواهیم پرداخت.

### ۶-۳-۳-۱- رمزبندی

با توجه به اینکه روش رمزبندی دودویی در مسائل بسیاری بکار گرفته شده است و به دلایلی از قبیل قابلیت تطابق با اکثر سیستم‌های موجود، ماهیت مسأله پیش رو و سادگی در عملیات، در الگوریتم استفاده شده در این پروژه از این روش استفاده شده است. پس کروموزوم ارائه شده شامل

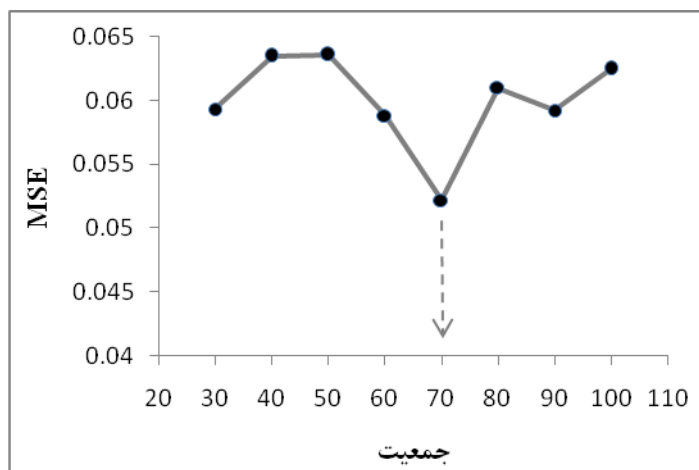
۶۰۵ بیت (صفر و یک) خواهد بود که یک‌های آن نشانه‌ی وجود و صفرهای آن نشانه‌ی عدم وجود یک توصیف‌کننده است.

### ۶-۳-۳-۲- تابع شایستگی

همانطور که در فصل سوم بیان شد، معیار انتخاب کروموزوم‌ها برای جامعه جدید از جامعه اولیه، توابع برازش یا شایستگی هستند که برای آنها تعریف می‌شوند. تابع هدف برابر با معکوس اختلاف میانگین مربعات خطا برای سری آموزش و آزمون، و تابع شایستگی برابر با مجموع MSE برای دو سری می‌باشد، که باید مینیمم شود. با وجودی که ملاک اصلی، شایستگی کروموزوم‌ها است، استفاده از این تابع شایستگی باعث می‌شود فاصله‌ی  $MSE_{train}$  و  $MSE_{test}$  خیلی زیاد نشود.

### ۶-۳-۳-۳- جمعیت

این برنامه کروموزوم ساخته شده به وسیله‌ی الگوریتم ژنتیکی را گرفته و بر اساس آن توصیف‌کننده‌های منتخب را به‌طور خودکار از مجموعه‌ی داده‌ها انتخاب می‌کند. هم‌چنین به صورت تصادفی ترکیبات را به دو سری آموزش و آزمون تقسیم می‌نماید. اکثر محققان برای اندازه جمعیت عددی بین ۵۰ تا ۱۰۰ پیشنهاد کرده‌اند. این مجموعه از جمعیت‌های انتخاب شده اولین نسل را ایجاد می‌کند و از نسل اولیه نیز نسل‌های بعدی ایجاد می‌شود. سرانجام در نسل نهایی، یک جمعیت انتخاب شده که موجب حل بهینه مسأله است بدست می‌آید. در این مطالعه به منظور کاهش زمان همگرایی به پاسخ مناسب، به‌طور حدسی از یک جامعه اولیه که دارای اندازه‌ی بزرگتری نسبت به جامعه تولیدی بعدی می‌باشد، استفاده شده است. این روش سبب می‌شود که نقاط بیشتری در بازه تعریف شده جستجو شود و از میان این نقاط، مجموعه‌ای که دارای خطای کمتری است انتخاب شود. این امر تا حدودی کاهش زمان انجام محاسبات را در پی دارد [۷۱]. استفاده از تعداد ماکزیمم نسل نیز از پیشرفت نامحدود و نامشخص الگوریتم وقتی که همگرایی نتواند حاصل شود جلوگیری می‌کند.



شکل (۶-۱) - تغییرات مقادیر خطا بر حسب تعداد جمعیت

#### ۴-۳-۳-۶- روش انتخاب

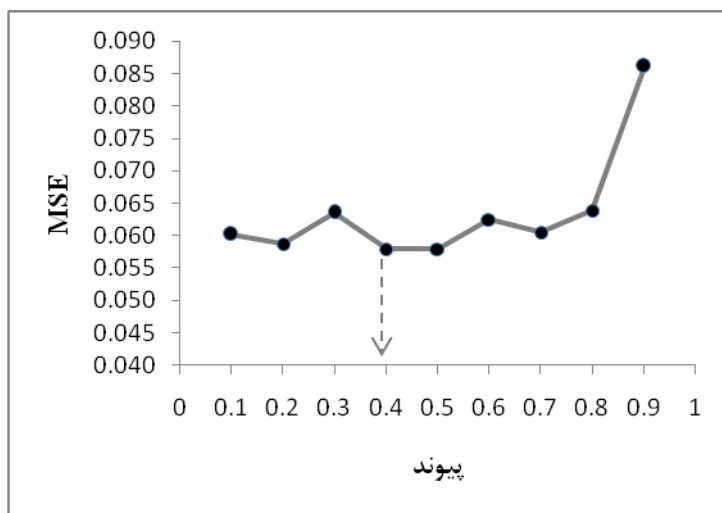
عملگر انتخاب مورد استفاده ترکیبی از روش‌های انتخاب چرخه‌ی رولت و نخبه‌گرایی<sup>۱</sup> است. بحث نخبه‌گرایی بسیار حائز اهمیت می‌باشد و می‌توان گفت به نوعی، حافظه‌دار کردن الگوریتم ژنتیک می‌باشد. بدین معنا که با نگهداری از بهترین جواب حاصله در الگوریتم در طی تکرارهای مختلف، عملکرد بهبود خواهد یافت.

#### ۵-۳-۳-۶- عملگر پیوند

عملگر پیوند می‌تواند دو نقطه‌ای باشد که در این صورت، دو نقطه به‌طور تصادفی انتخاب می‌شوند و پیوند بین آن دو اتفاق می‌افتد. این عملیات در چند نقطه انجام شد، که به آن بازترکیبی چند نقطه‌ای<sup>۲</sup> می‌گویند. معمولاً احتمال پیوند برای هر زوج کروموزوم ۰/۱ تا ۰/۹ در نظر گرفته می‌شود [۷۲]. برای یافتن بهترین احتمال این عملگر، مقدار آن بر حسب MSE رسم گردید.

<sup>۱</sup>- Elitism

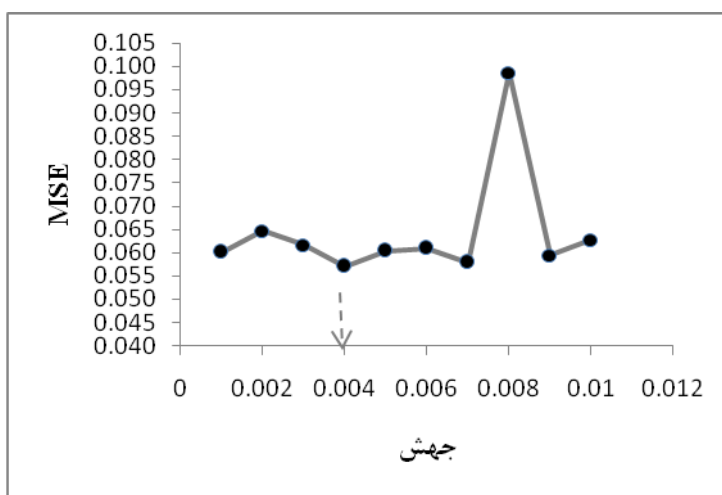
<sup>۲</sup>- Multipoint Crossover



شکل (۲-۶) - تغییرات خطا بر حسب مقادیر عملگر پیوند

### ۶-۳-۳-۶ - عملگر جهش

جهش در جمعیت تنوع ایجاد می کند و به پرهیز از همگرایی زودرس کمک می کند. احتمال وقوع این عملگر معمولاً بین ۰/۰۰۵ تا ۰/۰۱ انتخاب می شود (۵٪ تا ۱ درصد). احتمال این عملگر از ۰/۰۰۱ تا ۰/۰۱ بر حسب MSE بررسی گردید.



شکل (۳-۶) - تغییرات خطا بر حسب مقادیر عملگر جهش

الگوریتم استفاده شده سری داده‌ها را به صورت تصادفی به سری آموزشی (با ۱۰۴ ترکیب) و سری آزمون (با ۴۰ ترکیب) تفکیک کرده و تعداد توصیف‌کننده‌های خروجی از الگوریتم بعد از هر تکرار چهار در نظر گرفته شد. مقادیر بهینه شده‌ی پارامترهای الگوریتم ژنتیک در جدول آمده است.

جدول (۲-۶) - مقادیر پارامترهای الگوریتم ژنتیک

جمعیت	۷۰
نرخ پیوند	۰/۴
نرخ جهش	۰/۰۰۴

در ۱۵۰ تکرار انجام شده، علاوه بر مراحل بهینه‌سازی، چهار توصیف‌کننده انتخاب شدند که این توصیف‌کننده‌ها و مقادیر عددی آنها برای سری داده‌ها در جدول (۳-۶) آورده شده است.

جدول (۳-۶) - مقادیر توصیف‌کننده‌های انتخاب شده جهت مدل‌سازی با ANN و MLR

شماره	R <sup>2</sup> <sub>e</sub>	R <sup>2</sup> <sub>p</sub>	J <sub>3D</sub>	Mor <sub>15e</sub>	شماره	R <sup>2</sup> <sub>e</sub>	R <sup>2</sup> <sub>p</sub>	J <sub>3D</sub>	Mor <sub>15e</sub>
1	2/639	0/491	6/165	-0/975	20	1/528	0/475	4/103	-0/132
2t	2/141	0/384	4/298	0/117	21	1/66	0/717	3/662	-0/664
3v	2/103	0/832	3/76۰	-0/506	22	1/996	0/797	2/658	-0/549
4t	1/914	0/724	4/922	-0/295	23	1/775	0/73۰	3/268	-0/518
5	1/798	0/537	5/61۰	-0/324	24	2/53	0/463	5/303	0/124
6v	2/038	0/747	3/691	0/327	25	1/86۰	0/845	3/498	-0/203
7	1/898	0/789	4/296	-0/731	26	1/76۰	0/703	4/473	-0/061
8	1/842	0/554	5/700	-0/138	27t	1/705	0/704	3/208	-0/393
9t	1/883	0/597	3/623	-0/08۰	28	1/881	0/801	3/44۰	-0/696
10	1/887	0/738	6/177	-0/344	29v	1/859	0/627	3/352	0/482
11	2/119	0/947	3/887	-0/876	30	1/706	0/692	3/157	-0/236
12v	1/879	0/672	5/79۰	-0/341	31t	1/801	0/806	5/626	-0/174
13t	2/484	0/494	5/767	0/236	32	1/606	0/583	2/823	-0/252
14	1/66۰	0/545	5/300	0/162	33	1/736	0/742	3/309	-0/295
15	1/757	0/716	3/596	-0/118	34	1/804	0/661	3/367	-0/091
16	1/645	0/562	2/727	0/261	35v	2/452	0/502	5/743	0/006
17t	2/281	0/755	3/553	0/019	36	1/814	0/756	3/501	-0/078
18v	1/871	0/617	3/130	0/582	37	1/801	0/53۰	3/399	0/704
19	1/781	0/672	3/55	-0/261	38	1/591	0/539	3/324	0/049

ادامه‌ی جدول (۳-۶)

شماره	R 2 e	R 2 p	J <sup>3</sup> D	Mor15 e	شماره	R 2 e	R 2 p	J <sup>3</sup> D	Mor15 e
39	1/53·	0/577	3/013	-0/065	74	1/945	0/613	7/391	-0/152
40t	1/713	0/512	3/266	0/368	75v	1/94·	0/597	6/761	-0/969
41v	1/706	0/775	۳/۰۴۰	0/278	76	1/879	0/525	5/736	-0/662
42	1/691	0/683	3/279	0/06·	77	1/775	0/467	4/777	-0/313
43	1/767	0/529	3/188	0/396	78	1/936	0/463	4/676	-0/083
44v	1/921	0/934	3/435	0/003	79	2/774	0/303	3/44·	-0/05·
45v	1/839	1/105	3/411	-0/195	80v	1/981	0/406	4/372	-0/101
46	2/472	0/487	5/552	0/198	81	1/938	0/505	5/060	-0/056
47	1/505	0/61·	2/975	0/017	82t	1/946	0/486	5/314	-0/179
48	1/795	0/908	3/043	0/167	83v	1/902	0/529	5/47·	-0/871
49	1/814	0/706	3/305	-0/292	84	1/823	0/459	4/594	-0/177
50v	1/459	0/583	2/905	-0/178	85t	1/876	0/596	6/344	-0/542
51t	1/624	0/703	2/721	-0/372	86	1/888	0/601	6/569	-0/781
52	2/239	0/463	4/96·	-0/147	87	1/869	0/492	4/604	-0/294
53	2/365	0/515	5/536	0/411	88v	1/752	0/367	4/056	-0/356
54	2/086	0/521	6/059	0/165	89t	1/999	0/535	5/835	-0/361
55v	2/170	0/523	6/028	-0/397	90	2/537	0/462	5/677	-0/529
56t	2/34	0/482	5/424	-0/108	91	1/876	0/604	6/332	-0/624
57	2/569	0/452	5/511	-0/087	92t	2/007	0/644	6/607	-0/746
58	2/523	0/491	3/349	0/077	93	1/866	0/589	5/929	-0/649
59	2/372	0/35·	3/633	-0/02·	94	2/135	0/672	8/461	0/198
60	2/268	0/51·	5/772	-0/219	95v	1/997	0/583	6/075	-0/486
61	1/888	0/54·	5/975	-0/573	96t	1/966	0/592	6/142	-0/54·
62	1/776	0/442	4/816	-0/327	97v	2/026	0/719	4/543	-0/5··
63	1/849	0/503	5/543	-0/472	98	1/833	0/526	5/502	-0/628
64t	1/883	0/525	5/815	-0/587	99t	1/978	0/614	7/073	-0/631
65v	1/884	0/534	6/036	-0/643	100v	1/975	0/63·	8/46·	-0/676
66	1/848	0/553	6/337	-0/256	101	2/507	0/456	5/486	-0/261
67	2/096	0/529	3/274	-0/179	102	1/809	0/784	2/609	-0/002
68v	2/547	0/426	5/368	0/204	103t	1/866	0/464	5/256	-0/336
69v	1/909	0/56·	6/367	-0/717	104	1/873	0/612	7/088	-0/496
70v	1/923	0/577	7/542	-0/792	105t	1/746	0/379	3/962	-0/007
71	1/927	0/58·	6/596	-0/917	106v	2/042	0/598	6/506	-0/5··
72t	1/914	0/493	5/273	-0/588	107	2/097	0/792	4/166	-0/958
73t	1/822	0/531	5/422	-0/427	108	2/215	0/792	4/375	-1/06·

ادامه‌ی جدول (۳-۶)

شماره	R 2 e	R 2 p	J <sup>3</sup> D	Mor15 e	شماره	R 2 e	R 2 p	J <sup>3</sup> D	Mor15 e
109	2/126	0/808	4/177	-1/1··	127	1/61·	0/45·	4/91·	-0/212
110t	1/584	0/407	4/203	0/082	128	1/939	0/642	6/466	-0/309



111	2/053	0/690	3/054	0/362		129	1/584	0/407	4/203	0/082
112v	2/68	0/443	5/452	0/412		130v	1/940	0/625	4/657	-0/085
113	2/002	0/689	6/508	-0/515		131	1/919	0/627	5/805	-0/302
114	1/842	0/578	6/901	-0/319		132	1/986	0/616	7/038	-0/524
115t	1/779	0/611	4/748	-0/317		133	2/116	0/870	3/944	-0/916
116	1/750	0/527	6/592	-0/019		134t	2/559	0/466	5/524	0/432
117v	2/108	0/783	4/258	-0/986		135v	1/667	0/530	5/370	-0/299
118	1/965	0/567	5/443	-0/239		136	1/548	0/464	3/805	-0/221
119	1/744	0/552	7/395	0/381		137	1/562	0/521	4/680	-0/251
120t	2/029	0/724	3/808	-0/734		138	1/730	0/511	5/524	-0/289
121	2/108	0/703	3/574	-0/348		139	1/612	0/419	5/133	-0/030
122v	1/869	0/410	3/833	-0/214		140v	2/000	0/584	6/135	-0/069
123	2/552	0/425	4/839	-0/266		141t	1/605	0/418	4/190	-0/023
124t	2/088	0/754	4/511	-0/539		142	1/859	0/715	5/869	-0/246
125v	1/797	0/520	5/512	-0/474		143v	1/883	0/687	6/025	0/110
126	1/931	0/581	4/462	-0/317		144	1/695	0/457	4/160	-0/284

t = سری /آزمون

v = سری ارزیابی

### ۶-۳-۴- مدل رگرسیون خطی چندگانه

به منظور بدست آوردن ضرایب توصیف‌کننده‌ها برای بررسی ارتباط ساختاری ترکیبات با ضریب شکست آنها و کارایی توصیف‌کننده‌های منتخب الگوریتم ژنتیک از روی پارامترهای آماری، مدل رگرسیون خطی چندگانه توسط نرم افزار SPSS ایجاد گردید. برای انجام مقایسه بین مدل حاصل از این روش و مدل بدست آمده از شبکه عصبی مصنوعی، داده‌ها به‌طور تصادفی به سری آموزشی، آزمون و ارزیابی تقسیم شدند. با این چهار توصیف‌کننده و با استفاده از روش اجباری، رگرسیون بر روی سری آموزش انجام گرفت. توصیف‌کننده‌های مدل MLR به همراه پارامترهای آماری مربوطه در جدول (۴-۶) آمده است.

جدول (۴-۶) - مشخصات مدل حاصل از روش MLR

نماد توصیف‌کننده	مفهوم	اثر متوسط
R 2 e	خود ارتباط R با تاخیر ۲ / وزن دار شده با الکترونگاتیویته اتمی ساندerson <sup>۱</sup>	-۰/۲۰۱
R 2 p	خود ارتباط R با تاخیر ۲ / وزن دار شده با قطبش پذیری اتمی <sup>۱</sup>	۰/۰۹۳

<sup>۱</sup> - R autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities

J3D	شاخص بالابان سه بعدی <sup>۲</sup>	-۰/۰۳۳
Mor15 e	سیگنال ۱۵ 3D-MORSE / وزن دار شده با الکترونگاتیویته اتمی ساندرسون <sup>۴</sup>	-۰/۰۰۵
ضریب همبستگی		0/956
مقدار آماره F		215/966
خطای استاندارد		0/020
میانگین مربعات خطا		۰/۱۸۸

معادله‌ی خطی حاصل از این مدل بصورت رابطه‌ی (۱-۶) می‌باشد:

$$RI = 1/719 - 0/155 R_{2e} + 0/235 R_{2p} - 0/01 J_{3D} + 0/032 Mor_{15e} \quad (1-6)$$

ماتریس همبستگی متغیرهای مستقل حاصل از الگوریتم ژنتیکی برای سری داده‌ها در جدول (۵-۶) ضریب همبستگی پایین میان آنها را نشان می‌باشد.

جدول (۵-۶) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده با الگوریتم ژنتیک

توصیف‌کننده	R 2 e	R 2 p	J3D	Mor15 e
R 2 e	1/000			
R 2 p	-0/198	1/000		
J3D	0/208	-0/255	1/000	
Mor15 e	-0/117	-0/339	-0/158	1/000

### ۵-۳-۶- مدل شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های الگوریتم ژنتیک

در این بخش نیز از شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی استفاده شد. الگوریتم آموزشی لونیبرگ - مارکورت و برای لایه‌ی خروجی، تابع انتقال خطی انتخاب گردید.

<sup>۲</sup>- R autocorrelation of lag 2 / weighted by atomic polarizabilities

<sup>۳</sup>- 3D-Balaban index

<sup>۴</sup>- Molecule representation of structures based on electron diffraction

<sup>۵</sup>- 3D-MORSE-signal 15 / weighted by atomic Sanderson electronegativities

### 6-3-5-1- انتخاب تابع انتقال

برای بهینه‌سازی و شناسایی بهترین پارامترهای شبکه، ابتدا شبکه‌هایی با لگاریتم سیگموئید و تانژانت سیگموئید با تعداد گره‌های مختلف در لایه پنهان و تعداد دوره‌های آموزش یکسان و برابر ۲۰ ایجاد و هر شبکه با الگوریتم آموزشی لونبرگ - مارکورت آموزش داده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار مجذور میانگین مربعات خطا به عنوان معیار انتخاب گردید. با توجه به جداول (۶-۶) و (۷-۶) و نمودارهای (۴-۶) و (۵-۶)، تابع تبدیل لگاریتم سیگموئید نسبت به تانژانت سیگموئید دارای کمترین مقدار مجذور میانگین مربعات خطا می‌باشد.

جدول (۶-۶) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئید و تعداد گره‌های مختلف در لایه پنهان

	تعداد گره						
	2	3	4	5	6	7	8
سری آموزش	0/0188	0/0188	0/0178	0/0177	<b>0/0175</b>	0/0152	0/0163
سری آزمون	0/0237	0/0226	0/024	0/0401	<b>0/0215</b>	0/0258	0/0314

جدول (۷-۶) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل تانژانت سیگموئید و تعداد گره‌های مختلف در لایه پنهان

	تعداد گره						
	2	3	4	5	6	7	8
سری آموزش	0/0188	0/0183	0/0176	0/017	0/0152	0/015	0/0153
سری آزمون	0/0235	0/0231	0/0244	0/0248	0/0281	0/0311	0/0913



شکل (۴-۶) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل لگاریتم سیگموئید



شکل (۵-۶) - تغییرات مقادیر RMSE بر حسب تعداد گره‌ها برای سری آموزش و آزمون با تابع تبدیل تانژانت سیگموئید

### 6-3-5-2- انتخاب تعداد گره‌های لایه پنهان و تعداد دوره‌های آموزش

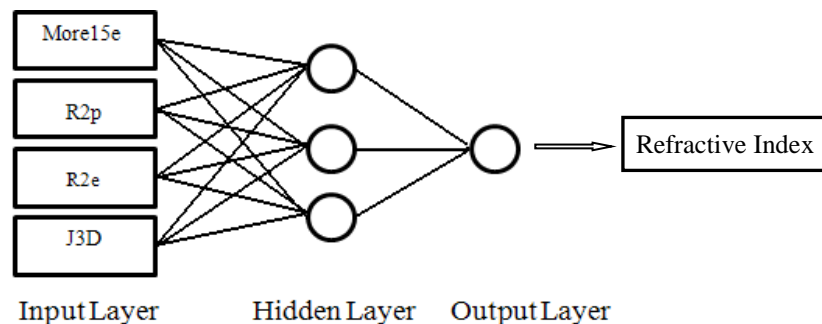
برای بهینه‌سازی دو پارامتر دیگر تعداد گره‌های لایه پنهان از ۲ تا ۸ در تعداد دوره‌های آموزش مختلف، از ۵ تا ۲۰۰، تغییر داده شدند و از میان شبکه‌های آموزش دیده، با توجه به مجذور میانگین مربعات خطا برای سری‌های آموزش و آزمون، مدلی با تابع انتقال لگاریتم سیگموئیدی و لایه‌ی پنهان حاوی سه نرون و تعداد دوره‌های آموزشی ۴۷ بدست آمد. نتایج حاصله در جدول (۶-۸) آمده است. در تعداد گره‌های بالا کاهش قابلیت تعمیم و افزایش خطا برای سری ارزیابی را داریم، بنابراین از تعداد گره‌های کمتر استفاده شد.

جدول (۶-۸) - مقادیر مجذور میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل لگاریتم سیگموئیدی و تعداد گره‌های مختلف در لایه پنهان و چرخه‌های مختلف

		تعداد دوره‌های آموزشی						
سری آموزش		37	39	41	43	45	47	49
تعداد گره‌ها	2	0/0187	0/0187	0/0187	0/0187	0/0187	0/0187	0/0187
	3	0/0181	0/0180	0/0180	0/0180	0/0180	<b>0/0179</b>	0/0179
	4	0/0177	0/0177	0/0177	0/0177	0/0177	0/0177	0/0177
	5	0/0173	0/0173	0/0173	0/0173	0/0172	0/0172	0/0171
	6	0/0160	0/0160	0/0160	0/0160	0/0159	0/0159	0/0159
	7	0/0150	0/0149	0/0149	0/0148	0/0147	0/0146	0/0146
	8	0/0158	0/0158	0/0158	0/0158	0/0157	0/0157	0/0156
سری آزمون		37	39	41	43	45	47	49
تعداد گره‌ها	2	0/0236	0/0236	0/0236	0/0236	0/0236	0/0236	0/0236
	3	0/0227	0/0227	0/0227	0/0227	0/0227	<b>0/0227</b>	0/0227
	4	0/0242	0/0242	0/0242	0/0242	0/0242	0/0242	0/0242
	5	0/0399	0/0397	0/0395	0/0392	0/0389	0/0383	0/0375
	6	0/0260	0/0263	0/0264	0/0261	0/0257	0/0223	0/0236
	7	0/0243	0/0244	0/0245	0/0245	0/0246	0/0248	0/0250
	8	0/0327	0/0327	0/0359	0/0447	0/0550	0/0620	0/0675

### ۶-۳-۵-۳ - معماری شبکه‌ی عصبی مصنوعی بهینه شده

با توجه به نتایج حاصل در بهینه‌سازی شبکه، شبکه‌ی عصبی پیشخور سه لایه و آموزش پس انتشار با الگوریتم آموزشی لونبرگ - مارکورت و تابع انتقال لگاریتم سیگموئیدی انتخاب گردید. ساختار هندسی شبکه‌ی بهینه شده در شکل (۶-۶) و پارامترهای آن در جدول (۶-۹) آمده است.



شکل (۶-۶) - ساختار هندسی شبکه‌ی عصبی استفاده شده

جدول (۶-۹) - پارامترهای شبکه‌ی بهینه

۳	تعداد نرون‌های لایه‌ی پنهان
۴۷	تعداد دوره‌های آموزش
۰/۰۵	ممنتوم ( $\mu$ )
۰/۱	فاکتور کاهش برای پارامتر ممنتوم
۱۰	فاکتور افزایش برای پارامتر ممنتوم
$e^{-5}$	مقدار ماکزیمم برای پارامتر ممنتوم

### ۶-۳-۵-۴- فرایند ایجاد تصادفی Y-

برای بررسی احتمال همبستگی تصادفی مدل بدست آمده، از فرایند ایجاد تصادفی Y- استفاده شد. در این آزمون، ضریب شکست تجربی ترکیبات به صورت تصادفی جابجا و مدل‌های جدیدی ایجاد شدند. مقادیر پایین ضریب تعیین در این آزمون، برای این مدل‌ها که در جدول (۶-۱۰) آمده است بیان‌گر تصادفی نبودن مدل اصلی می‌باشد.

جدول (۶-۱۰) - نتایج تصادفی کردن Y-

مدل	۱	۲	۳	۴	۵	۶
$R^2$ برای سری آزمون	۰/۰۰۵	۰/۰۰۳	۰/۱۰۴	۰/۰۳۸	۰/۰۰۲	۰/۰۷۰
$R^2$ برای سری ارزیابی	۰/۰۰۱	۰/۰۰۱	۰/۰۷۲	۰/۰۱۰	۰/۰۶۳	۰/۲۰۹

مقادیر تجربی و پیش‌بینی شده‌ی ضریب شکست پلیمرها بوسیله‌ی روش‌های GA-MLR و

GA-ANN در جدول (11-6) آمده است.

جدول (11-6) - مقادیر تجربی و پیش‌بینی شده‌ی ضریب شکست بوسیله‌ی روش‌های GA-MLR و GA-ANN

شماره	Refractive Index			شماره	Refractive Index		
	Exp	GA-MLR	GA-ANN		Exp	GA-MLR	GA-ANN
1	1/339	1/331	1/340	31t	1/574	1/567	1/581
2t	1/375	1/438	1/426	32	1/578	1/571	1/574
3v	1/532	1/534	1/532	33	1/582	1/581	1/583
4t	1/533	1/533	1/540	34	1/585	1/558	1/563
5	1/537	1/499	1/495	35v	1/390	1/399	1/392
6v	1/538	1/552	1/560	36	1/586	1/578	1/582
7	1/540	1/543	1/544	37	1/587	1/553	1/566
8	1/540	1/501	1/500	38	1/587	1/567	1/575
9t	1/540	1/528	1/531	39	1/591	1/585	1/589
10	1/542	1/526	1/544	40t	1/593	1/553	1/563
11	1/542	1/546	1/542	41v	1/596	1/615	1/608
12v	1/543	1/516	1/525	42	1/597	1/586	1/590
13t	1/380	1/399	1/383	43	1/597	1/550	1/560
14	1/548	1/541	1/550	44v	1/600	1/606	1/602
15	1/549	1/575	1/579	45v	1/608	1/653	1/621
16	1/550	1/577	1/584	46	1/392	1/401	1/390
17t	1/550	1/508	1/506	47	1/610	1/600	1/599
18v	1/552	1/561	1/572	48	1/612	1/629	1/613
19	1/554	1/557	1/561	49	1/615	1/561	1/564
20	1/556	1/548	1/554	50v	1/625	1/595	1/595
21	1/557	1/572	1/573	51t	1/657	1/593	1/591
22	1/558	1/552	1/550	52	1/407	1/426	1/420
23	1/559	1/566	1/567	53	1/412	1/431	1/422
24	1/385	1/386	1/379	54	1/418	1/462	1/446
25	1/562	1/587	1/588	55v	1/419	1/432	1/425
26	1/567	1/564	1/572	56t	1/419	1/411	1/405
27t	1/568	1/575	1/577	57	1/346	1/368	1/350
28	1/568	1/558	1/557	58	1/420	1/412	1/428
29v	1/571	1/560	1/570	59	1/420	1/396	1/404
30	1/571	1/578	1/581	60	1/437	1/422	1/415

ادامه‌ی جدول (11-6)

شماره	Refractive Index			شماره	Refractive Index		
	Exp	GA-MLR	GA-ANN		Exp	GA-MLR	GA-ANN
61	1/451	1/474	1/468	96t	1/484	1/474	1/477
62	1/454	1/488	1/477	97v	1/485	1/512	1/512

63	1/456	1/479	1/469
64t	1/458	1/473	1/464
65v	1/459	1/471	1/463
66	1/459	1/490	1/484
67	1/460	1/480	1/474
68v	1/348	1/377	1/346
69v	1/461	1/467	1/465
70v	1/463	1/454	1/452
71	1/463	1/460	1/469
72t	1/463	1/466	1/456
73t	1/463	1/493	1/487
74	1/464	1/482	1/485
75v	1/464	1/459	1/477
76	1/465	1/472	1/464
77	1/467	1/495	1/487
78	1/467	1/478	1/469
79	1/356	1/324	1/352
80v	1/467	1/460	1/448
81	1/469	1/484	1/478
82t	1/471	1/472	1/460
83v	1/471	1/465	1/460
84	1/472	1/492	1/485
85t	1/473	1/486	1/490
86	1/473	1/476	1/485
87	1/503	1/489	1/482
88v	1/476	1/481	1/469
89t	1/477	1/464	1/457
90	1/356	1/360	1/351
91	1/477	1/486	1/492
92t	1/481	1/468	1/492
93	1/483	1/487	1/489
94	1/483	1/466	1/482
95v	1/483	1/469	1/471

98	1/485	1/482	1/475
99t	1/487	1/465	1/481
100v	1/489	1/453	1/473
101	1/360	1/374	1/363
102	1/630	1/597	1/595
103t	1/489	1/475	1/461
104	1/490	1/485	1/494
105t	1/500	1/497	1/491
106v	1/490	1/461	1/469
107	1/495	1/507	1/501
108	1/497	1/483	1/482
109	1/498	1/502	1/495
110t	1/500	1/529	1/533
111	1/500	1/544	1/551
112v	1/360	1/366	1/334
113	1/500	1/488	1/510
114	1/500	1/489	1/486
115t	1/502	1/529	1/532
116	1/503	1/504	1/493
117v	1/503	1/501	1/496
118	1/505	1/485	1/483
119	1/506	1/516	1/500
120t	1/507	1/512	1/506
121	1/510	1/510	1/506
122v	1/510	1/480	1/471
123	1/364	1/366	1/372
124t	1/511	1/510	1/509
125v	1/512	1/492	1/484
126	1/512	1/501	1/498
127	1/513	1/519	1/516
128	1/514	1/494	1/505
129	1/515	1/529	1/533
130v	1/517	1/515	1/518

ادامی جدول (۶-۱۱)

شماره	Refractive Index		
	Exp	GA-MLR	GA-ANN
131	1/517	1/500	1/505
132	1/517	1/468	1/482
133	1/518	1/526	1/520
134t	1/367	1/390	1/368
135v	1/520	1/521	1/522
136	1/520	1/543	1/547
137	1/520	1/544	1/549

شماره	Refractive Index		
	Exp	GA-MLR	GA-ANN
138	1/520	1/506	1/501
139	1/521	1/515	1/509
140v	1/525	1/482	1/481
141t	1/527	1/525	1/527
142	1/527	1/531	1/545
143v	1/527	1/531	1/546
144	1/529	1/512	1/509



## ۴-۶- بررسی نتایج

### ۴-۶-۱- ارزیابی مدل های GA-ANN و GA-MLR

پارامترهای آماری حاصل از مدل های الگوریتم ژنتیک بر اساس شبکه ی عصبی و رگرسیون

خطی چندگانه در جدول (۴-۱۲) آورده شده است:

جدول (۴-۱۲)- پارامترهای آماری مدل GA-ANN و GA-MLR

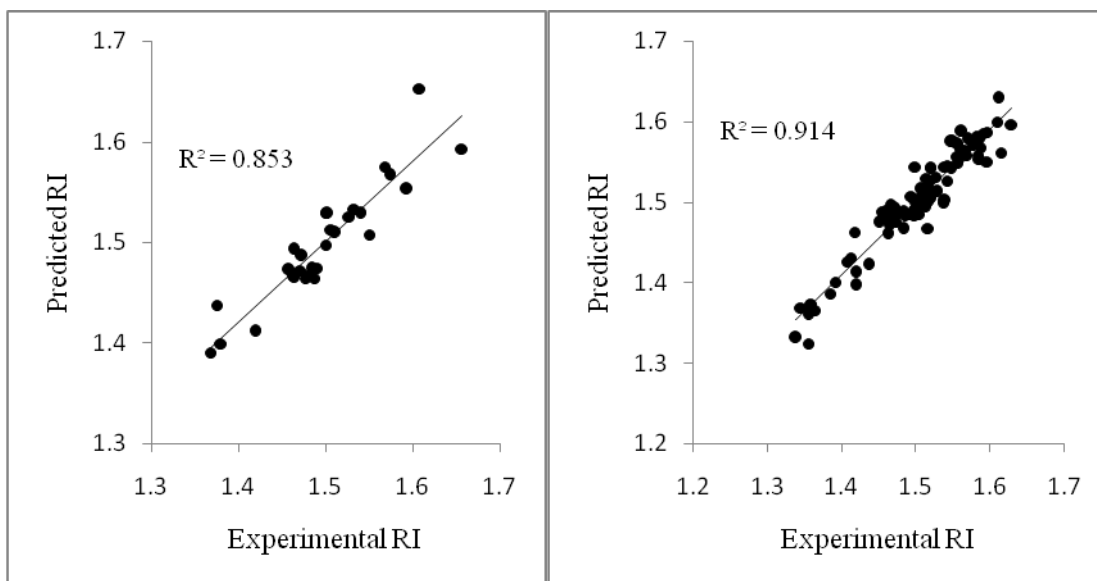
GA-ANN	R	SE	F	RMSE
سری آموزش	0/964	0/018	1103/211	0/018
سری آزمون	0/939	0/023	200/750	0/023
سری ارزیابی	0/965	0/017	364/136	0/018
GA-MLR	R	SE	F	RMSE
سری آموزش	0/956	0/020	215/966	0/020
سری آزمون	0/924	0/026	157/166	0/025
سری ارزیابی	0/958	0/019	304/012	0/019

رابطه ی خطی ضریب شکست پیش بینی شده در دو روش بر حسب مقادیر واقعی نیز در شکل های

(۴-۷) و (۴-۸) نشان داده شده است. با توجه به پارامترهای آماری بهتر و نمودار خطی متراکم تر،

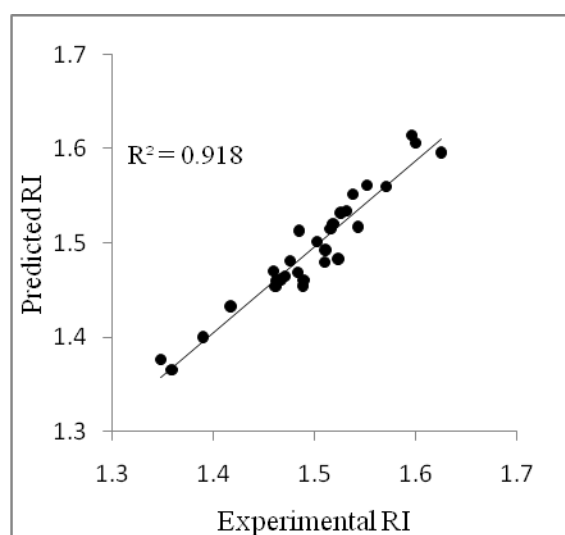
می توان بیان کرد که توانایی مدل حاصل از روش شبکه عصبی مصنوعی در پیش بینی ضریب شکست

سری داده ها بیشتر می باشد.



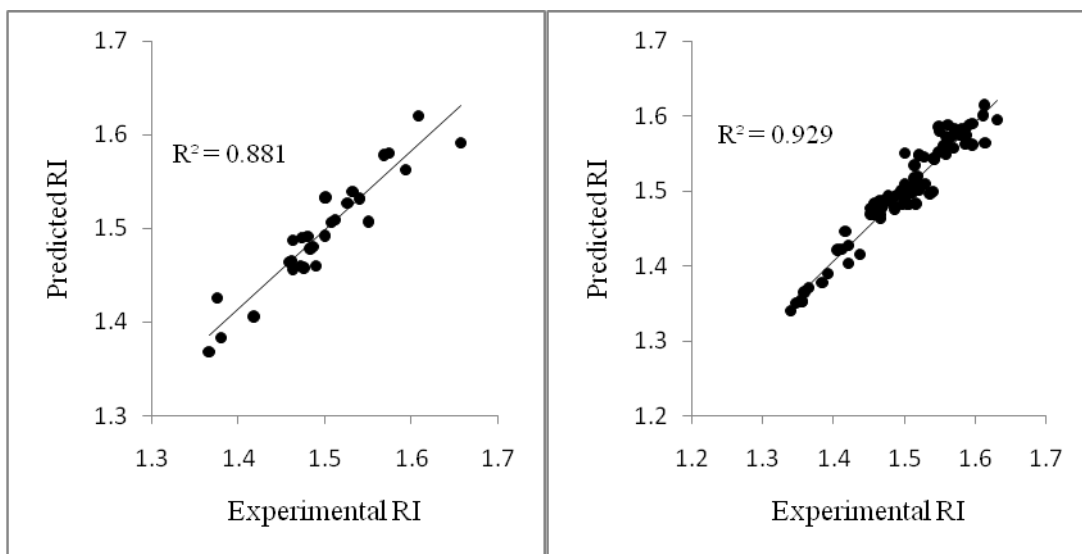
(ب)

(الف)



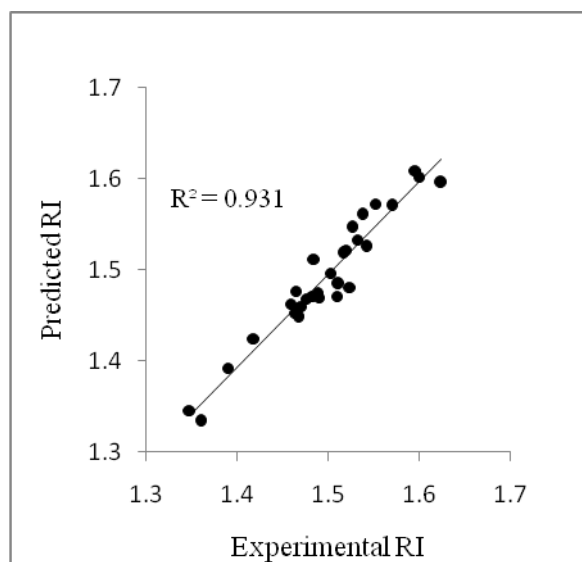
(ج)

شکل (7-6) - نمودار مقادیر پیش‌بینی شده‌ی ضریب شکست بر حسب مقادیر واقعی با روش GA-MLR برای  
 الف) سری آموزش، ب) سری آزمون و ج) سری ارزیابی



(ب)

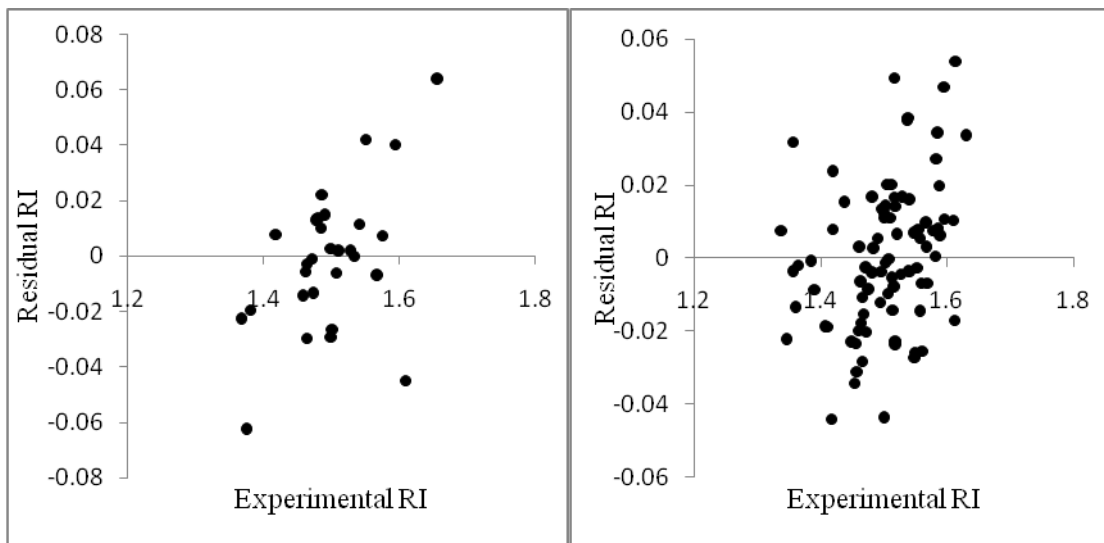
(الف)



(ج)

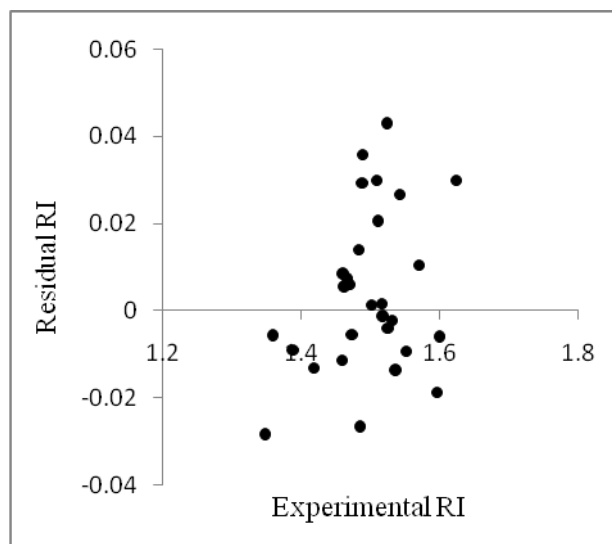
شکل (6-8) - نمودار مقادیر پیش‌بینی شده‌ی ضریب شکست بر حسب مقادیر واقعی با روش GA-ANN برای (الف) سری آموزش، (ب) سری آزمون و (ج) سری ارزیابی

همچنین به منظور بررسی وجود خطای معین در این مدل‌ها باقیمانده مقادیر محاسبه شده از ضریب شکست تجربی بر حسب مقادیر تجربی برای ترکیبات در شکل‌های (6-9) و (6-10) رسم شده است.



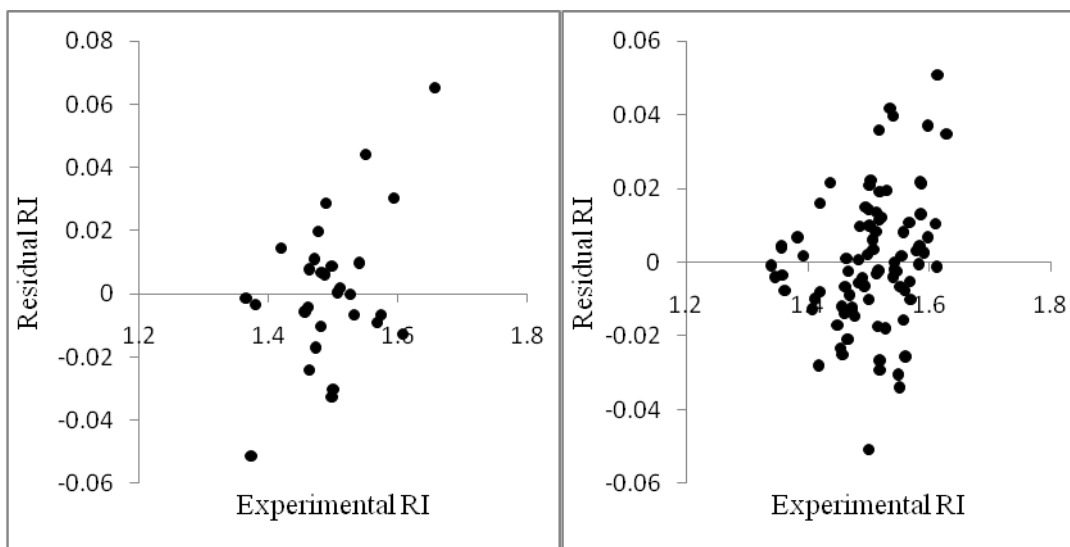
(ب)

(الف)



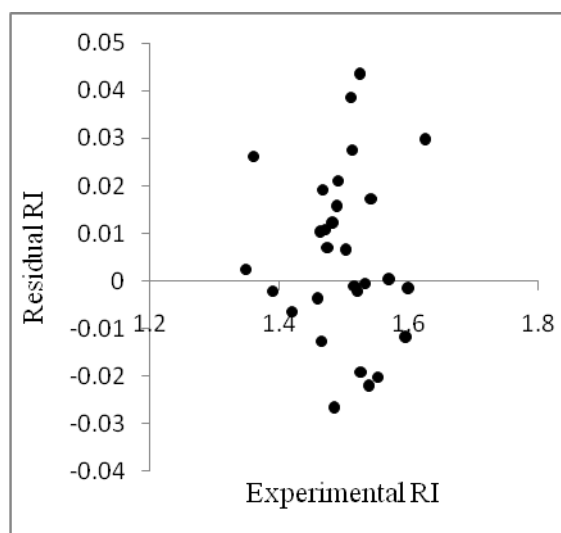
(ج)

شکل (۶-۹) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی ضریب شکست با روش GA-MLR از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، ب) سری آزمون و ج) سری ارزیابی



(ب)

(الف)



(ج)

شکل (6-10) - نمودار باقیمانده‌ی مقادیر پیش‌بینی شده‌ی ضریب شکست با روش GA-ANN از مقادیر واقعی بر حسب مقادیر تجربی برای الف) سری آموزش، ب) سری آزمون و ج) سری ارزیابی

پراکندگی نقاط باقیمانده‌ها در دو طرف خط صفر نشان‌دهنده میزان خطای معین در مقادیر پیش‌بینی شده‌ی روش استفاده شده است. پراکندگی نسبتاً یکنواخت‌تر در شکل‌های (۶-۱۰) بیان‌گر این نکته است که خطای معین کمتری در مدل حاصل از روش شبکه عصبی مصنوعی اتفاق افتاده است.

## ۶-۴-۲- تجزیه و تحلیل توصیف‌کننده‌های مدل

در یکی از روش‌های دسته‌بندی توصیف‌کننده‌های مولکولی، ابعاد بررسی مولکول مد نظر قرار می‌گیرد و بر این اساس چهار گروه، توصیف‌کننده‌های صفربعدی، یک بعدی، دو بعدی و سه بعدی خواهیم داشت. تمام توصیف‌کننده‌های وارد شده در مدل از نوع سه بعدی می‌باشند که در زیر گروه‌های توصیف‌کننده‌های هندسی، 3 D-MORSE و 1 GETAWAY قرار می‌گیرند.

توصیف‌کننده‌های GETAWAY بر اساس ماتریس قدرت نفوذ<sup>۲</sup> تعریف می‌شوند و بیان‌کننده‌ی ویژگی‌های هندسی و توپولوژیک مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تأثیر مولکول<sup>۳</sup> با رابطه‌ی زیر بیان می‌شود:

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T \quad (۲-۶)$$

که  $M$  ماتریس مختصات اتمی،  $M^T$  به معنای ماتریس ترانپوز و  $H$  ماتریس قدرت نفوذ است.  $H$  یک ماتریس متقارن  $A \times A$  است که  $A$  تعداد اتم‌هاست. عناصر قطری ماتریس  $H$  لوریج‌ها نام دارند که هر یک بیان‌گر اثر یک اتم در ایجاد شکل کلی یک مولکول است. توصیف‌کننده‌های GETAWAY به دو زیر مجموعه‌ی  $H$  و  $R$  تقسیم می‌شوند.

توصیف‌کننده‌های R-GETAWAY بر اساس ماتریس تأثیر- فاصله به صورت رابطه‌ی (۳-۶) تعریف می‌شوند:

$$[R]_{ij} = \begin{cases} \sqrt{h_{ii} \cdot h_{jj}} \\ r_{ij} \end{cases} \quad i \neq j \quad (۳-۶)$$

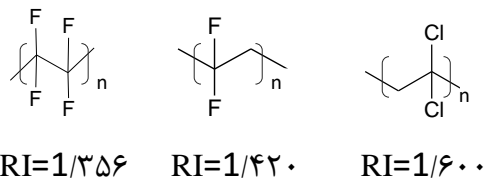
که  $h_{ii}$  و  $h_{jj}$  لوریج‌های دو اتم  $i$  و  $j$  و  $r_{ij}$  فاصله‌ی آن دو است و عناصر قطری این ماتریس صفر است. این توصیف‌کننده‌ها به ساختار سه‌بعدی مولکول‌ها حساس بوده و در برگیرنده‌ی اطلاعاتی نظیر

<sup>۱</sup>- Geometry, Topology, and Atom-Weights Assembly

<sup>۲</sup>- Leverage matrix

<sup>۳</sup>- Molecular Influence Matrix

شکل، اندازه، تقارن و نحوه توزیع اتم‌ها در مولکول می‌باشند [۷۳]. از این گروه، توصیف‌کننده‌های  $R2e$  و  $R2p$  در مدل ظاهر شده‌اند.  $R2e$  با افزایش الکترونگاتیوی افزایش می‌یابد. ضریب منفی این توصیف‌کننده بیان‌گر آن است که پلیمر دارای اتم الکترونگاتیوتر دارای ضریب شکست کمتری خواهد بود.  $R2p$  با جایگزینی الکترونگاتیویته اتمی سندرسون بوسیله‌ی قطبش‌پذیری اتمی در  $R2e$  بدست می‌آید که با افزایش قطبش‌پذیری اتمی مقدار این توصیف‌کننده افزایش یافته و در مدل دارای ضریب مثبت است. از آنجایی که اتم الکترونگاتیوتر، قطبش‌پذیری کمتری دارد، پلیمرهای حاوی اتم‌های با الکترونگاتیویته کمتر، ضریب شکست بزرگ‌تری خواهند داشت. به عنوان مثال مولکول‌های زیر قابل مقایسه هستند:



توصیف‌کننده‌های  $D-MORSE^3$  از معادله‌ای که در مطالعات پراش الکترونی بکار می‌رود و امکان نمایش سه بعدی مولکول را با مقادیری ثابت فراهم می‌کند، مشتق شده‌اند. خصوصیات اتمی گوناگونی مانند: جرم اتمی، حجم واندروالسی، الکترونگاتیویته و قطبش‌پذیری در محاسبه این توصیف‌کننده‌ها بکار می‌روند. این توصیف‌کننده‌ها قادرند تفاوت‌های ساختاری ترکیبات هم‌خانواده را که بسیار به هم نزدیک هستند نشان داده و امکان مطالعه ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی، شیمیایی، و بیولوژیکی را فراهم می‌سازند [۷۴]. این توصیف‌کننده‌ها با رابطه‌ی (۴-۶) نشان داده می‌شوند:

$$Mor(s, e) = I(s, e) = \sum_{i=2}^N \sum_{j=1}^{i-1} e_i e_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (4-6)$$

$$1 \leq s \leq 32$$

که  $s$  نشان دهنده‌ی پراش در جهات مختلف،  $e$  الکترونگاتیوی،  $N$  تعداد اتم‌ها و  $r$  فاصله‌ی میان اتم‌های  $i$  و  $j$  است.

در تحقیق حاضر، برای پیش‌بینی ضریب شکست پلیمرها، توصیف‌کننده‌ی  $Mor15e$  از این گروه وارد مدل شد که با توجه به اثر متوسط منفی آن در مدل می‌توان گفت که با افزایش شدت پراش الکترونی در فاصله‌ی ۱۵ آنگسترومی مولکول بر اساس خاصیت الکترونگاتیوی اتم‌ها، ضریب شکست نیز کاهش می‌یابد.

شاخص بالابان سه بعدی از نوع توصیف‌کننده‌های هندسی است که از ماتریس فاصله هندسی مشتق می‌شود. ماتریس هندسی  $G$  یک ماتریس متقارن مربعی است که در آن  $j$  امین ورودی، مسافت هندسی میان  $i$  امین و  $j$  امین اتم‌ها می‌باشد. درجه‌ی مسافت هندسی  $\sigma_i^G$  برای هر  $i$  برابر است با مجموع  $i$  امین سطر در ماتریس هندسی  $G$ :

$$\sigma_i^G = \sum_{j=1}^N G_{ij} \quad (5-6)$$

بنابراین شاخص بالابان سه بعدی به صورت رابطه‌ی زیر تعریف می‌شود:

$$J3D = (M / (M - N + 2)) \sum_M (\sigma_i^G \cdot \sigma_j^G)^{-0.5} \quad (6-6)$$

که  $\sigma_i^G$  و  $\sigma_j^G$  درجه‌های فاصله هندسی دو اتم مجاور  $i$  و  $j$  بوده و جمع روی تمام پیوندها در مولکول انجام می‌گیرد.  $M$  تعداد پیوندها و  $N$  نیز تعداد رئوس  $G$  می‌باشد [۳۰]. افزایش حلقه‌ها و شاخه‌دار شدن مولکول باعث کاهش این توصیف‌کننده می‌گردد. با توجه به ضریب منفی آن در مدل با کاهش این شاخص، برهم‌کنش‌ها در ساختار پلیمر بیشتر شده و باعث افزایش ضریب شکست ترکیب پلیمری می‌شود.



## آینده‌نگری

مدل‌های بدست آمده در این پروژه بخوبی قادر به پیش‌بینی خواص مشتقات ترکیبات بکار رفته خواهند بود و با گسترش روش‌های ترکیبی جدید دامنه‌ی مدل‌سازی را می‌توان به اندازه‌ای وسیع نمود که برای ترکیبات متفاوت نیز با موفقیت بکار رود.

برای بهبود انتخاب توصیف‌کننده‌ها و کاهش زمان انجام محاسبات می‌توان از تلفیق روش‌هایی از قبیل الگوریتم ژنتیک و شبکه‌ی عصبی مصنوعی بگونه‌ای استفاده کرد که شبکه نقش تابع شایستگی را در مدل ایفا کرده و در هر مرحله با توصیف‌کننده‌های منتخب الگوریتم ژنتیک آموزش ببیند. اخیراً بکارگیری الگوریتم‌های غیر خطی PC-ANN و ماشین بردار پشتیبان حداقل مربعات جزئی اجزای اصلی<sup>۱</sup> در رگرسیون و طبقه‌بندی، نتایج خوبی حاصل نموده است.

---

<sup>۱</sup> - principal component-least squares-support vector machine (PC-LS-SVM)

- [۱]- Levine Ira. N. (2002) "*Physical Chemistry*", McGraw-Hill, NewYork. pp 271.
- [۲]- Nirmalakhandan N. N. and Speece R. E. (1988) "*QSAR Model for Predicting Henry's Constant*", Environ. Sci. Technol., 22, pp 1349-1357.
- [۳]- Mackay D. and Shiu W.Y. (2000) "*Handbook of Property Estimation Methods for Chemicals: Environmental and Health Sciences*", CRC Press, Boca Raton, Chapter 4.
- [۴]- Knoll W. (1998) "*Interfaces and Thin Films as Seen by Bound Electromagnetic Waves*", Rev. Phys. Chem., 49, pp 569-639.
- [۵]- Krevelen D. W.van. (1976) "*Properties of polymers: correlation with chemical structure*", 2nd ed., Elsevier, Amsterdam.
- [۶]- Liu Q., et.al. (2005) "*Long-period gratings in polymer ridge waveguides*", Optics Express, 13, pp 1150-1160.
- [۷]- Kuzyk M. G., Garvey D. W., Canfield B. K., Vigil S. R., Welker D. J., Tostenrude J., Breckon C. (1999) "*Characterization of single-mode polymer optical fiber and electrooptic fiber devices*", Chemical Physics., 245, pp 327-340.
- [۸]- Ziemann O. ( 2007) "*Polymer Optical Fibers*", Specialty Optical Fibers Handbook , pp 617-650.
- [۹]- Safamirzaei M., Modarress H., Mohsen-Nia M. (2008) "*Modeling and predicting the Henry's law constants of methyl ketones in aqueous sodium sulfate solutions with artificial neural network*", Fluid Phase Equilibria., 266, pp 187-194.
- [۱۰]- Pablo R. D., J. Garro C. M., Castro E. A (2008) "*QSPR study of the Henry's Law constant for hydrocarbons*", J. chemolab., 91, pp 133-140.
- [۱۱]- Hu Z., Yao X. , Liu M., Zhang X. , Fan B. (2002) "*Radial basis function network-based quantitative structure–property relationship for the prediction of Henry's law constant*", ACA., 462, pp 101-117.
- [۱۲]- Xu J., Liang H., Chen B., Xu W., Shen X., Liu H. (2008) "*Linear and nonlinear QSPR models to predict refractive indices of polymers from cyclic dimer structures*", Chemolab., 92, pp 152-156.
- [۱۳]- Yu X., Yi B., Wang X. (2007) "*Prediction of Refractive index of vinyl polymer by using Density functional Theory*", J.Coput.Chem., 28, pp 2336-2341.
- [۱۴]- Garcia-Domenech R., Julian-Ortiz J.V. (2002) "*Prediction of Indices of Refraction and Glass Transition Temperatures of Linear Polymers by Using Graph Theoretical Indices*", J. Phys. Chem. B., 106, pp 1501-1507.

[١٥]- Wold S. (1995) “ *Chemometrics; what do we mean with it and what do we want from it*”, J.Chemolab., 30, pp 109-115.

[١٦]- Manssnat D. L., Vandeginste B. G., Deming S. N. and Kaufman L. (1998) “*Chemometrics, A Text Book*”, Elsevier, Amesterdom.

[١٧]- Jalali-Heravi M.,Garakani-Nejad Z. (1993) “*Prediction of gas chromatographic retention indices of some benzene derivatives*”, J. of . Chroma., 648, pp 389-393.

[١٨]- Charton M., Charton B. (2002) “*Advances in Quantitative Structure-Property Relationship*”, University of Bristol, UK, 3, pp 140.

[١٩]- Katritzky A. R., Dobchev D. A., Karelson M. (2006) “*Physical,chemical and technological property correlation with chemical structure: potential of QSPR*”, Z.Naturforsch. B.,61, pp 373-384.

[٢٠]- Gundertofte K., Jorgensen F. S. (2000) “*Molecular modeling and prediction of bioactivity*”, Plenum Pub, New York.

[٢١]- Jurs P. C. (2005) “*Assessing the reliability of a QSAR model's predictions*”, J.Mol Graph Model, 23(6), pp 503.

[٢٢]- *Introduction to molecular descriptors*, < <http://www.qsar.pe.kr/descriptors> > .

[٢٣]- Hansch C., Fujita T. (1964) “ *$\rho$ - $\sigma$ - $\pi$  Analysis: a method for the correlation of biological activity and chemical structure*”, J. Am. Chem. Soc, 86, pp 1616.

[٢٤]- Galton F. (1894) “*Natural Inheritance*”, New York, Mcmillan and Company.

[٢٥]- Stephens L. (2004) “*Advancsd Statistics Demystified*”, McGRAW-HILL., pp 125.

[٢٦]- Frank I.E and Todeschini R. (1994) “*The Data Analysis Handbook*”, Elsevier, Amsterdam(The Netherlands) , pp 366.

[٢٧]- Karelson M. (2000) “*Molecular Descriptors in QSAR/QSPR*”, Wiley-VCH, New York.

[٢٨]- Katritzky A. R., Karelson M., Petrukhin R. (2002) “*Topological Descriptors*”. University of Florida.

[٢٩]- Timmerman H., Todeschini R., Consonni V., Mannhold R., Kubinyi H. (2002) “*Handbook of Molecular Descriptors*”, Wiley-VCH. Weinheim.

[٣٠]- Todeschini R. And Consonni V. (2000) “*Handbook of Molecular Descriptors*”, Wiley-VCH,Weinheim(Germany).

[٣١]- Balaban A. T. (1997) “*From Chemical Topology to 3D molecular Geometry*”, Plenum Press, New York.

[٣٢]- *SPSS for windows*, Statistical Package for IBM PC, SPSS Inc., < <http://www.SPSS.com> >

[۳۳]- Pople J. A., Beveridge D. (1970) "Approximate Molecular Orbital Theory", McGraw-Hill.

[۳۴]- *HyperChem 8.0 for Windows* (2007), Toronto, Canada: HyperCube Inc, < <http://www.hyper.com> >.

[۳۵]- Todeschini R. (2001) Milano Chemometrics and QSAR Group, Italy, *Dragon software version 2.1.*, < <http://www.disat.unimib.it/vhml> >.

[۳۶]- MATLAB R2008a, The Math Works, Inc., Natick, MA, USA.

[۳۷]- نیروئی م، (۱۳۷۵)، پایان نامه ارشد: "طراحی یک مدل هیبرید و شبکه عصبی برای تشخیص سرطان سینه در ماموگرافی"، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران.

[۳۸]- Melanie M. (1996) "An introduction to Genetic Algorithms", MIT Press. Chapter 1.

[۳۹]- Holland J. H. (1975) "Adaptation in natural and artificial systems", University of Michigan Press.

[۴۰]- Goldberg D. E. (1989) "Algorithms in Search, Optimization and Machine Learning", Addison Wesley.

[۴۱]- Sipper M. (1996) "A Brief Introduction To Genetic Algorithms", University of Leeds. pp 121.

[۴۲]- Chen D. (1997) "Using Genetic Algorithms for the optimal design of structural systems", Arizona state university, pp 70-93.

[۴۳]- Nanakorn P. and Meesomklin K. (2001) "An adaptive penalty function in genetic algorithms for structural design optimization", Computers & Structures., 79. pp 2527-2539.

[۴۴]- Imam H. and Al-Shihri M. A. (2000) "A primitive crossover for improving the reliability of genetic algorithms for structural optimization, in Computational Engineering Using Metaphors from Nature", Civil-Comp Press, Edit. B.H.V. Topping, pp 91-97.

] Gen M., Cheng R. (2000) "Genetic algorithms and engineering optimization", [۴۵] 28-34. Wiley Canada., pp

[۴۶]- Rajeev S. and Krishnamoorthy C.S. (1992) "Discrete optimization of structures using Genetic algorithms", J. of Structural Engineering, ASCE, 118, pp 1233-1250.

[۴۷]- Sakamoto J. and Oda J. (1993) "A technique of optimal layout design for truss structures using genetic algorithm", Proc. 35th AIAA/ASCE/ASME/AHS SDM Conference, ASCE, New York, pp 2402-2408.

[۴۸]- Erbatur F., Hasancebi O., Tutuncu I. and Kilic H. (2000) "Optimal design of planar and space structures with Genetic algorithms", Computers and Structures, 75, pp 209-224.

[۴۹]- Leardi R., Boggia R. and Terrile M. (1992) "Genetic Algorithms as a Strategy for Feature Selection", J. Chemom., 6, pp 267-281.

[۵۰]- Fisz J. J. (2006) "Combined Genetic Algorithm and Multiple Linear Regression (GA-MLR) Optimizer: Application to Multiexponential Fluorescence Decay Surface", J. Phys. Chem., 110, 48, pp 12977-12985.

[۵۱]- McCulloch W.W. and Pitts W. (1943) "A Logical Calculus of the Ideas Imminent in Nervous Activity", Bull. Math. Biophys., 5, pp 115-133.

[۵۲]- Rosenblatt F. (1958) "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Cornell Aeronautical Lab, Psychological Review, v65, 6, pp 386-408.

[۵۳]- Hopfield J. J. (1982) "Neural networks and physical systems with emergent collective computational abilities", Proceedings of the National Academy of Sciences of the USA, vol. 79 no. 8 pp 2554-2558.

[۵۴]- Rumelhart D. E. and McClelland J. L. (1986) "Parallel Distributed Processing: Exploration in the Microstructure of Cognition", vol.1, Cambridge, MA:MIT Press, Chapter 1.

[۵۵]- Zupan J. and Gasteiger J. (1993) "Neural Networks for Chemists", VCH Publishers, New York, Chapter 2.

[۵۶]- Hagan M., Demuth H, Howard B. (2000) "Neural network design", Mark Beale MHB, Inc. PWS Publishing Company. Chapter 2.

[۵۷]- منهای م، (۱۳۸۷) "مبانی شبکه های عصبی (هوش محاسباتی)" جلد اول، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

[۵۸]- بن کویین، داورپناه ه، میرزایی ر، (۱۳۸۵) "هوش مصنوعی" انتشارات سخن گستر، مشهد، ص ۲۳۴.

[۵۹]- منهای م، سیفی پور ن، (۱۳۷۷) "کاربرد هوش محاسباتی در کنترل"، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

[۶۰]- Hornik K., Stinchcombe M., White, H. (1989) "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, 2, pp 359-366.

[۶۱]- اصغری اسکویی م.ر، (۱۳۸۱) "کاربرد شبکه های عصبی در پیش بینی سری های زمانی"، فصلنامه پژوهشهای اقتصادی ایران، شماره ۱۲، ص ۶۹.

[۶۲]- Hongxia P., Qingfeng Ma., Xiuye W. "Research on Fault Diagnosis of Gearbox Based on Particle Swarm Optimization Algorithm" ICM 2006 \* IEEE 3<sup>rd</sup> International Conference on Mechatronics.

[۶۳]- Dayhoff J. E. (1990) "Neural Network Principles", Prentice-Hall International, New York.

[۶۴]- Khanna T. (1990) "Foundation of neural networks", Addison-Wesley Publishing Company, U.S.A.

- [٦٥]- Hagan M. T. and Menhaj M. (1994) “*Training Feedforward Networks with the Marquardt Algorithm*”, IEEE Trans. Neural Networks, 5, No. 6, pp. 989-993.
- [٦٦]- Demuth H., Beale M., Hagan M. (2007) “*Neural network toolbox user's guide*”, Math Works Inc., Natick, MA, USA.
- [٦٧]- Yaws C. L. (1999) “*Chemical Properties Handbook*”, McGraw-Hill, New York.
- [٦٨]- شیرپور ا، (١٣٨٨)، پایان نامه ارشد: " کاربرد مطالعات رابطه کمی ساختار – فعالیت برای پیش‌بینی و مدل‌سازی فعالیت دارویی برخی از داروهای سنتز شده جدید"، دانشگاه صنعتی شاهرود.
- [٦٩]- Moriguchi I., Hirono S., Liu Q., Nakagome I., Matsushita Y. (1992) “*Simple method of calculating Octanol/Water Partition Coefficient*”, Chem. Pharm. Bull. 40, pp 127-130.
- [٧٠]- Brandrup J., Immergut E. H., Grulke E. A. (1999) “*POLYMER HANDBOOK*”, Vol. 1, 4nd ed, New York Public Library.
- [٧١]- Haupt R. L. and Haupt S. E. (1998) “*Practical Genetic Algorithm*”, John Wiley and Sons, Inc., New York.
- [٧٢]- Colt W., Smith D., Smith E. (1999) “*Using a Neural Network as a function Evaluator During GA Search for reliability optimization*”, department of industrial engineering university of Pittsburg.
- [٧٣]- Consonni V, Todeschini R., Pavan M. (2002) “*Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors.1. Theory of the Novel 3D Molecular Descriptors*”, J. Chem. Inf. Comput. Sci., 42, pp 682-692.
- [٧٤]- Schuur J.H., Selzer P., Gasteiger J. (1996) “*The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity*”, J. Chem. Inf. Comput. Sci., 36, pp 334-344.

## Abstract

Nowadays quantitative structure-property relationship (QSPR) is used for studying the dependence of various physico-chemical properties of compounds to molecular structure. In the first section of this project, Multiple linear regression (MLR) as a linear method and artificial neural network (ANN) as nonlinear method used to predicting the Henry constant of 333 organic compounds. The data set was divided into a training (263 compounds), test (40 compounds) and an external validation (30 compounds) sets. A three-layered feed forward ANN with back-propagation of error was generated using four molecular descriptors appearing in the MLR model. The statistical characteristics provided by multiple linear model ( $R^2 = 0.942$ ;  $MSE = 0.109$  for test set) indicated predictive ability, while the predictive ability of ANN model is somewhat superior ( $R^2 = 0.982$ ;  $MSE = 0.039$  for test set).

In the second part of this project, genetic algorithm – multiple linear regression (GA-MLR) and genetic algorithm–artificial neural network (GA-ANN) methods were used to generate QSPR models between the descriptors and refractive index of 144 polymers. The  $R2e$ ,  $R2p$ ,  $Mor15e$  and  $J3D$  descriptors selected by genetic algorithm for constructing of linear and nonlinear models to predicting the refractive index. The root mean square error of the training, test, and validation sets for the ANN model are 0.018, 0.023, 0.018 and also determination coefficients ( $R^2$ ) of 0.929, 0.881, and 0.931 respectively.

Keywords: QSPR, Henry constant, artificial neural network, refractive index, genetic algorithm– multiple linear regression



دانشگاه صنعتی شاهرود

Shahrood University of Technology

Faculty of Chemistry

Prediction of Henry constants of some organic compounds  
using of linear and nonlinear QSPR methods

Mohammad Ali Ferdowsi

Supervisors:

Dr. Hossein Nikoofard

Dr. Nasser Goudarzi

Advisor:

Dr. Zahra Kalantar

June 2010