

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده: شیمی

## پیش بینی اثرات بازدارنده بعضی ترکیبات دارویی باروشهای QSAR

دانشجو: فاطمه عامری

استاد راهنما:

دکتر ناصر گودرزی

استاد مشاور:

دکتر منصور عرب چم‌جنگلی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

اردیبهشت ماه ۹۰

## خدایا:

ای سرانجام نیک هر نیاز خواهی،

ای آنکه توفیق رسیدن به خواسته‌ها از توست،

گنجینه‌های نعمت را تهی نمی‌گرداند، ای آنکه خواسته‌های بسیار،

پس بر این نعمت‌ها تو را سپاس ای پروردگار هستی.

## تقدیم به:

روح پدر بزرگوارم و مادر مهربانم که همیشه حامی و پشتیبان من بودند

همسر عزیزم که صبورانه همراهیم کرد و

خواهران و برادر نازنینم که همیشه باعث دلگرمی‌ام بودند

## تشکر و قدردانی

اکنون که با عنایت و یاری خداوند مهربان توانسته ام این دوره از تحصیلاتم را با موفقیت به پایان برسانم، بر خود لازم می‌دانم، از اساتید محترمی که در به ثمر رساندن این پایان نامه مرا یاری کردند تشکر و قدردانی کنم.

از جناب آقای دکتر گودرزی به عنوان استاد راهنما که در طول انجام این پروژه از هیچ کمکی نسبت به من دریغ نکردند و از جناب آقای دکتر عرب نیز به عنوان استاد مشاور نهایت سپاسگزاری را دارم و از خداوند منان توفیق روز افزون و سلامتی این بزرگواران را خواستارم.

## چکیده:

در این پروژه، مطالعات ارتباط کمی ساختار- فعالیت (QSAR) بر روی ثابت بازداري (ki) ۷۷ ترکیب دارویی از مشتقات پایپرازینیل گلوتامات پیریدین که به عنوان ترکیبات ضد لخته شدن خون عمل می‌کنند، انجام شد. برای انتخاب مهمترین توصیف کننده‌ها رگرسیون خطی مرحله‌ای مورد استفاده قرار گرفت، سپس توصیف کننده‌های مهم به عنوان ورودی برای ایجاد مدل‌های QSAR با استفاده از رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN)، به کار گرفته شد. اعتبار این مدل‌ها توسط سری تست و تکنیک‌های حذف مرحله‌ای تک تک (LOO) و  $-Y$  تصادفی بررسی شد. نتایج نشان داد ضریب تعیین برای پیش بینی ثابت بازداري ترکیبات سری تست با مدل MLR و ANN به ترتیب ۰/۹۴۵۸ و ۰/۹۶۶۱ است.

در بخش دوم این تحقیق، روش‌های MLR و ANN برای مدل‌سازی و پیش بینی فعالیت ضد HIV یکسری ترکیبات از مشتقات ۵،۶-دی هیدروکسی پیریمیدین -۴- کربوکسامید، استفاده شد. ارزیابی تقاطعی بر روی سری آموزش برای انتخاب بهترین مدل MLR و ANN به کار گرفته شد. ضریب تعیین به دست آمده برای سری تست توسط روش‌های MLR و ANN به ترتیب ۰/۹۸۳۶ و ۰/۹۸۳۷ بودند. نتایج به دست آمده توانایی پیش بینی مناسب مدل‌های پیشنهادی را نشان می‌دهد.

**کلمات کلیدی:** QSAR، ثابت بازداري، ضد HIV، شبکه عصبی مصنوعی، رگرسیون خطی

چندگانه.

## فهرست مطالب

### فصل اول

- ۱-۱- تعریف تحقیق ..... ۲
- ۲-۱- ثابت بازداری فعالیت ..... ۳
- ۳-۱- ضرورت تحقیق ..... ۳
- ۴-۱- مروری بر کارهای انجام شده ..... ۵

### فصل دوم

- ۱-۲- کمومتریکس ..... ۹
- ۲-۲- ارتباط کمی ساختار- فعالیت ..... ۱۱
- ۱-۲-۲- انتخاب سری داده‌ها ..... ۱۲
- ۲-۲-۲- رسم و بهینه سازی ساختار ترکیبات ..... ۱۳
- ۳-۲-۲- محاسبه توصیف کننده‌ها ..... ۱۵
- ۴-۲-۲- انتخاب توصیف کننده های مناسب ..... ۱۶
- ۵-۲-۲- انتخاب مدل مناسب ..... ۱۶
- ۶-۲-۲- رگرسیون خطی چندگانه ..... ۱۷
- ۱-۶-۲-۲- روش ورود اجباری ..... ۲۱
- ۲-۶-۲-۲- روش جلو برنده ..... ۲۱
- ۳-۶-۲-۲- روش عقب برنده ..... ۲۱
- ۴-۶-۲-۲- روش مرحله‌ای ..... ۲۲
- ۷-۲-۲- تکنیک‌های اعتبارسنجی ..... ۲۲
- ۱-۷-۲-۲- ارزیابی متقاطع ..... ۲۳
- ۲-۷-۲-۲- اعتبارسنجی بیرونی ..... ۲۳
- ۳-۷-۲-۲- آزمون Y- تصادفی ..... ۲۴
- ۳-۲- نرم افزارهای مورد نیاز ..... ۲۴
- ۱-۳-۲- بسته‌ی نرم افزاری Hyperchem ..... ۲۴

|         |                                     |
|---------|-------------------------------------|
| ۲۵..... | Dragon نرم افزاری بسته‌ی ۲-۳-۲..... |
| ۲۶..... | SPSS نرم افزاری بسته‌ی ۳-۳-۲.....   |
| ۲۷..... | MATLAB نرم افزار ۴-۳-۲.....         |

### فصل سوم

|         |  |
|---------|--|
| ۲۹..... | ۱-۳-۱- مقدمه ای بر شبکه های عصبی مصنوعی..... |
| ۳۱..... | ۲-۳-۲- نرون بیولوژیکی.....                   |
| ۳۲..... | ۳-۳-۳- ساختار و عملکرد نرون مصنوعی.....      |
| ۳۴..... | ۴-۳- معماری شبکه‌های عصبی.....               |
| ۳۵..... | ۵-۳- آموزش شبکه عصبی.....                    |
| ۳۶..... | ۶-۳- انواع شبکه های عصبی.....                |
| ۳۶..... | ۱-۶-۳- شبکه‌های جلوبرنده (پیش خور).....      |
| ۳۷..... | ۲-۶-۳- شبکه های برگشتی (پس خور).....         |
| ۳۸..... | ۷-۳- طراحی شبکه پیش خور.....                 |
| ۳۹..... | ۸-۳- آموزش شبکه های پیش خور.....             |
| ۳۹..... | ۱-۸-۳- ایجاد یک شبکه پیش خور.....            |
| ۳۹..... | ۲-۸-۳- مقدار دهی آغازین به وزن‌ها.....       |
| ۳۹..... | ۳-۸-۳- شبیه سازی.....                        |
| ۴۰..... | ۴-۸-۳- آموزش شبکه.....                       |
| ۴۰..... | ۹-۳- الگوریتم‌های آموزشی شبکه‌های عصبی.....  |
| ۴۰..... | ۱-۹-۳- الگوریتم آموزشی پس انتشار خطا.....    |
| ۴۲..... | ۲-۹-۳- الگوریتم لومبرگ- مارکورات.....        |
| ۴۴..... | ۱۰-۳- ارتقای عمومیت.....                     |

### فصل چهارم

|         |   |
|---------|---|
| ۴۹..... | ۱-۴- سری داده‌ها.....                       |
| ۵۶..... | ۲-۴- رسم و بهینه‌سازی ساختار مولکول‌ها..... |

- ۳-۴- محاسبه توصیف کننده های مولکولی.....۵۶
- ۴-۴- انتخاب توصیف کننده های مناسب با استفاده از روش رگرسیون خطی چندگانه.....۵۶
- ۵-۴- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN).....۵۹
- ۴-۵-۱- بهینه سازی تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه و تعداد نرون های لایه مخفی.....۶۰
- ۴-۵-۲- بهینه سازی تعداد دورهای آموزشی.....۶۸
- ۴-۵-۳- بهینه سازی پارامتر  $\mu$ .....۶۹
- ۴-۵-۴- ساختار شبکه عصبی مصنوعی بهینه شده.....۷۰
- ۴-۶-۱- ارزیابی مدل های خطی و غیرخطی.....۷۱
- ۴-۶-۱- ارزیابی مدل های ANN و MLR با استفاده از سری تست.....۷۱
- ۴-۶-۲- ارزیابی مدل های ANN و MLR به روش حذف مرحله ای تک تک داده ها.....۷۶
- ۴-۶-۳- ارزیابی مدل های برتر خطی و غیرخطی با استفاده از پارامترهای آماری.....۸۳
- ۴-۶-۴- ارزیابی شبکه به روش  $Y$ -تصادفی.....۸۳
- ۴-۷- بررسی ارتباط توصیف کننده های منتخب با ثابت بازدارندگی فعالیت ترکیبات.....۸۴
- ۴-۸- نتیجه گیری.....۸۹

### فصل پنجم

- ۵-۱- سری داده ها.....۹۱
- ۵-۲- رسم و بهینه سازی ساختار مولکول ها.....۹۴
- ۵-۳- محاسبه و انتخاب بهترین توصیف کننده ها.....۹۴
- ۵-۴- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN).....۹۷
- ۵-۴-۱- بهینه سازی تابع آموزش، تابع انتقال و تعداد نرون های لایه مخفی.....۹۷
- ۵-۴-۲- بهینه سازی پارامتر  $\mu$ .....۱۰۴
- ۵-۴-۳- بهینه سازی تعداد دورهای آموزشی.....۱۰۵
- ۵-۴-۴- ساختار شبکه عصبی مصنوعی بهینه شده.....۱۰۷
- ۵-۵- ارزیابی مدل های خطی و غیرخطی.....۱۰۸
- ۵-۵-۱- ارزیابی مدل های ANN و MLR با استفاده از سری تست.....۱۰۸



|          |  |
|----------|--|
| ۱۰۹..... | ۲-۵-۵- ارزیابی مدل‌های ANN و MLR به روش حذف مرحله‌ای تک تک داده‌ها.....      |
| ۱۱۳..... | ۳-۵-۵- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری..... |
| ۱۱۴..... | ۴-۵-۵- ارزیابی شبکه به روش Y- تصادفی.....                                    |
| ۱۱۵..... | ۶-۵- بررسی ارتباط توصیف‌کننده‌های منتخب با فعالیت دارویی ترکیبات.....        |
| ۱۱۹..... | ۷-۵- نتیجه‌گیری.....   |
| ۱۲۰..... | آینده‌نگری.....  |
| ۱۲۱..... | منابع.....   |

## فهرست اشکال

- شکل ۳-۱- ساختار نرون طبیعی..... ۳۱
- شکل ۳-۲- مدل ریاضی نرون..... ۳۲
- شکل ۳-۳- برخی از توابع انتقال مورد استفاده در شبکه‌های عصبی..... ۳۴
- شکل ۳-۴- نمایش یک نرون تک‌لایه..... ۳۵
- شکل ۳-۵- شیوه‌ی یادگیری تحت نظارت..... ۳۶
- شکل ۳-۶- شبکه عصبی پیش‌خور..... ۳۷
- شکل ۳-۷- شبکه عصبی برگشتی..... ۳۷
- شکل ۳-۸- شبکه پیش‌خور  $tansig/purelin$ ..... ۳۸
- شکل ۴-۱- ساختار کلی ترکیبات دارویی..... ۴۹
- شکل ۴-۲- تأثیر تعداد توصیف‌کننده‌ها بر مقادیر  $R^2$  مدل‌های MLR..... ۵۸
- شکل ۴-۳- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۶۲
- شکل ۴-۴- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۶۳
- شکل ۴-۵- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال خطی..... ۶۴
- شکل ۴-۶- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم‌بازین و تابع انتقال لگاریتم سیگموئیدی..... ۶۵
- شکل ۴-۷- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم‌بازین و تابع انتقال تانژانت سیگموئیدی..... ۶۶
- شکل ۴-۸- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم‌بازین و تابع انتقال خطی..... ۶۷
- شکل ۴-۹- منحنی مقدار میانگین مربع خطای سری ارزیابی در برابر تعداد دوره‌های آموزشی متفاوت..... ۶۹
- شکل ۴-۱۰- نمودار میانگین مربع خطای سری ارزیابی بر حسب پارامتر  $\mu$ ..... ۷۰
- شکل ۴-۱۱- ساختار شبکه عصبی بهینه شده..... ۷۰
- شکل ۴-۱۳- نمودار مقادیر پیش‌بینی شده‌ی  $pK_a$  بر حسب مقادیر تجربی برای سری ارزیابی توسط الف) روش MLR ،  
ب) روش ANN..... ۷۳
- شکل ۴-۱۴- نمودار مقادیر پیش‌بینی شده‌ی  $pK_a$  بر حسب مقادیر تجربی برای سری تست توسط الف) روش MLR ،  
ب) روش ANN..... ۷۵

- شکل ۴-۱۵- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی  $pK_i$  به روش حذف مرحله‌ای تک تک داده‌ها الف)
- ۸۱..... روش MLR ب) روش ANN.....
- شکل ۴-۱۶- نمودار باقیمانده‌ها بر حسب مقدار تجربی  $pK_i$  برای کل داده‌ها الف) روش MLR ب) روش ANN..... ۸۲
- شکل ۵-۱- ساختار کلی ترکیبات ۵، ۶- دی هیدروکسی پیریمیدین-۴-کربوکسامید..... ۹۱
- شکل ۵-۲- تأثیر تعداد توصیف‌کننده‌ها بر مقادیر  $R^2$  مدل‌های MLR..... ۹۵
- شکل ۵-۳- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۹۸
- شکل ۵-۴- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۹۹
- شکل ۵-۵- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال خطی..... ۱۰۰
- شکل ۵-۶- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی..... ۱۰۰
- شکل ۵-۷- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی..... ۱۰۲
- شکل ۵-۸- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال خطی..... ۱۰۳
- شکل ۵-۹- نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر  $\mu$ ..... ۱۰۵
- شکل ۵-۱۰- منحنی مقدار مربع میانگین خطاها در برابر تغییر تعداد دوره‌های آموزشی..... ۱۰۶
- شکل ۵-۱۱- تصویر شماتیک ساختار شبکه عصبی مصنوعی به دست آمده پس از بهینه‌سازی..... ۱۰۷
- شکل ۵-۱۲- نمودار مقادیر پیش‌بینی شده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری تست توسط الف) روش MLR ،  
ب) روش ANN..... ۱۰۹
- شکل ۵-۱۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی  $pIC_{50}$  به روش حذف مرحله‌ای تک تک داده‌ها  
الف) روش MLR ب) روش ANN..... ۱۱۲
- شکل ۵-۱۴- نمودار باقیمانده‌ها بر حسب مقدار تجربی  $pIC_{50}$  برای کل داده‌ها الف) روش MLR ب) روش ANN..... ۱۱۳

## فهرست جداول

- جدول ۲-۱- انواع توصیف کننده های محاسبه شده توسط نرم افزار Dragon.....۲۶
- جدول ۴-۱- سری ترکیبات مورد بررسی.....۴۹
- جدول ۴-۲- مقایسه آماره های مدل های به دست آمده از روش MLR.....۵۷
- جدول ۴-۳- توصیف کننده های وارد شده در مدل منتخب MLR.....۵۸
- جدول ۴-۴- ماتریس همبستگی توصیف کننده های انتخاب شده در مدل نهایی MLR.....۵۹
- جدول ۴-۵- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع تبدیل لگاریتم سیگموئید (logsig) با تعداد متغیر های ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۲
- جدول ۴-۶- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع تبدیل تانژانت سیگموئید (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۳
- جدول ۴-۷- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع تبدیل خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۴
- جدول ۴-۸- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی بایزین (trainbr) و تابع تبدیل لگاریتم سیگموئید (logsig) با تعداد متغیر های ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۵
- جدول ۴-۹- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی بایزین (trainbr) و تابع تبدیل تانژانت سیگموئید (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۶
- جدول ۴-۱۰- مقادیر میانگین مربع خطاها برای شبکه هایی با الگوریتم آموزشی بایزین (trainbr) و تابع تبدیل خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره های مختلف در لایه پنهان.....۶۷
- جدول ۴-۱۱- مقادیر خطای شبکه با تغییر تعداد دوره های آموزشی.....۶۸
- جدول ۴-۱۲- مقادیر میانگین مربع خطای سری ارزیابی در مقادیر متفاوت  $\mu$ .....۶۹
- جدول ۴-۱۳- توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....۷۱
- جدول ۴-۱۴- مقادیر تجربی و پیش بینی شده ی  $pK_i$  به وسیله ی روش MLR برای سری ارزیابی.....۷۲
- جدول ۴-۱۵- مقادیر تجربی و پیش بینی شده ی  $pK_i$  به وسیله ی روش ANN برای سری ارزیابی.....۷۲
- جدول ۴-۱۶- مقادیر تجربی و پیش بینی شده ی  $pK_i$  به وسیله ی روش MLR برای سری تست.....۷۴
- جدول ۴-۱۷- مقادیر تجربی و پیش بینی شده ی  $pK_i$  به وسیله ی روش ANN برای سری تست.....۷۴

- جدول ۴-۱۸- ارزیابی مدل MLR به روش حذف مرحله‌ای تک تک داده‌ها.....۷۶
- جدول ۴-۱۹- ارزیابی مدل ANN به روش حذف مرحله‌ای تک تک داده‌ها.....۷۸
- جدول ۴-۲۰- پارامترهای آماری دو روش MLR و ANN.....۸۳
- جدول ۴-۲۱- مقادیر  $R^2$  برای سری تست و ارزیابی با استفاده از آزمون Y- تصادفی.....۸۴
- جدول ۵-۱- سری ترکیبات مورد مطالعه.....۹۱
- جدول ۵-۲- توصیف کننده‌های انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله.....۹۵
- جدول ۵-۳- توصیف کننده‌های وارد شده در مدل منتخب MLR.....۹۶
- جدول ۵-۴- ماتریس همبستگی توصیف کننده‌های وارد شده در مدل منتخب.....۹۶
- جدول ۵-۵- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان.....۹۸
- جدول ۵-۶- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان.....۹۹
- جدول ۵-۷- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال خطی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان.....۱۰۰
- جدول ۵-۸- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان..
- .....۱۰۱
- جدول ۵-۹- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان....
- .....۱۰۲
- جدول ۵-۱۰- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال خطی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان.....۱۰۳
- جدول ۵-۱۱- مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت  $\mu$ .....۱۰۵

- جدول ۵-۱۲- مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دورهای آموزش مختلف  
۱۰۶.....
- جدول ۵-۱۳- توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....  
۱۰۷.....
- جدول ۵-۱۴- مقادیر تجربی و پیش‌بینی شده ی  $pIC_{50}$  به وسیله‌ی روش MLR برای سری تست.....  
۱۰۸.....
- جدول ۵-۱۵- مقادیر تجربی و پیش‌بینی شده ی  $pIC_{50}$  به وسیله‌ی روش ANN برای سری تست.....  
۱۰۸.....
- جدول ۵-۱۶- ارزیابی مدل MLR به روش حذف مرحله‌ای تک تک داده‌ها.....  
۱۱۰.....
- جدول ۵-۱۷- ارزیابی مدل ANN به روش حذف مرحله‌ای تک تک داده‌ها.....  
۱۱۱.....
- جدول ۵-۱۸- پارامترهای آماری دو روش MLR و ANN.....  
۱۱۴.....
- جدول ۵-۱۹- مقادیر  $R^2$  برای سری تست با استفاده از آزمون Y- تصادفی.....  
۱۱۴.....

## فصل اول

### مقدمه

## ۱-۱- تعریف تحقیق

کوموتریکس شاخه‌ای از شیمی است که ابزارهای ریاضی و آمار را برای سامان بخشیدن به داده‌های شیمیایی و بدست آوردن اطلاعات بیشتر از آنها به کار می‌گیرد. از نظر شیمی‌دانان فعالیت‌ها و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. این نوع از مطالعات به بررسی کمی ارتباط ساختار با فعالیت (QSAR)<sup>۱</sup> و همچنین بررسی کمی ارتباط ساختار با خاصیت (QSPR)<sup>۲</sup> معروف می‌باشد.

نتایج این مطالعات علاوه بر مشخص شدن نحوه‌ی ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمانی آنها به محققان در پیش بینی رفتار مولکول‌های جدید براساس رفتار مولکول‌های مشابه کمک می‌کند.

یکی از شاخه‌های بسیار مهم کوموتریکس که در این پایان‌نامه به کار گرفته می‌شود، ارتباط کمی ساختار - فعالیت است. این تکنیک در این پایان‌نامه برای پیش‌بینی ثابت بازدارنده ( $K_i$ )<sup>۳</sup> تعدادی از مشتقات پایپرازینیل گلوتامات پیریدین<sup>۴</sup> در فصل چهارم، و پیش‌بینی فعالیت دارویی ( $IC_{50}$ )<sup>۵</sup> تعدادی از مشتقات ۵، ۶-دی‌هیدروکسی-۴-پیریمیدین کربوکسامید<sup>۶</sup> در فصل پنجم به کار گرفته شده است. از جمله روش‌های خطی که بدین منظور استفاده می‌شوند می‌توان به روش‌های رگرسیون خطی چندگانه<sup>۷</sup>، آنالیز اجزای اصلی<sup>۸</sup> و روش حداقل مربعات جزئی<sup>۹</sup> اشاره کرد. از مهمترین روش‌های غیر خطی نیز می‌توان به شبکه‌های عصبی مصنوعی<sup>۱۰</sup> اشاره کرد. در تحقیق حاضر مطالعه ارتباط کمی

---

۱. Quantitative Structure-Activity Relationship (QSAR)

۲. Quantitative Structure-Property Relationship (QSPR)

۳. Constant inhibition

۴. Piperazinyll Glutamate Pyridine

۵. Half maximal inhibitory concentration

۶. 5,6-Dihydroxypyrimidine-4-carboxamide

۷. Multiple Linear Regression (MLR)

۸. Principal Components Analysis (PCA)

۹. Partial Least Square (PLS)

۱۰. Artificial Neural Network (ANN)



ساختار - فعالیت ترکیبات مورد نظر با استفاده از روش‌های خطی و غیرخطی انجام گرفته و نتایج با هم مقایسه شده‌اند.

### ۲-۱- ثابت بازدارندگی فعالیت ( $K_i$ )

$IC_{50}$  یا غلظت مهارکنندگی نصف ماکزیمم، غلظتی از دارو است که ۵۰٪ اثر بازدارندگی داشته باشد. به عبارت دیگر به غلظتی از ترکیب (دارو) که از فعالیت بیولوژیکی و یا بیوشیمیایی آنزیم تا حدود ۵۰٪ جلوگیری کند،  $IC_{50}$  گفته می‌شود.

مقدار  $IC_{50}$  با استفاده از معادله چنگ-پروسوف<sup>۱</sup> به ثابت بازدارندگی  $K_i$  تبدیل می‌شود [۱]:

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \quad (1-1)$$

$K_i$  تمایل چسبیدن بازدارنده (دارو) به آنزیم است در حالی که  $IC_{50}$  قدرت بازدارنده را نشان می‌دهد. همچنین  $S$  غلظت بازدارنده (دارو) بوده و  $K_m$  نیز غلظت بازدارنده (دارو) در نصف ماکزیمم فعالیت آنزیم می‌باشد.

### ۳-۱- ضرورت تحقیق

بیماری‌های مغزی-عروقی<sup>۲</sup> و قلبی-عروقی<sup>۳</sup> یکی از رایج‌ترین علل مرگ و میر در جهان است [۲]. اگر خون رسانی به قسمتی از مغز دچار اختلال شود و متوقف گردد، این قسمت از مغز دیگر

---

۱. Cheng-Prusoff

۲. Cerebrovascular

۳. Cardiovascular

نمی‌تواند عملکرد طبیعی خود را داشته باشد، این وضعیت را اصطلاحاً سکته<sup>۱</sup> می‌نامند. سکته مغزی می‌تواند به عللی مانند بسته شدن یا پاره شدن یکی از رگ‌های خون رسان مغز ایجاد شود. یک حمله قلبی (سکته قلبی) نیز زمانی رخ می‌دهد که یک منطقه از عضله قلب به علت نرسیدن اکسیژن کافی به آن محل دچار صدمه دائمی یا مرگ شود.

بیشتر حملات قلبی توسط یک لخته که یکی از رگ‌های کرونری<sup>۲</sup> (رگ‌های خونی که خون و اکسیژن اکسیژن را به عضله قلب می‌رسانند) را مسدود می‌کند ایجاد می‌شوند. لخته درون رگ کرونری جریان خون و اکسیژن رسانی به عضله قلبی را مختل نموده، که سبب مرگ سلول‌های قلبی در آن ناحیه می‌شود [۱].

ایدز (AIDS)<sup>۳</sup> نیز یکی از دلایل اصلی مرگ و میر در جهان است به طوری که در سال ۲۰۰۹ بیش از چهل میلیون نفر به این بیماری مبتلا شده‌اند [۳]. این آمار همچنان در حال گسترش بوده و اکثر مبتلایان به ایدز در کشورهای در حال توسعه زندگی می‌کنند.

ایدز به معنی نشانگان نقص ایمنی اکتسابی یک بیماری عفونی کشنده است که اولین بار در سال ۱۹۸۱ شناسایی شد. این بیماری حاصل تکثیر ویروسی به نام (HIV)<sup>۴</sup> در بدن میزبان است که باعث تخریب جدی دستگاه ایمنی بدن انسان شده که خود زمینه‌ساز بروز عفونت‌های فرصت طلب<sup>۵</sup> است که یک بدن سالم عموماً قادر به مبارزه با آنها نبوده و در نهایت پیشرفت همین عفونت‌ها منجر به مرگ بیمار می‌گردد [۱].

بنابراین نیاز به درمان توسط مواد دارویی ارزان و مؤثر امری ضروری است. از آن جایی که بیشتر داروهای شیمیایی توسط آزمون و خطا سنتز می‌شوند که این امر بسیار وقت‌گیر و هزینه‌بر است [۴ و ۵]. توسعه روش‌های تئوری به عنوان یک ابزار کارآمد برای رفع این محدودیت‌ها مورد توجه

---

۱. Stroke

۲. Coronary

۳. Acquired Immune Deficiency Syndrome

۴. Human Immunodeficiency Virus

۵. Infection opportunistic

قرار گرفته است. به کارگیری این روش‌ها می‌تواند علاوه بر پیش‌بینی فعالیت بیولوژیکی مورد نظر، به هدفمندتر شدن انجام آزمایش‌های تجربی و توضیح پارامترهای مؤثر بر نتایج آزمایش‌ها کمک کند.

از جمله روشهای نظری، رابطه کمی ساختار-فعالیت (QSAR) است که در آن می‌توان از روشهای مختلفی برای ایجاد روابط خطی و غیرخطی میان ساختار و خاصیت دارویی ترکیباتی که فعالیت دارویی آنها با استفاده از روشهای تجربی اندازه‌گیری شده‌اند، استفاده نمود. مبنای روش QSAR، به دست آوردن روابط کمی بهینه بین ساختار و فعالیت دارویی ترکیبات مورد مطالعه است و هنگامی که رابطه معتبری به دست آمد امکان پیش‌بینی فعالیت دارویی برای ساختارهای مشابه اندازه‌گیری شده یا ساختارهایی که فعالیت آنها هنوز اندازه‌گیری نشده است، وجود دارد. بر این اساس QSAR به راحتی می‌تواند برای کمک به بهینه‌سازی سنتز داروها مورد استفاده قرار گیرد.

#### ۱-۴- مروری بر کارهای انجام شده

در سال ۲۰۰۲، کاندیا<sup>۱</sup> و همکارانش مطالعات QSAR بر روی ترکیبات فنیل آمید و استرهای پایپیریدین کربوکسیلیک اسید که از لخته شدن خون جلوگیری می‌کنند را بررسی کردند. ضریب تعیین برای کل سری داده‌ها ۰/۷۴ به دست آمد [۶].

دوبل<sup>۲</sup> و همکارانش در سال ۲۰۰۵ مطالعات QSAR بر روی آنالوگ‌های پانول را مورد بررسی قرار دادند. آنالوگ‌های پانول شامل مشتقات ۱- (۲- هیدروکسی فنیل) اتانول هستند که از لخته شدن خون جلوگیری می‌کنند. در این تحقیق از مدل شبکه عصبی پس انتشار و رگرسیون خطی چند گانه جهت ارتباط ساختار-فعالیت استفاده شد [۷].

---

۱ . Candia

۲ . Doble

کاتریتزکی<sup>۱</sup> و همکارانش در سال ۲۰۰۶، مطالعات QSAR بر روی ۶۰ مشتق از بنزوکسازینون را که به عنوان عوامل ضد لخته شدن خون به شمار می روند بررسی کردند. ضریب تعیین گزارش شده برای روش MLR و ANN به ترتیب ۰/۸۲۱ و ۰/۹۰۹ بودند [۸].

در سال ۲۰۰۶ کابالرو<sup>۲</sup> و همکارانش ثابت بازداري تعدادی از مشتقات کتوننی را با استفاده از شبکه عصبی ژنتیک تنظیم بایزین (BRGNN)<sup>۳</sup> بررسی کردند و ضریب تعیین گزارش شده نیز ۰/۸۷۴ بود [۹].

در سال ۲۰۰۰ جلالی هروی و همکارش مطالعات QSAR را برای ایجاد مدل‌هایی که ارتباط دهنده‌ی فعالیت ضد HIV با ساختار ۱۰۷ مشتق (۱-۲- (هیدروکسی اتوکسی) متیل) -۶- (فنیل تیو) تیمین (HEPT) باشد، انجام دادند. در این کار مدل‌سازی توسط شبکه عصبی مصنوعی نسبت به روش رگرسیون خطی چندگانه نتایج بهتری را ارائه داد به طوری که میانگین مربع خطا (MSE)<sup>۴</sup> برای سری تست با استفاده از روش خطی چندگانه ۰/۷۸۰ و با استفاده از شبکه عصبی ۰/۳۷۲ به دست آمد [۱۰].

گودرزی و همکارانش در سال ۲۰۰۹ مدل‌سازی QSAR را بر روی مشتقات استریل کوینولین به عنوان ترکیبات ضد HIV با استفاده از الگوریتم ژنتیک کوپل شده با رگرسیون خطی چند گانه انجام دادند ضریب تعیین برای سری تست ۰/۷۹۷۷ بود [۱۱].

در سال ۲۰۰۷ عرب چم جنگلی و همکارانش ارتباط میان ساختار و فعالیت دارویی ۴۲ ترکیب از مشتقات ۵- فنیل-۱- فنیل‌آمینو-۱-هیدروژن-ایمیدازول را با استفاده از شبکه مصنوعی به دست آوردند و نشان دادند که مدل به دست آمده با شبکه عصبی مصنوعی نسبت به مدل خطی که توسط روی<sup>۵</sup> و لئوناردو<sup>۶</sup> در سال ۲۰۰۵ ارائه کرده بود [۱۲]، قدرت پیش‌بینی بهتری برای این دسته از

---

۱ . Katritzky

۲ . Caballero

۳ . Bayesian Regularized Genetic Neural Network

۴ . Mean Square Error

۵ . Roy

۶ . Leonardo

ترکیبات دارد [۱۳]. عرب چم جنگلی همچنین در سال ۲۰۰۹ با استفاده از روش غیرخطی بر مبنای شبکه عصبی، مدلی مناسب برای فعالیت ضد HIV مشتقات پیرویل آریل سولفون‌ها ارائه داد، که ضریب همبستگی برای سری تست ۰/۹۸۹ بود [۵].

جلالی هروی نیز در سال ۲۰۰۹ مدل‌سازی QSAR ۱- (۳ و ۳ دی فنیل پروپیل) - پایپیریدینیل آمیدها را با استفاده از شبکه عصبی ژنتیک تنظیم بایزین (BRGNN) انجام داد. در این تحقیق هروی از الگوریتم ژنتیک برای انتخاب متغیرهای ورودی شبکه استفاده کرد و مدلی مناسب برای فعالیت ضد HIV این ترکیبات ارائه داد و برتری این روش را نسبت به روش MLR اعلام کرد [۱۴].

در سال ۲۰۰۸، راویچاندران<sup>۱</sup> و همکارانش مطالعات QSAR را بر روی مشتقات PETT (phenyl ethyl thiourea) به عنوان ترکیبات ضد ایدز انجام دادند. ضریب تعیین گزارش شده برای کل سری داده‌ها به ترتیب ۰/۸۳۵ بودند [۱۵].

پاتل<sup>۲</sup> و همکارانش در سال ۲۰۱۰ مطالعات QSAR را بر روی ۴۶ مشتق دی‌هیدرو-آلکوکسی-بنزیل-اکسوپیریمیدین (DABOs) که به عنوان ترکیبات ضد HIV عمل می‌کنند، انجام دادند. آنها با توجه به علامت ضرایب توصیف‌کننده‌ها در مدل نهایی به دست آمده نتیجه گرفتند که استخلاف‌های آگریز و استخلاف‌های با شاخه بزرگتر، تأثیر مثبتی بر بازدارندگی دارند [۱۶].

در بخش اول این تحقیق، ارتباط کمی ساختار - فعالیت بر روی ثابت بازدارندگی فعالیت ۷۷ ترکیب دارویی که از لخته شدن خون جلوگیری می‌کنند و نیز فعالیت ضد HIV ۳۴ ترکیب دارویی در بخش دوم به کار گرفته شده است. مدل‌سازی به وسیله روش‌های رگرسیون خطی چندگانه و شبکه عصبی در مورد این سری داده‌ها تا کنون انجام نشده و این کار به عنوان تحقیق جدیدی مطرح می‌باشد. این روش‌ها توانایی بسیار بالایی در یافتن ارتباط خطی و غیر خطی ساختار - فعالیت برای ترکیبات مشابه دارند.

---

۱ . Ravichandran

۲ . Patel

## 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 فرآیندهای اندازه‌گیری و استخراج اطلاعات شیمیایی بیشتر از داده‌های موجود بسیار مفید می‌باشد. تعدادی از گروه‌های تجزیه‌ای در دهه‌ی ۷۰ میلادی از روش‌های ریاضی و آماری که امروزه به روش‌های کمومتریکس مشهور هستند، استفاده کرده‌اند.

عبارت کمومتریکس اولین بار توسط اسوانت ولد<sup>۱</sup> و بروس کووالسکی<sup>۲</sup> تعریف شد که منجر به تأسیس انجمن بین‌المللی کمومتریکس در سال ۱۹۷۴ گردید [۱۷]. پیش از این عباراتی مانند بیومتریکس<sup>۳</sup> و اکونومتریکس<sup>۴</sup> در زمینه‌های علوم زیستی و اقتصاد معرفی شده بودند و پس از آن انجمن بین‌المللی کمومتریکس تأسیس شد.

تعریف فعلی از کمومتریکس عبارت است از شاخه‌ای از شیمی که از روش‌های ریاضی و آماری برای انتخاب و طراحی مطلوب‌ترین روش‌ها و آزمایش‌ها، فراهم آوردن بیشترین اطلاعات شیمیایی به وسیله‌ی تجزیه و تحلیل داده‌های شیمیایی و شناخت سیستم‌های شیمیایی استفاده می‌کند [۱۸].

هاوری<sup>۵</sup> و هیرش<sup>۶</sup> در اوایل سال ۱۹۸۰ پیشرفت‌های کمومتریکس را در مراحل مختلف رده‌بندی کردند. اولین مرحله قبل از ۱۹۷۰ بوده که در آن تعدادی از روش‌های ریاضی توسعه یافته و در زمینه‌های مختلف ریاضی، علم رفتار و مهندسی به کار گرفته شدند. در این محدوده زمانی، شیمیدانان خود را به روش‌های آنالیز داده‌ها، یعنی محاسبه پارامترهای آماری مانند میانگین، انحراف استاندارد و حدود اطمینان محدود کرده بودند. هاوری و هیرش تحقیقات خود را به طور ویژه روی همبستگی تعداد زیادی از داده‌های شیمیایی و ویژگی‌های مولکولی مربوطه، متمرکز کردند. این اقدامات اولیه اساس یکی از شاخه‌های مهم کمومتریکس یعنی ارتباط کمی ساختار-فعالیت را فراهم کرد. دومین

---

۱. SvantWold

۲. Bruce R.Kowalski

۳. Biometrics

۴. Econometrics

۵. Howery

۶. Hirsch

مرحله تکامل کمومتریکس در سال ۱۹۷۰ واقع شد، هنگامی که کمومتریکس توجه شیمیدانان به خصوص متخصصین تجزیه را به خود جلب کرد، که به این ترتیب روش‌های بیشتری برای آنالیز داده‌های در دسترس و همچنین روش‌های نوینی برای به دست آوردن اطلاعات پیش روی خود می‌دیدند [۱۹].

هم‌اکنون کمومتریکس ابزاری قوی و کارآمد برای شیمیدانان به شمار رفته و به عنوان یک زیر-شاخه رشته شیمی در بسیاری از دانشگاه‌های آمریکا و اروپا و بعضی دانشگاه‌های چین و سایر کشورها تدریس می‌شود. پیشرفت‌ها در علوم کامپیوتر، فناوری اطلاعات، آمار و ریاضی کاربردی عناصر جدیدی را در کمومتریکس معرفی کرده است.

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 بازداري کروماتوگرافي گازی تعدادی از مشتقات بنزن در سال ۱۹۹۳ به چاپ رسیده‌است [۲۰].

با رشد سریع کمومتریکس در جهان، استفاده‌ی شیمیدانان تجزیه در ایران از کمومتریکس نیز در تحقیقاتشان افزایش یافت. تا سال ۲۰۰۵ حدود ۲۰۰ مقاله علمی توسط محققین کمومتریکس در ایران منتشر و بیش از ۸ گروه تحقیقاتی فعال در این زمینه مشغول فعالیت می‌باشند. انجمن کمومتریکس ایران نیز به عنوان یکی از شاخه‌های فعال انجمن شیمی ایران در سال ۲۰۰۰ تأسیس شد و همچنین سمینارهای داخلی و بین‌المللی در زمینه کمومتریکس در ایران برگزار می‌گردد. زمینه‌هایی که تاکنون بیشتر در ایران مورد توجه قرار گرفته اند شامل مطالعات (QSPR/QSAR)، کالیبراسیون چندمتغیره و هوش مصنوعی می‌باشند [۲۱].



## ۲-۲- 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 [۲۲]. 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 [۲۳]. 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 [۲۴].

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

۱ . Hansch

۲ . Quantitative descriptors

ترکیبات دارویی را داشتند. هانش و فوجیتا<sup>۱</sup> از روش‌های رگرسیون خطی چندگانه در مطالعات QSAR استفاده کردند [۲۵].

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>۲</sup> می‌گویند. 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 فعالیت ترکیبات شش مرحله دنبال می‌شود که این شش مرحله عبارتند از:

- ۱- جمع‌آوری سری داده‌ها
- ۲- ترسیم ساختار هندسی ترکیبات
- ۳- محاسبه‌ی توصیف‌کننده‌ها
- ۴- انتخاب بهترین توصیف‌کننده‌ها با روش‌های انتخاب متغیر برای ورود به مدل
- ۵- مدل‌سازی
- ۶- اعتبارسنجی مدل

---

۱ . Fujita  
۲ . Parametric methods

## ۲-۲-۱- انتخاب سری داده‌ها

اولین مرحله جمع‌آوری و انتخاب یک سری مولکولی است که مقادیر تجربی خاصیت یا فعالیت مورد نظر آنها در دسترس باشد. زمانی مدل‌سازی‌ها می‌تواند داده‌های مورد نظر را به خوبی پیش‌بینی کند که مولکول‌های مورد استفاده تقریباً مشابه یا گروهی از مشتقات یک ترکیب باشند. که این کار از طریق جستجو در مقالات صورت می‌گیرد.

هرچه سری داده‌ها بزرگتر و متنوع‌تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش‌بینی بالاتری خواهد داشت. اگر تعداد ترکیبات سری داده‌ها کم باشد، برای ارزیابی مدل از روش ارزیابی تقاطعی<sup>۱</sup> استفاده می‌شود.

## ۲-۲-۲- رسم وبهینه سازی ساختار ترکیبات

کمومتریکس از روش‌های مفیدی که قادر به استخراج اطلاعات موجود در داده هستند، استفاده می‌کند [۲۶]. به این منظور از شیمی محاسباتی<sup>۲</sup> استفاده می‌شود. شیمی محاسباتی ساختارهای مولکولی را به صورت پارامترهای عددی معرفی و رفتار آنها را با معادلات کوانتومی و فیزیک کلاسیک شبیه‌سازی می‌نماید. این امر به دانشمندان امکان می‌دهد که بتوانند از این طریق به اطلاعات مولکول از جمله ساختار هندسی، انرژی‌ها و ... دست پیدا کنند. در سال‌های اخیر روش‌های محاسباتی در بین شیمیدانان بسیار رواج یافته است. در این روش‌ها می‌توان به راحتی محاسبات را انجام داد، بدون اینکه از اصول اولیه و روش محاسبه آگاهی دقیق داشت.

روش‌های مختلفی که در شیمی محاسباتی برای بهینه‌سازی ساختار مولکول مورد استفاده قرار می‌گیرند شامل موارد زیر می‌باشند [۲۷]:

---

۱ . Cross-validation

۲ . Computational chemistry

۱- روش مکانیک مولکولی

۲- روش مکانیک کوانتومی

روش‌های مکانیک مولکولی بر مبنای روابط کلاسیک بنا شده است و به طور خیلی ساده ساختمان مولکولی به صورت گوی و فنر در نظر گرفته می‌شود. در این روش انرژی مولکول به صورت مجموعه‌ای از انرژی‌های کششی، خمشی، الکتروستاتیکی و ... بیان می‌شود. سپس با استفاده از برنامه‌های کامپیوتری مربوط به روش مکانیک مولکولی طول پیوندها، زوایای پیوندی و صورتبندی تغییر می‌کند تا ساختاری که عبارت انرژی مکانیک مولکولی را مینیمم کند، پیدا شود. این روش بهینه سازی بیشتر برای ماکرومولکول‌ها به کار می‌رود [۲۸] ولی در مکانیک کوانتومی از معادله شرودینگر<sup>۱</sup> برای محاسبه انرژی مولکول استفاده می‌شود و نسبت به روش مکانیک مولکولی دارای صحت بیشتری است زیرا اثرات الکترونی اعمال شده روی مولکول را نیز در محاسبات وارد می‌کند در حالی که مکانیک مولکولی الکترون‌ها را در نظر نمی‌گیرد. از طرف دیگر این محاسبات که پیچیده می‌باشند، وقت بیشتری را طلب می‌کند. از جمله روش‌های کوانتومی می‌توان به روش-های آغازین<sup>۲</sup> و روش‌های نیمه تجربی<sup>۳</sup> اشاره کرد [۲۷].

۱. در روش‌های آغازین از هیچ داده تجربی استفاده نشده است بلکه ثابت‌هایی مانند جرم الکترون، ثابت پلانک، ثابت سرعت نور و بار الکترون به کار گرفته می‌شوند و محاسبه انرژی مولکول از حل معادله شرودینگر با استفاده از یکسری تقریب به دست می‌آید. این روش چون کل الکترون‌ها را در محاسبات وارد می‌کند، بیشتر منحصر به مولکول‌های کوچک است و برای مولکول‌های بزرگتر محاسبات پیچیده‌تر و نیازمند ابر رایانه است [۲۱ و ۲۳].

۲. روش‌های نیمه تجربی روش‌هایی هستند که در برنامه‌هایی نظیر MOPAC و Hyperchem وارد شده‌اند. در این روش‌ها برخلاف روش‌های آغازین، فقط الکترون‌های لایه ظرفیت را در محاسبات وارد

---

۱ . Schrodinger Equation

۲ . Ab Initio

۳ . Semi empirical

می‌کنند. در نتیجه زمان محاسبات در این روش‌ها کوتاهتر از روش‌های آغازین است و می‌تواند برای مولکول‌های بزرگتر به کار رود. روش‌های نیمه تجربی مختلفی مانند چشم‌پوشی کامل از انتگرال‌های همپوشانی دیفرانسیلی (CNDO)<sup>۱</sup>، چشم‌پوشی از برخی از انتگرال‌های همپوشانی دیفرانسیلی (INDO)<sup>۲</sup>، اصلاح شده صرف نظر کردن دو اتمی از انتگرال همپوشانی (MNDO)<sup>۳</sup> و AM1<sup>۴</sup> وجود دارد [۳۰]. روش AM1 جزء دقیق‌ترین روش‌های نیمه تجربی است که در این تحقیق از آن استفاده شده است [۲۷].

## ۲-۲-۳- محاسبه توصیف‌کننده‌ها

اگر چه چندین پارامتر مولکولی از زمان شروع شیمی کوانتومی و نظریه گراف تعریف شدند اما اصطلاح "توصیف‌کننده مولکولی" با پیشرفت مدل‌های ارتباط دهنده ساختار- خاصیت مشهور شد. عدد پلات<sup>۵</sup> [۳۱] و شاخص وینر<sup>۶</sup> [۳۲] که در سال ۱۹۴۷ تعریف شدند، در بعضی مراجع به عنوان اولین توصیف‌کننده‌های مولکولی در نظر گرفته می‌شوند. توصیف‌کننده‌های مولکولی مقادیر عددی هستند که ساختار یا شکل مولکول را توصیف می‌کنند و به پیش‌بینی فعالیت و خصوصیات مولکول‌ها در آزمایش‌های پیچیده کمک می‌کنند. به عبارت دیگر توصیف‌کننده‌های مولکولی نتیجه نهایی یک فرآیند منطقی و ریاضی هستند که اطلاعات شیمیایی مربوط به ساختار یک مولکول را به اعداد تبدیل می‌کنند. این اعداد می‌توانند برای تفسیر خواص مولکولی استفاده شوند و یا برای پیش‌بینی تعدادی از ویژگی‌های مولکولی در یک مدل شرکت کنند.

برخی از ویژگی‌های لازم برای یک توصیف‌کننده مولکولی مناسب عبارتند از:

- 
- ۱ . Complete Neglect of Differential Overlap
  - ۲ . Intermediate Neglect of Differential Overlap
  - ۳ . Modified Neglected of Diatomic Overlap
  - ۴ . Austian Method 1
  - ۵ . Platt Number
  - ۶ . Wiener Index

- ❖ ساده بودن
- ❖ مستقل بودن
- ❖ تفسیر ساختار مولکول (غنی بودن از نظر اطلاعات)
- ❖ عدم همبستگی با سایر توصیف‌کننده‌ها
- ❖ تغییر منظم با تغییر تدریجی در ساختارها
- ❖ وابستگی صحیح به اندازه مولکول
- ❖ تمایز بین ایزومرهای مختلف مولکول [۳۳]

#### ۲-۲-۴- انتخاب توصیف‌کننده‌های مناسب

یکی از روش‌های انتخاب توصیف‌کننده‌های مناسب، روش مرحله‌ای<sup>۱</sup> است که از طریق محاسبه ضریب همبستگی<sup>۲</sup> بین توصیف‌کننده و فعالیت بیولوژیکی انجام می‌شود. اگر توصیف‌کننده‌ای دارای ضریب همبستگی برابر یک باشد، بدین معنی است که می‌تواند خواص مورد نظر را به درستی توصیف کند و ضریب همبستگی صفر بدین معنی است که هیچ ارتباطی بین توصیف‌کننده و خاصیت دارویی وجود ندارد. پس توصیف‌کننده‌هایی با ضرایب همبستگی بزرگ در برآزش منحنی، برای مدل‌سازی استفاده می‌شود. از دیگر روش‌های انتخاب توصیف‌کننده‌ها می‌توان به الگوریتم ژنتیک<sup>۳</sup> و روش جایگزینی<sup>۴</sup> اشاره کرد.

#### ۲-۲-۵- انتخاب مدل مناسب

مدل یک رابطه ریاضی است که بیانگر رابطه بین متغیر وابسته و متغیر(های) مستقل می‌باشد و با استفاده از روش‌های آماری گوناگون، می‌توان آن را به دست آورد. مرحله اصلی در مطالعات QSAR ساختن مدل می‌باشد. مدل‌های QSAR را می‌توان بر اساس پارامترهایی که به صورت تجربی بدست

---

۱ . Stepwise  
 ۲ . Correlation coefficient  
 ۳ . Genetic Algorithm  
 ۴ . Replacement Method

آمده و یا از طریق تئوری محاسبه شده‌اند، به زبان ریاضی بیان نمود. به طور کلی اگر نتایج این نوع مطالعات همبستگی 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 [۳۴]. چندین روش متفاوت برای ساختن مدل QSAR وجود دارد که می‌توان به رگرسیون خطی چندگانه (MLR)، آنالیز اجزای اصلی (PCA) و حداقل مربعات جزئی (PLS) به عنوان روش‌های خطی و شبکه عصبی مصنوعی (ANN) به عنوان یک روش غیرخطی اشاره کرد.

## ۲-۲-۶- 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 [۳۵]. 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 به طور 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.

ساده‌ترین و متداول‌ترین مدلی که برای ارتباط چند متغیر به کار می‌رود، مدل خطی چندگانه است، که با رابطه خطی زیر تعریف می‌شود:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots + \beta_p X_p \quad (1-2)$$

<sup>۱</sup>. Francis Galton

که در آن  $Y$  متغیر وابسته (فعالیت دارویی)،  $X_i$  متغیر مستقل (توصیف‌کننده) و  $\beta_i$  ضرایب رگرسیون نامیده می‌شوند. باید توجه کرد که مقدار ضریب متغیر مستقل در معادله، بیانی از میزان اهمیت آن متغیر در پیش بینی مقادیر متغیر وابسته است.

غالباً روش حداقل مربعات برای ساخت معادله خطی به کار می‌رود و مدلی که با این روش بدست می‌آید، دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است، هر چند ممکن است همه‌ی مشاهدات تجربی بر روی این معادله واقع نشوند. 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 ( $y$ ) از هر خط مفروض دیگر است.

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

آماره  $R$ : ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی ضریب همبستگی<sup>۱</sup> آنهاست. ضریب همبستگی دو متغیر  $x, y$  با رابطه (۲-۲) تعریف می‌شود. مقدار این آماره بین ۱ تا -۱ متغیر است. مقدار بزرگ‌تر آن نشان دهنده‌ی این مسئله است که ارتباط خطی بیشتری میان متغیر وابسته و متغیرهای مستقل وجود دارد.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (2-2)$$

آماره  $R^2$ : به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته توضیح داده شده توسط متغیر مستقل است و به آن ضریب تعیین<sup>۲</sup> نیز می‌گویند، که به این صورت محاسبه می‌شود:

---

۱ . Correlation Coefficient  
۲ . Determination Coefficient



$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (3-2)$$

که  $SSR^1$  طبق رابطه (۲-۴) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (4-2)$$

$SST^2$  طبق رابطه (۲-۵) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5-2)$$

$SSE^3$  نیز مبین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6-2)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7-2)$$

طبق رابطه (۲-۷) اگر به ازای تمام نقاط  $y_i = \hat{y}_i$  باشد، مقدار  $R^2$  برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار  $R^2$  از یک کوچکتر شود.

---

۱ . Sum Square Regression  
 ۲ . Sum Square Total  
 ۳ . Sum Square Error

آماره F: آزمون F یا آزمون فیشر، آزمون معنادار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین مربعات رگرسیون ( $MSR$ ) به میانگین مربعات باقیمانده‌ها ( $MSE$ ) است. و از رابطه‌ی (۸-۲) بدست می‌آید:

$$F = \frac{MSR}{MSE} = \frac{SSR/df_m}{SSE/df_{res}} \quad (۸-۲)$$

$$df_m = p, \quad df_{res} = n - p - 1$$

$df_m$  درجه آزادی مدل،  $p$  تعداد متغیرهای مستقل مدل،  $n$  تعداد کل ترکیبات مربوط به مدل،  $df_{res}$  درجه آزادی باقیمانده‌ها است.  $SSR$ ،  $SSE$  در روابط قبل توضیح داده شد.

مجموع مربع باقیمانده‌ها<sup>۱</sup> ( $PRESS$ ): برابر مجموع مربعات تفاوت بین مقدار کمیت مشاهده شده ( $y_i$ ) و مقدار تخمین زده شده ( $\hat{y}_i$ ) است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (۹-۲)$$

خطای استاندارد پیش‌بینی<sup>۲</sup> ( $SEP$ )

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (۱۰-۲)$$

خطای مطلق میانگین<sup>۳</sup> ( $MAE$ )

$$(۱۱-۲)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

۱ . Predictive Residual Sum of Squares

۲ . Standard Error of Prediction

۳ . Mean Absolute Error

خطای نسبی پیش بینی<sup>۱</sup> (REP)

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}} \quad (12-2)$$

میانگین مربع خطاها<sup>۲</sup> (MSE)

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n} \quad (13-2)$$

میانگین خطای نسبی<sup>۳</sup> (MRE)

$$MRE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right|}{n} \times 100 \quad (14-2)$$

در روش MLR چندین روش رگرسیون مختلف برای ساختن مدل وجود دارد که عبارتند از:

۲-۲-۶-۱-روش ورود اجباری<sup>۴</sup>

فرآیندی برای انتخاب متغیرها است که طی آن تمام توصیف‌کننده‌ها در یک مرحله وارد می‌شوند. از آنجا که در این روش اثرات مثبت و منفی متغیرهای مستقل و میزان همبستگی آن‌ها با خاصیت مورد مطالعه (متغیر وابسته) برای ورود در مدل نادیده گرفته می‌شود، روش چندان مناسبی نمی‌باشد.

---

۱ . Relative Error of Prediction

۲ . Mean Square Error

۳ . Mean Relative Error

۴ . Enter

## ۲-۲-۶-۲- روش جلو برنده<sup>۱</sup>

تکنیکی مرحله‌ای برای انتخاب متغیرها است که طی آن متغیرها یکی پس از دیگری وارد مدل می‌شوند. در هر مرحله اثر ورود هر متغیر مورد بررسی قرار می‌گیرد یعنی اگر با ورود متغیر جدید تغییر در آماره  $F$  و  $R$  از نظر آماری معنادار باشد، متغیر مورد نظر در مدل نگه داشته می‌شود و در غیر این صورت آن متغیر از مدل حذف می‌شود. این روند تا زمانی که دیگر متغیری وجود نداشته باشد که دارای شرط ورود باشد و باعث بهبود مدل گردد، ادامه دارد.

## ۲-۲-۶-۳- روش عقب برنده<sup>۲</sup>

در این تکنیک تمام توصیف‌کننده‌ها در مرحله مدل سازی وارد می‌شوند. در مرحله بعد، اثر حذف تک تک متغیرها از مدل مورد بررسی قرار می‌گیرد. به این ترتیب که اگر با حذف یک متغیر مستقل آماره  $F$  و  $R$  به مقدار قابل توجهی تغییر نکرد، آن متغیر از مدل حذف می‌شود. این کار برای همه متغیرها تکرار می‌شود تا در نهایت متغیرهای مستقلی که بیشترین ارتباط با متغیر وابسته داشته باشند، به دست آیند.

## ۲-۲-۶-۴- روش مرحله‌ای

این روش ترکیبی از روش‌های جلو برنده و عقب برنده می‌باشد. بدین ترتیب که نخستین متغیر مستقل به روش انتخاب جلو برنده وارد مدل می‌گردد (یعنی متغیری که بیشترین همبستگی با متغیر وابسته داشته باشد) در مرحله دوم تمامی مدل‌های ممکن با دو متغیر انتخاب می‌شود، به طوری که یکی از متغیرها همان  $X_1$  و متغیر دیگر از سایر متغیرهای خارج از مدل باشد. مقدار  $R$  برای هر یک از مدل‌های دو متغیره محاسبه می‌شود و مدلی که دارای بزرگترین ضریب همبستگی است، انتخاب می‌شود. در این هنگام جهت بررسی معنادار بودن ضریب رگرسیون دو متغیر انتخاب شده، آزمون  $F$  انجام

---

۱ . Forward

۲ . Backward

می‌شود و متغیری که ضریب رگرسیون آن از نظر آماری بی‌معنا باشد، از مدل حذف می‌شود. در صورت معنادار بودن هر دو ضریب، آن مدل دو متغیره انتخاب می‌شود. در مرحله بعدی تمامی مدل‌های ممکن با سه متغیر انتخاب می‌شوند و کلیه مراحل دوم تکرار می‌شود. در این روش با توجه به فرایندی که رخ می‌دهد، هیچ تضمینی وجود ندارد که متغیر وارد شده در مدل تا مدل نهایی باقی بماند [۳۶].

## ۲-۲-۷- تکنیک‌های اعتبارسنجی

این تکنیک‌ها برای ارزیابی اعتبار مدل‌های به دست آمده به کار می‌روند. تکنیک‌های اعتبارسنجی برای بررسی قدرت پیش‌بینی مدل‌ها استفاده می‌شوند، بدین معنی که توانایی آنها را برای انجام پیش‌بینی‌های مطمئن برای موارد جدیدی که پاسخ آنها ناشناخته است می‌سنجند. یک شرط لازم برای اعتبار یک مدل رگرسیون آن است که ضریب تعیین ( $R^2$ ) یا ضریب همبستگی چند متغیره ( $R$ ) آن تا حد ممکن به یک نزدیک باشد و خطای استاندارد ( $SE$ ) آن کوچک باشد. به هر حال این شرط (توانایی برازش)، زمانی که مدلها برازش نزدیک‌تری ( $SE$  کوچک‌تر و  $R^2$  بزرگ‌تر) را برای تعداد زیادتری از پارامترها و متغیرها در مدل به دست می‌دهند، برای اعتبارسنجی مدل کافی نیست؛ زیرا این پارامترها ارتباطی به توانایی مدل برای ایجاد پیش‌بینی‌های مطمئن بر روی داده‌های بعدی ندارند. مسائل دیگر برای اعتبارسنجی مدل‌ها زمانی پیش می‌آیند که لازم است مدل‌هایی با متغیرهای اندک، با استفاده از فرآیندهای بر پایه انتخاب متغیر به دست آیند. در حقیقت زمانی که یک مجموعه با تعداد زیادی از توصیف‌کننده‌ها برای انتخاب از آنها در دسترس باشد، مدل‌های به دست آمده دارای برازش خوب اما همراه با همبستگی شانسی<sup>۱</sup> هستند، یعنی منحنی مقادیر نظری محاسبه شده از مدل بر حسب مقادیر تجربی در یک خط مستقیم واقع شده بدون این که توانایی پیش‌بینی داشته باشد.

---

۱ . Chance Correlation

برای جلوگیری از مدل‌های با همبستگی شانسی، یک اعتبارسنجی متفاوت باید انجام شود. بهتر است یک سری ارزیابی مناسب از ترکیباتی که در فرایند آموزش و بهینه‌سازی دخالت نداشته (سری ارزیابی بیرونی<sup>۱</sup>)، در دسترس باشد. تعدادی از تکنیک‌های اعتبارسنجی در زیر آمده است.

### ۲-۲-۷-۱- ارزیابی متقاطع

رایج‌ترین تکنیک اعتبارسنجی است که در آن در هر بار یک یا یک گروه کوچک از داده‌ها کنار گذاشته شده و سپس برای داده‌های باقی مانده، مدلی محاسبه می‌شود و پاسخ برای داده‌های کنارگذاشته شده از روی این مدل پیش‌بینی می‌شود.

### ۲-۲-۷-۲- اعتبارسنجی بیرونی<sup>۲</sup>

تکنیکی است که در آن یک سری تست نیز برای انجام یک بررسی اضافی بر روی قدرت پیش‌بینی مدل به دست آمده از سری آموزش و ارزیابی، در نظر گرفته می‌شود. سری تست نباید دخالتی در روند آموزش و انتخاب مدل داشته باشد.

### ۲-۲-۷-۳- آزمون Y- تصادفی<sup>۳</sup>

این تکنیک برای مطالعه همبستگی شانسی مدل طراحی شده است. اگر در مدلی متغیرهای مستقل به صورت تصادفی با متغیرهای پاسخ همبستگی داشته باشند، گفته می‌شود دارای همبستگی شانسی است. در این آزمون، مقادیر پاسخ (که در اینجا بردار Y نامیده می‌شود) به صورت تصادفی در محدوده پاسخ‌ها، تغییر داده می‌شود. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از یکی از شاخص‌های آماری (معمولاً  $R^2$ ) مورد بررسی قرار می‌گیرد. اختلاف زیاد بین

---

۱ . External Validation Set

۲ . External Validation

۳ . Y-Randomization

شاخص آماری به دست آمده از این روش و شاخص آماری به دست آمده از مدل اصلی، نشان‌دهنده عدم وجود همبستگی شانس می‌باشد. این فرایند چندین بار تکرار می‌شود [۳۷].

## ۲-۳- نرم افزارهای مورد نیاز

در این تحقیق از بسته‌های نرم افزار 8 HyperChem برای رسم ساختار ترکیبات مورد بررسی، Dragon برای محاسبه توصیف‌کننده‌های مولکولی، SPSS 17 برای انجام آنالیزهای آماری رگرسیون‌های خطی چندگانه و MATLAB 2008a به منظور برنامه نویسی شبکه عصبی مصنوعی استفاده شد، که در ادامه به معرفی آنها پرداخته خواهد شد.

## ۲-۳-۱- بسته‌ی نرم افزاری Hyperchem

از بسته‌ی نرم‌افزاری Hyperchem [۲۷] برای رسم شکل مولکول‌ها و بهینه‌سازی ساختار با استفاده از روش‌های کوانتومی و مکانیکی، استفاده می‌شود. این بسته، یک نرم افزار مدل‌سازی و شبیه‌سازی است که به ما اجازه می‌دهد محاسبات پیچیده‌ی شیمیایی را با رایانه انجام دهیم. از این نرم افزار می‌توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها و زوایای پیچشی را بدست آورد. داده‌های حاصل از این نرم‌افزار را می‌توان به عنوان ورودی به سایر نرم افزارها معرفی نمود. به عنوان مثال ساختار سه بعدی مولکول‌ها که با استفاده از روش مکانیک کوانتومی در Hyperchem بهینه شده است، را جهت محاسبه توصیف‌کننده‌های مختلف به وسیله نرم‌افزار Dragon، به کار برد.

## ۲-۳-۲- بسته‌ی نرم افزاری Dragon

از بسته نرم افزاری Dragon برای محاسبه توصیف‌کننده‌های مولکولی استفاده می‌شود که در سال ۲۰۰۱ توسط گروه تحقیقاتی کمومتریکس دانشگاه میلانو طراحی شده است [۳۸]. این توصیف‌کننده‌های مولکولی به منظور ارزیابی روابط ساختار-فعالیت یا ساختار-ویژگی مولکولی به کار می‌روند.

نرم افزار Dragon قادر به محاسبه ۱۴۸۱ توصیف کننده می باشد که به ۱۸ دسته اصلی تقسیم می - شوند که در جدول ۱-۲ آورده شده است. کاربر علاوه بر محاسبه ساده ترین نوع اتم ها، گروه های عاملی و شمارش اجزا می تواند تعداد زیادی توصیف کننده های توپولوژیکی و هندسی را نیز توسط این نرم افزار محاسبه کند.

برای محاسبه توصیف کننده به کمک این نرم افزار لازم است ساختار هندسی بهینه مولکول مورد استفاده قرار گیرد. برای این منظور می توان ساختار بهینه مولکول ها را به صورت فایل هایی با فرمت mol، solf، hin و ... به عنوان اطلاعات ورودی به کار برد.

جدول ۱-۲- انواع توصیف کننده های محاسبه شده توسط نرم افزار Dragon

|   |   |
|---|---|
| ۱. توصیف کننده های زیر ساختاری <sup>۱</sup> | ۱۰. توصیف کننده های هندسی <sup>۲</sup>    |
| ۲. توصیف کننده های توپولوژیکی <sup>۳</sup>  | ۱۱. توصیف کننده های RDF <sup>۴</sup>      |
| ۳. شمارنده های مولکولی مورس <sup>۵</sup>    | ۱۲. توصیف کننده های سه بعدی <sup>۶</sup>  |
| ۴. توصیف کننده های BCUT <sup>۷</sup>        | ۱۳. توصیف کننده های WHIM <sup>۸</sup>     |
| ۵. شاخص های بار <sup>۹</sup>                | ۱۴. توصیف کننده های GETAWAY <sup>۱۰</sup> |
| ۶. خود ارتباطی های دو بعدی <sup>۱۱</sup>    | ۱۵. گروه های عاملی <sup>۱۲</sup>          |
| ۷. توصیف کننده های بار <sup>۱۳</sup>        | ۱۶. اجزای میان اتمی <sup>۱۴</sup>         |
| ۸. شاخص های آروماتیسیت <sup>۱۵</sup>        | ۱۷. توصیف کننده های تجربی <sup>۱۶</sup>   |
| ۹. پروفایل های مولکولی راندیک <sup>۱۷</sup> | ۱۸. خصوصیات مولکولی <sup>۱۸</sup>         |

- 
- ۱ . Constitutional Descriptors
  - ۲ . Geometrical Descriptors
  - ۳ . Topological Descriptors
  - ۴ . RDF Descriptors
  - ۵ . Molecular walk count
  - ۶ . Molecule representation of structure backed on electron diffraction
  - ۷ . BCUT Descriptors
  - ۸ . Weighted holistic invariant molecular descriptors
  - ۹ . Charge indices
  - ۱۰ . Geometry, topology, and atom weights assembly
  - ۱۱ . 2D autocorrelation
  - ۱۲ . Functional group
  - ۱۳ . Charge descriptors.
  - ۱۴ . Atom-centered fragments
  - ۱۵ . Aromatic indices
  - ۱۶ . Empirical descriptor
  - ۱۷ . Randic Molecular profiles
  - ۱۸ . Molecular Properties



## ۲-۳-۳- بسته‌ی نرم‌افزاری SPSS

کلمه SPSS مخفف عبارت Statistical Package for Social Science نرم‌افزار آماری برای علوم اجتماعی می‌باشد. این بسته نرم‌افزاری، توانایی تحلیل کلی اطلاعات را دارا بوده و عمومی‌ترین نرم‌افزار آماری است. امروزه آمار، مجموعه‌ای از مفاهیم و روش‌هایی است که در هر زمینه پژوهشی، برای گردآوری و تحلیل اطلاعات مربوط به آن و انجام نتیجه‌گیری‌ها، در شرایطی که عدم قطعیت و تغییر وجود دارد، به کار می‌رود. SPSS دارای توانایی‌های بسیار بالایی است که به تعدادی از آنها اشاره می‌شود [۳۹].

- ❖ تهیه خلاصه‌های آماری مانند گراف‌ها، جداول، آماره‌ها و ...
- ❖ تهیه انواع جداول سفارشی مانند جداول فراوانی، فراوانی تجمعی، درصدفراوانی و ...
- ❖ محاسبه انواع آماره‌های توصیفی
- ❖ قابلیت مبادله اطلاعات با نرم‌افزارهای دیگر
- ❖ برآزش انواع مختلف رگرسیون

## ۲-۳-۴- نرم‌افزار MATLAB

MATLAB نرم‌افزاری با کارایی زیاد برای محاسبات تخصصی است. اولین نسخه‌های آن در دانشگاه نیومکزیکو و استنفورد در سال‌های دهه ۱۹۷۰ میلادی جهت حل مسائل نظریه ماتریس‌ها، جبر خطی و آنالیز عددی به وجود آمد. به تدریج و با افزودن امکانات مختلف، MATLAB به عنوان یک نرم‌افزار قوی و در عین حال ساده برای کاربران تبدیل شد. این نرم‌افزار امکان حل مسائل تخصصی، به خصوص مسائلی که به صورت ماتریسی هستند را فراهم می‌کند. یکی از قابلیت‌های آن برنامه نویسی در محیطی ساده است که مسائل و راه‌حل‌ها با علائم ریاضی مشابهی تعریف می‌شوند. کاربردهای نوعی آن شامل ریاضیات و الگوریتم محاسباتی، شبیه‌سازی، مدل‌سازی و آنالیز اطلاعات

و... می‌باشد. یکی از کاربردهای مهم آن که مورد استقبال شیمیدانان قرار گرفته است، امکان استفاده از آن برای آنالیز آماری و مدل‌سازی داده‌های شیمیایی است. مدل‌سازی به روش شبکه عصبی مصنوعی از جمله کاربردهای مهم دیگر آن است که در تحقیق حاضر مورد استفاده قرار گرفته است. اطلاعات شیمیایی ترکیبات به صورت ماتریس به نرم افزار ارائه می‌شوند. در محیط MATLAB با استفاده از فرامین، آرایه‌ها و امکانات ویژه، مدل‌سازی غیر خطی فعالیت شیمیایی ترکیبات امکان پذیر است. سایر روش‌های کمومتریکس مانند آنالیز رگرسیون اجزای اصلی، روش کمترین مربعات جزئی، الگوریتم ژنتیکی و غیره نیز توسط این نرم افزار قابل اجرا هستند [۴۰].

## فصل سوم

### شبکه عصبی مصنوعی

### ۳-۱- مقدمه‌ای بر شبکه‌های عصبی مصنوعی

شبکه‌ی عصبی مصنوعی، شبکه‌ای محاسباتی است که، تلاش می‌کند به منظور پردازش اطلاعات سیستمی مانند سلول‌های عصبی زیستی را شبیه‌سازی کند. سیستم عصبی موجود زنده از تجمع و آرایش ویژه‌ی عناصر پردازش‌کننده‌ای 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 پاسخ دهی 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 ذخیره‌سازی 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 نموده 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

---

<sup>۱</sup> . Neuron

تنظیم یا همان آموزش شبکه بر مبنای تطابق<sup>۱</sup> و همسنجی میان ورودی و هدف خروجی شبکه بر هدف منطبق می‌گردد. استفاده از زوج داده‌های ورودی و هدف بیانگر یادگیری با ناظر<sup>۲</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، نوروفیزیولوژیست‌ها سعی کردند سیستم یادگیری و تجزیه و تحلیل مغز را کشف کنند و از سویی دیگر ریاضیدانان تلاش می‌نمودند 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>۳</sup> و پیتز<sup>۴</sup> مطرح شد [۴۱] و نخستین کاربرد عملی شبکه‌های عصبی برای شناسایی الگوها در اواخر دهه ۵۰، زمانیکه روزنبلات<sup>۵</sup> در سال ۱۹۵۸ شبکه پرسپترون<sup>۶</sup> را معرفی کرد، مطرح شد [۴۲]. سال ۱۹۶۹ آغاز افول موقت شبکه‌های عصبی بود و این به خاطر عدم توانایی شبکه‌های عصبی در حل مسائل غیرخطی بود. 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 جداکنند. در سال ۱۹۸۲ هاپفیلد<sup>۷</sup> با معرفی شبکه‌های چند لایه و برگشتی، راه حلی برای حل مسائل غیرخطی ارائه کرد [۴۳] و سپس دیوید راملهارت<sup>۸</sup> و جیمز مک کلاند<sup>۹</sup> الگوریتم پس‌انتشار<sup>۱۰</sup> خطا را در سال ۱۹۸۶ مطرح کردند [۴۴]. 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، به کار روند.

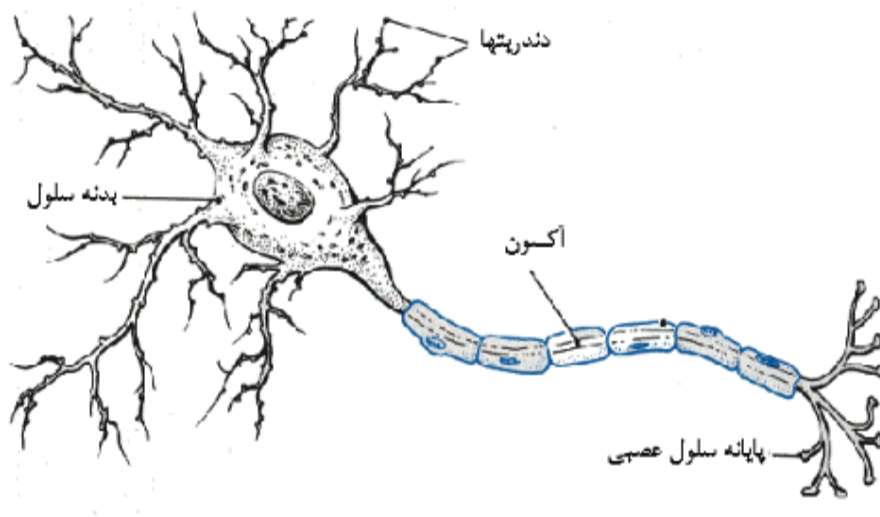
---

۱ . Adaptive  
 ۲ . Supervised learning  
 ۳ . McCulloch  
 ۴ . Pitts  
 ۵ . Rosenblatt  
 ۶ . Perceptrons  
 ۷ . Hopfield  
 ۸ . Rumelhart  
 ۹ . James McClelland  
 ۱۰ . Back Propagation

### ۳-۲- 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 ما بین نرون‌ها توسط سیناپس‌ها برقرار می‌شود.

درحقیقت 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 نموده و آنها را با هم جمع می‌کنند. 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 شده و می‌تواند باعث ارسال پیام‌های الکتریکی بین نرون‌های دیگر گردد[۴۴]. شکل (۳-۱) نمای کلی یک نرون را نشان می‌دهد.

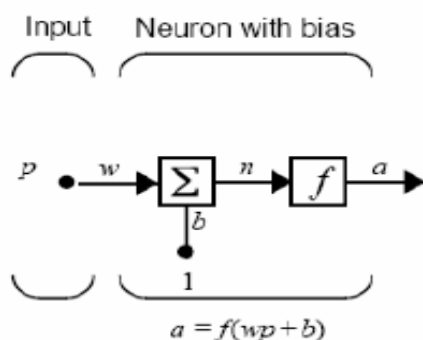


شکل ۳-۱- ساختار نرون طبیعی

- 
۱. Dendrite
  ۲. Cell body(Soma)
  ۳. Axon

### ۳-۳- 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 [۴۵]. 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. معمولاً برای نشان دادن یک مدل ریاضی ساده از نرون شبکه عصبی مصنوعی از شکل زیر استفاده می‌شود:



شکل ۳-۲- مدل ریاضی نرون

در شکل فوق کمیت‌های  $p$  و  $a$  به ترتیب ورودی و خروجی نرون می‌باشد. میزان تأثیر ورودی  $p$  روی خروجی  $a$  به وسیله پارامتر وزن ( $w$ ) تعیین می‌شود. ورودی دیگر یک مقدار ثابت ۱ است که در جمله بایاس<sup>۱</sup> ضرب شده و سپس با  $wp$  جمع می‌شود. این حاصل جمع، ورودی خالص  $n$  برای تابع محرک  $f$  را تشکیل می‌دهد.

$$n = wp + b \quad (۱-۳)$$

سپس خروجی نرون به کمک معادله زیر به دست می‌آید:

$$a = f(wp + b) \quad (۲-۳)$$

۱ . Bias

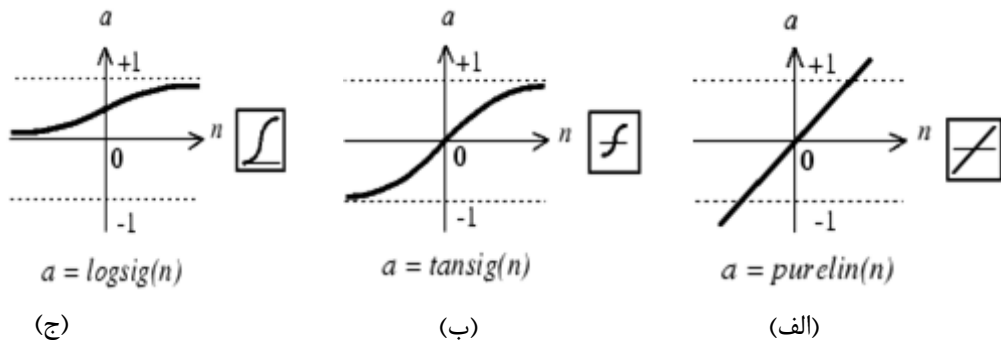
بنابراین اجزای یک شبکه‌ی عصبی مصنوعی عبارتند از:

- ❖ ورودی‌ها: می‌توانند خروجی سایر لایه‌ها بوده و یا به حالت خام در اولین لایه به صورت داده‌های عددی و رقمی، متون ادبی، تصویر و یا شکل باشند.
- ❖ وزن‌ها: میزان تأثیر ورودی  $X_i$  بر خروجی  $y$ ، توسط وزن اندازه‌گیری شده 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 نموده و در شبکه‌های چند نرونی میزان سطح فعالیت نرون در لایه‌های درونی را مشخص می‌سازد.
- ❖ تابع انتقال: وظیفه محدودسازی خروجی گره به یک بازه متناهی را بر عهده دارد. انواع متفاوتی از توابع انتقال 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. از توابع انتقال مشتق پذیری که مخصوصاً در شبکه‌های عصبی پس انتشار به کار می‌رود و در این تحقیق نیز مورد استفاده قرار گرفته، می‌توان به توابع انتقال خطی، لگاریتم سیگموئیدی<sup>۱</sup> و یا تانژانت سیگموئیدی<sup>۲</sup> اشاره کرد. تابع انتقال خطی همان مقدار ورودی را به عنوان خروجی بر می‌گرداند (شکل ۳-۳ الف). تابع انتقال تانژانت سیگموئید (tansig) مقادیر ورودی را در محدوده  $-\infty$  و  $+\infty$  دریافت کرده و خروجی بین ۱ و -۱ تولید می‌نماید (شکل ۳-۳ ب) و تابع انتقال لگاریتم سیگموئید (logsig) مقادیر ورودی را در محدوده  $-\infty$  و  $+\infty$  دریافت کرده و خروجی بین ۰ و ۱ تولید می‌نماید (شکل ۳-۳ ج).

---

۱. Log-sigmoid  
۲. Tan-sigmoid





شکل ۳-۳- برخی از توابع انتقال مورد استفاده در شبکه‌های عصبی [۴۶]

❖ خروجی: منظور از خروجی، پاسخ مسأله است.

### ۳-۴- معماری شبکه‌های عصبی<sup>۱</sup>

اگر بردار ورودی  $p$  را با عناصر  $p_1, p_2, \dots, p_R$  و بردار وزن‌ها،  $w$ ، را با عناصر  $w_{1,1}, w_{1,2}, \dots, w_{1,R}$  در نظر بگیریم، برای اعمال مقادیر وزن در مقادیر ورودی دو بردار  $p$  و  $w$  در هم ضرب ماتریسی می‌کنیم، بنابراین ورودی تابع انتقال  $F$  یعنی  $n$  به صورت زیر خواهد بود:

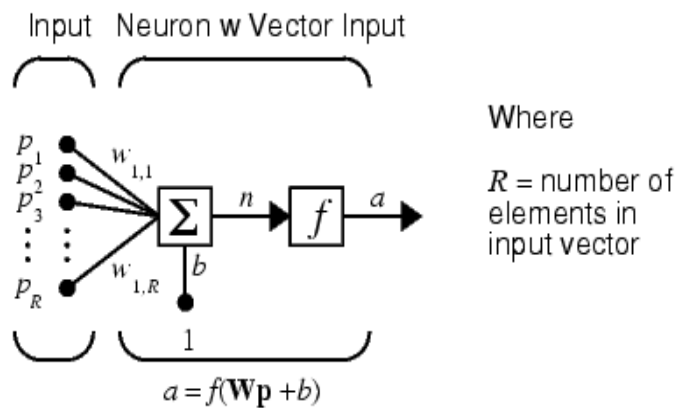
$$n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b \quad (۳-۳)$$

و خروجی نرون از طریق رابطه زیر بدست می‌آید:

$$a = f(w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b) \quad (۴-۳)$$

در نهایت  $b$  مقدار بایاس است که به حاصلضرب ماتریسی  $w$  در  $p$  اضافه شده است. به ترکیب وزن‌ها، بایاس، تابع انتقال و عملیات ضرب و جمع انجام شده یک لایه از شبکه می‌گویند. شایان توجه است که بردار ورودی در بیشتر نوشتارها، یک لایه محسوب نمی‌شود. اگر از تابع انتقال خاصی استفاده شود، نماد آن تابع در کادر مربوط به تابع قرار می‌گیرد. شکل ۳-۴ نمایش خلاصه‌ای از یک نرون تک-لایه را نشان می‌دهد.

<sup>۱</sup> . Network architecture



شکل ۳-۴- نمایش یک نرون تک‌لایه

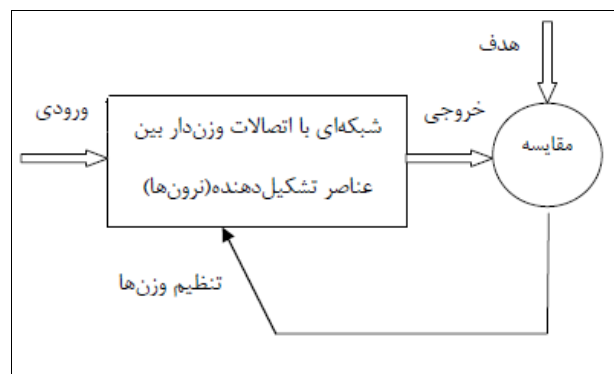
دو یا چند نرون می‌توانند در یک لایه با هم ترکیب شوند و یک شبکه می‌تواند از یک یا چند لایه این‌چنینی تشکیل شود.

### ۳-۵- آموزش شبکه عصبی

معمولاً شبکه‌های عصبی تنظیم می‌شوند یا آموزش می‌یابند تا اعمال یک ورودی خاص منجر به دریافت خروجی خاص موردنظر یا همان هدف شود. همانطور که در شکل ۳-۶ دیده می‌شود شبکه بر مبنای تطابق و همسنگی بین ورودی و هدف سازگار می‌شود تا اینکه خروجی شبکه و هدف بر هم منطبق گردند. عموماً تعداد زیادی از این زوج‌های ورودی و خروجی به کار گرفته می‌شوند تا در این روند که یادگیری تحت نظارت<sup>۱</sup> نامیده می‌شود، شبکه آموزش داده شود. در یادگیری تحت نظارت قاعده یادگیری با استفاده از مجموعه‌ای از مثال‌ها (مجموعه آموزشی) به دست می‌آید و شبکه را آموزش می‌دهد. زوج داده‌های زیر را در نظر بگیرید:

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\}$$

<sup>۱</sup>. Supervised Training



شکل ۳-۵- شیوه ی یادگیری تحت نظارت

که  $pQ$  یک ورودی شبکه و  $tQ$  هدف مورد نظر متناظر با هر ورودی می باشد. زمانی که ورودی به شبکه اعمال می شود، خروجی آن با هدف مقایسه می شود. سپس قواعد یادگیری برای تنظیم وزن ها و بایاس ها به کار برده می شود تا خروجی شبکه را به هدف نزدیک نماید. اما در یادگیری نظارت نشده<sup>۱</sup> ورودی ها و بایاس ها تنها در مقابل ورودی شبکه اصلاح می شوند و در واقع هیچ هدفی وجود ندارد. این نوع الگوریتم ها غالباً برای دسته بندی سری داده ها مورد استفاده قرار می گیرند. عموماً برای آموزش شبکه های عصبی از قواعد یادگیری نظارت شده استفاده می شود اما می توان شبکه ها را با روش آموزش غیر نظارتی آموزش داد [۴۶].

### ۳-۶- انواع شبکه های عصبی

#### ۳-۶-۱- شبکه های جلوبرنده (پیش خور)<sup>۲</sup>

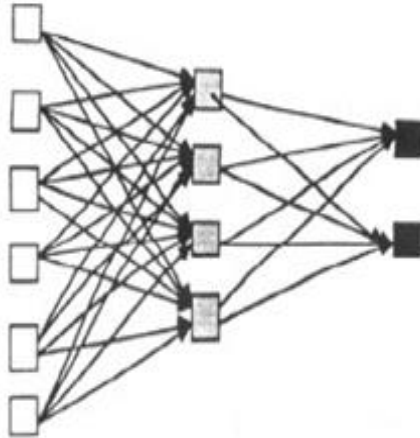
شبکه های پیش خور، شبکه هایی هستند که مسیر پاسخ در آنها همواره رو به جلو پردازش می شود و به نرون های لایه قبل باز نمی گردد. در این نوع شبکه ها به سیگنال اجازه می دهد که در مسیر یک طرفه عبور کند یعنی از ورودی تا خروجی. بنابراین بازخوردی<sup>۳</sup> وجود ندارد بدین معنی که

<sup>۱</sup> . Unsupervised training

<sup>۲</sup> . Feed forward

<sup>۳</sup> . Feed back

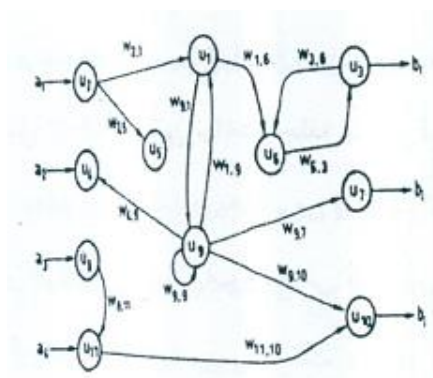
خروجی هر لایه تأثیری بر همان لایه و همچنین لایه‌های قبلی ندارد. شکل (۳-۶) نمونه‌ای از شبکه پیش‌خور را ارائه می‌دهد.



شکل ۳-۶- شبکه عصبی پیش‌خور

### ۳-۶-۲- شبکه‌های برگشتی (پس‌خور)<sup>۱</sup>

در شبکه‌های برگشتی حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های دیگر همان لایه و یا نرون‌های لایه قبل وجود دارد. به عبارتی در این نوع شبکه‌ها بازخورد وجود دارد. شکل (۳-۷) نمونه‌ای از شبکه برگشتی را نمایش می‌دهد [۴۷ و ۴۸].



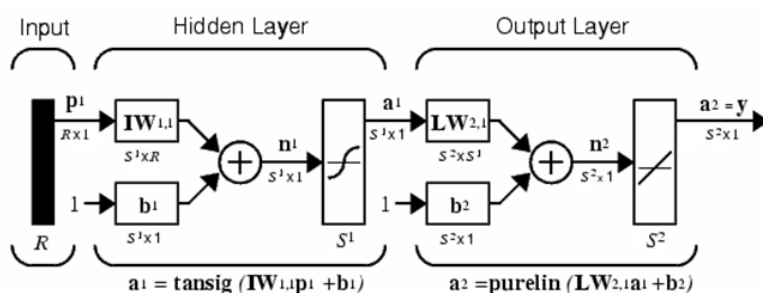
شکل ۳-۷- شبکه عصبی برگشتی

<sup>۱</sup> . Recurrent

### ۳-۷- طراحی شبکه پیش‌خور

در شبکه‌های پیش‌خور، نرون‌ها در چند لایه سازمان‌دهی می‌شوند. اولین و آخرین لایه به ترتیب لایه ورودی<sup>۱</sup> و لایه خروجی<sup>۲</sup> نام دارند، در حالی که لایه‌های بین این دو لایه، لایه‌های پنهان<sup>۳</sup> نامیده می‌شوند. در این نوع شبکه‌ها ابتدا اطلاعات ورودی از محیط خارج به نرون‌های لایه ورودی ارسال می‌شود. در این لایه هیچ پردازشی بر روی ورودی‌ها صورت نمی‌گیرد و در واقع لایه ورودی نقش انتقال‌دهنده را بر عهده دارد. خروجی این نرون‌ها به لایه بعدی که همان لایه پنهان است، منتقل می‌شود. در نهایت خروجی شبکه از طریق نرون‌های لایه خروجی به محیط خارج ارسال می‌گردد.

طبق شکل (۳-۸)، شبکه‌های پیش‌خور اغلب از یک یا چند لایه مخفی از نرون‌های سیگموئیدی و از یک لایه پایانی خطی استفاده می‌کنند. شبکه چند لایه از نرون‌ها با یک تابع انتقال غیرخطی به شبکه اجازه می‌دهد که توانایی یادگیری رابطه خطی و غیرخطی را بین ورودی‌ها و خروجی‌ها داشته باشد. لایه خروجی خطی به شبکه این امکان را می‌دهد که خروجی در هر محدوده دلخواهی از  $-\infty$  تا  $+\infty$  باشد. البته اگر بخواهیم خروجی در دامنه محدودی قرار گیرد، می‌توان از توابع محدودسازی مانند توابع سیگموئیدی در لایه خروجی استفاده کرد [۴۹].



شکل ۳-۸- شبکه پیش‌خور tansig/purelin

- 
- ۱. Input layer
  - ۲. Output layer
  - ۳. Hidden layer

### ۳-۸- آموزش شبکه‌های پیش‌خور

به طور کلی مراحل که در این فرایند دنبال می‌شوند، شامل ایجاد شبکه و تعیین مجموعه داده‌های آموزش، مقدار دهی آغازین به وزن‌ها، آموزش شبکه و شبیه‌سازی<sup>۱</sup> پاسخ شبکه می‌باشد [۴۹].

#### ۳-۸-۱- ایجاد یک شبکه پیش‌خور

اولین مرحله در طراحی یک شبکه پیش‌خور، ایجاد شبکه می‌باشد و تابع `newff` نیز یک شبکه پیش‌خور ایجاد می‌کند. این تابع چهار ورودی دارد و شبکه تهیه شده را به عنوان خروجی برمی‌گرداند. اولین پارامتر ماتریس بردار ورودی، دومین پارامتر ماتریس مقادیر مطلوب خروجی‌ها سومین پارامتر نوع تابع انتقال و آخرین پارامتر نوع تابع آموزش مورد استفاده، می‌باشد.

#### ۳-۸-۲- مقداردهی آغازین به وزن‌ها

قبل از آموزش یک شبکه پیش‌خور، مقادیر ابتدایی وزن‌ها و بایاس‌ها باید تعیین گردند. تابع `newff` به صورت خودکار به وزن‌ها مقدار می‌دهد ولی ممکن است با دستور `init` آن را به صورت غیر خودکار نیز انجام داد.

#### ۳-۸-۳- شبیه‌سازی

تابع `sim` یک شبکه را شبیه‌سازی می‌کند. این تابع شبکه و بردار ورودی را به عنوان ورودی دریافت کرده و خروجی شبکه را بر می‌گرداند.

---

۱ . Simulation

### ۳-۸-۴- آموزش شبکه

پس از مقداردهی به وزن‌ها و بایاس‌ها نوبت به آموزش شبکه می‌رسد. شبکه می‌تواند برای تقریب توابع، تشخیص الگو، یا طبقه بندی الگوها مورد استفاده قرار گیرد. فرایند آموزش به یک سری مثال‌ها از رفتار مورد انتظار شبکه نیاز دارد که شامل ورودی شبکه (P) و هدف (T) می‌باشد. در طول فرایند آموزش وزن‌ها و بایاس‌ها تنظیم می‌شوند تا تابع کارایی شبکه، حداقل شود. تابع کارایی پیش‌فرض برای شبکه‌های پیش‌خور، میانگین مجموع مربعات خطاها (MSE) می‌باشد [۴۶].

### ۳-۹- 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 [۴۷]. برای آموزش شبکه‌های عصبی از الگوریتم‌های یادگیری متفاوتی مانند الگوریتم پس انتشار خطا و الگوریتم لونیبرگ-مارکواریت<sup>۱</sup> و... استفاده می‌شود. انتخاب هر الگوریتم بر سرعت یادگیری و دقت شبکه موثر است.

### ۳-۹-۱- الگوریتم آموزشی پس انتشار خطا

در شبکه‌های عصبی پیش‌خور که با انتخاب معماری مناسب و انتخاب الگوریتم آموزش مناسب مواجه هستیم، از الگوریتم‌های مختلف تکرار که متداول‌ترین آنها الگوریتم پس انتشار خطا است، استفاده می‌شود. دو روش مختلف برای ارائه الگوهای آموزشی و پیاده‌سازی الگوریتم وجود دارد:

---

۱ . Levenberg - Marquardt (LM)

❖ آموزش گام به گام<sup>۱</sup>

❖ آموزش دسته‌ای<sup>۲</sup>

در آموزش گام به گام وزن‌ها و بایاس‌ها بعد از اعمال هر ورودی به روز<sup>۳</sup> می‌شوند در حالی که در روش دسته‌ای پس از اعمال تمام ورودی‌ها (اعضای مجموعه آموزشی) عملیات به روزرسانی وزن‌ها انجام می‌شود. بدین صورت که شیب‌های محاسبه شده برای هر ورودی با هم جمع می‌شوند تا در نهایت وزن‌ها و بایاس‌ها از طریق آن به روز شود. در این تحقیق شیوه آموزش دسته‌ای مورد استفاده قرار گرفته است [۴۹].

به طور کلی آموزش به کمک تکنیک پس‌انتشار بر طبق مراحل زیر انجام می‌شود [۵۰]:

- ۱- انتشار ورودی‌ها از نرون‌های ورودی به سمت نرون‌های خروجی
  - ۲- اختصاص ماتریس وزن‌های تصادفی به هر یک از اتصالات
  - ۳- مقایسه خروجی‌های شبکه با مقادیر واقعی (مقادیر هدف) و محاسبه‌ی خطای شبکه
  - ۴- پس‌انتشار خطا از نرون‌های خروجی به سمت نرون‌های ورودی و اصلاح وزن‌ها
  - ۵- ارزیابی عملکرد شبکه با توجه به تابع کارایی تعیین شده
- مراحل فوق تا زمانی تکرار می‌شود که به حداکثر تکرار<sup>۴</sup> مجاز رسیده باشد یا مقدار تابع کارایی از مقداری که تعیین شده کمتر باشد.

به دلیل فرایندی که در طی این تکنیک رخ می‌دهد، این روش انتشار به عقب نامیده می‌شود زیرا خطای خروجی برای اصلاح وزن‌ها به شبکه برگشت داده می‌شود. و ابتدا وزن‌های ارتباطی بین لایه خروجی و لایه پنهان و سپس وزن‌های ارتباطی بین لایه پنهان و لایه ورودی اصلاح می‌گردند.

---

۱. Incremental training

۲. Batch training

۳. Update

۴. Epoch



متداولترین الگوریتمی که در الگوی آموزشی پس‌انتشار به کار می‌رود، الگوریتم آموزشی نزول گرادینانی<sup>۱</sup> است. در این الگوریتم آموزشی پارامترهای شبکه از قبیل وزن‌ها و بایاس‌ها در جهتی که تابع کارایی حداقل شود، تغییر می‌کند. به روز رسانی وزن‌ها در گام آموزشی (n+1) به صورت زیر است:

$$w_{ji,n+1} = w_{ji,n} + \alpha \Delta w_{ji,n} \quad (5-3)$$

$$\Delta w_{ji,n} = - \frac{\partial E}{\partial w_{ji,n}} \quad (6-3)$$

که در روابط فوق  $w_{ji,n}$  وزن ارتباطی بین نرون  $i$  ام و نرون  $j$  ام در گام آموزشی  $n$  ام،  $\alpha$  سرعت یادگیری و  $E$  تابع خطاست. سرعت یادگیری اندازه گام‌های برداشته شده در اصلاح وزن‌ها را نشان می‌دهد. سرعت یادگیری زیاد باعث نوسان‌های زیاد در حین آموزش شده در حالی که مقدار کم آن باعث آهستگی سرعت همگرایی می‌شود. علامت منفی در رابطه نشان می‌دهد که حرکت به سمتی است که تابع خطا حداقل شود به بیان دیگر هدف از تصحیح وزن‌ها کاهش مقدار خطاست.

### ۳-۹-۲- الگوریتم لونیبرگ- مارکوارت

همان‌طور که گفته شد نزول گرادینانی با وجود متداول بودنش دارای دو ضعف عمده و اساسی است. اول سرعت همگرایی پایین [۵۱] و دوم ساکن شدن و متوقف شدن پارامترهای شبکه به دلیل به دام افتادن آنها در نقاط مینیمم محلی<sup>۲</sup> [۵۲ و ۵۳] قبل از پایان یادگیری کامل شبکه می‌باشد. الگوریتم دیگری که در الگوی آموزشی پس‌انتشار استفاده می‌شود، الگوریتم آموزشی لونیبرگ مارکوارت است که نسبت به روش پس انتشار صدها برابر سریعتر است. این الگوریتم آموزشی که تقریبی از روش نیوتن است، اولین بار توسط لونیبرگ در سال ۱۹۹۴ پیشنهاد شد.

۱. Gradient descent  
۲. Local minimum

در این الگوریتم ماتریس هسین که مشتقات مرتبه دوم خطا نسبت به پارامترهای شبکه است، به صورت زیر تقریب زده می‌شود:

$$H(x) = \nabla^2 F(x) \approx 2J^T(x)J(x) \quad (۷-۳)$$

$$g(x) = \nabla F(x) \approx 2J^T(x)E(W) \quad (۸-۳)$$

که  $E(W)$  بردار خطای شبکه و  $J$  ماتریس ژاکوبین که شامل مشتقات مرتبه اول خطای شبکه نسبت به پارامترهای شبکه است و به فرم زیر نوشته می‌شود:

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_2(x)}{\partial x_n} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\partial e_n(x)}{\partial x_1} & \frac{\partial e_n(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_n(x)}{\partial x_n} \end{bmatrix}_{n \times n} \quad (۹-۳)$$

طبق این الگوریتم، تغییر پارامترها در هر گام آموزشی برابر است با [۵۱ و ۵۴]:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)E(x_k) \quad (۱۰-۳)$$

این الگوریتم حالت بینابین روش نزول گرادیانی و روش گوس-نیوتن است که در آن  $I$  ماتریس یکانی و  $\mu_k$  ضریب یادگیری مارکوارت و عددی غیر منفی می‌باشد. هر قدر  $\mu_k$  بزرگ‌تر شود، نحوه آموزش

به نزول گرادیانی با سرعت یادگیری کوچک ( $\frac{1}{2\mu_k}$ ) نزدیک‌تر می‌شود:

$$x_{k+1} = x_k - \frac{1}{\mu_k} J^T(x_k)E(x_k) = x_k - \frac{1}{2\mu_k} g(x_k) \quad (۱۱-۳)$$

و هر چه  $\mu_k$  کوچک‌تر انتخاب شود، نحوه آموزش به گوس-نیوتن نزدیک‌تر می‌شود:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k)]^{-1} J^T(x_k)E(x_k) \quad (12-3)$$

این ضریب پس از هر گام موفق کاهش می‌یابد و تنها در صورتی افزایش می‌یابد که گام آزمایشی تابع کارایی را افزایش دهد. به این ترتیب تابع کارایی همیشه در حال کاهش است. مراحل‌هایی که در حین آموزش با این الگوریتم دنبال می‌شود، به شرح زیر است:

۱- ارائه ورودی به شبکه و تعیین مقدار اولیه وزن‌ها و  $\mu$

۲- محاسبه خروجی شبکه و خطای مربوط به آن

۳- محاسبه ماتریس ژاکوبین

۴- به روز کردن وزن‌ها بر طبق رابطه (۱۰-۳)

۵- محاسبه مجدد خطای شبکه

اگر خطای محاسبه شده در این حالت کمتر از حالت اول باشد، مقدار  $\mu$  توسط  $\mu_{dec}$  کاهش داده می‌شود و سپس با توجه به مقدار جدید  $\mu$  و مقادیر جدید وزن‌ها به مرحله اول برمی‌گردد و چنانچه خطای افزایش یافته باشد، نتایج مربوط به آن دور نادیده گرفته می‌شود (تغییرات مربوط به وزن‌ها اعمال نمی‌شود) و  $\mu$  در فاکتوری بزرگتر ( $\mu_{inc}$ ) ضرب می‌شود و به مرحله ۴ برمی‌گردد. تکرار مراحل این الگوریتم تا زمانی ادامه دارد که تابع کارایی به کمترین مقدار خود برسد و یا به حداکثر تعداد چرخه‌های آموزش رسیده باشد.

### ۱۰-۳- 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 نموده و از پیدا کردن ارتباط موجود در داده‌ها دور می‌شود. بدین صورت که در هنگام آموزش شبکه، خطا به

۱ . Improving Generalization

۲ . Overtraining

۳ . Overfitting

مقدار قابل قبول می‌رسد ولی هنگام ارزیابی داده‌های جدید، خطای شبکه به مراتب از خطای داده‌های 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. درحین آموزش شبکه، زمانی که خطای بین داده‌های آموزش و ارزیابی می‌خواهد زیاد شود، فرآیند آموزش قطع می‌شود.

روش دیگر برای ارتقای تعمیم شبکه، تنظیم نامیده می‌شود که با اصلاح تابع کارایی از طریق افزایش یک عبارت که شامل میانگین مربعات وزن‌ها و بایاس‌های شبکه است، انجام می‌گیرد.

$$F = \beta E_D + \alpha E_W \quad (13-3)$$

استفاده از این تابع کارایی اصلاح شده سبب می‌شود که شبکه، وزن‌ها و بایاس‌های کوچکتری داشته باشد و این پاسخ شبکه را هموارتر می‌کند. در این رابطه  $\alpha$  و  $\beta$  پارامترهای تابع کارایی،  $E_D$  و  $E_W$  به ترتیب میانگین مربعات خطاها و میانگین مربعات وزن‌ها بوده و طبق روابط (۱۵-۳) و (۱۶-۳) تعریف می‌شوند.

$$E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2 \quad (14-3)$$

---

۱. Early stopping  
۲. Regularization

$$E_w = \frac{1}{N} \sum_{j=1}^N (w_j)^2 \quad (15-3)$$

در الگوریتم آموزشی تنظیم بایزین، اصلاح وزن‌ها و بایاس‌ها همانند الگوریتم لوبنبرگ مارکوارت انجام می‌شود با این تفاوت که تابع خطا، همان تابع کارایی اصلاح شده است یعنی تابع خطا ترکیبی از میانگین مربع خطاها و میانگین مربع وزن‌هاست [۵۵] و هدف از آموزش به روش پس انتشار، کاهش این تابع کارایی اصلاح شده است. مراحل طی شده حین آموزش شبکه با الگوریتم تنظیم بایزین این چنین است:

۱- تعیین مقدار اولیه وزن‌ها،  $\alpha$  و  $\beta$

۲- اجرای الگوریتم لوبنبرگ مارکوارت جهت کاهش تابع کارایی (رابطه (۳-۱۳))

۳- محاسبه تعداد مؤثر پارامترهای تابع کارایی ( $\gamma$ )

$$\gamma = N - 2\alpha \text{tr}(H)^{-1} \quad (16-3)$$

که در این رابطه  $N$  تعداد کل پارامترها و  $\text{tr}(H)^{-1}$  اثر ماتریس معکوس هسیان است. که منظور از اثر ماتریس همان مجموع عناصر قطری ماتریس می‌باشد.

۴- تنظیم مجدد پارامترهای تابع کارایی با استفاده از روابط (۳-۱۷) و (۳-۱۸)

$$\alpha = \frac{\gamma}{2E_w} \quad (17-3)$$

$$\beta = \frac{n - \gamma}{2E_D} \quad (18-3)$$

که در رابطه (۳-۱۸)  $n$  تعداد خطاها می‌باشد.

۵- تکرار مراحل ۲ تا ۴ برای رسیدن به همگرایی

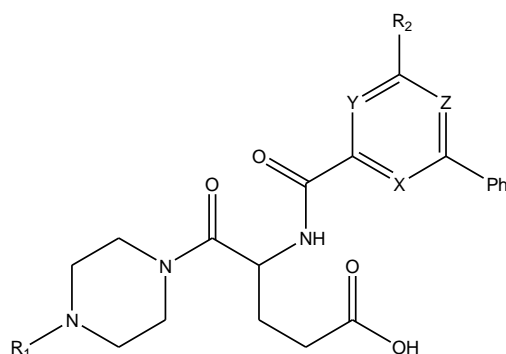
شرایط پایان آموزش و توقف این الگوریتم همانند الگوریتم لوبنبرگ مارکوارت است.

## فصل چهارم

پیش بینی QSAR ثابت بازداري فعاليت  
(pK<sub>i</sub>) مشتقات پایپرازینیل گلوتامات  
پیریدین با استفاده از روش رگرسیون خطی  
چندگانه و شبکه عصبی مصنوعی

#### ۴-۱- سری داده‌ها

در این مرحله از تحقیق، ثابت بازداري فعالیت ۷۷ pK<sub>i</sub> ترکیب از مشتقات پایپرازینیل گلوتامات پیریدین به عنوان سری داده‌ها مورد بررسی قرار گرفت [۲]. ساختار کلی این ترکیبات در شکل (۴-۱) آمده است. همچنین مقادیر عددی pK<sub>i</sub> این ترکیبات برحسب مولار به همراه استخلاف‌های آنان در جدول (۴-۱) آمده است. داده‌های جدول (۴-۱) نشان می‌دهد که نوع استخلاف‌ها و نحوه‌ی آرایش آنها در اطراف اسکلت مولکولی تأثیر زیادی بر روی ثابت بازداري فعالیت دارد. لذا در این تحقیق سعی شده است مدلی ارائه شود که بتواند ثابت بازداري فعالیت هر ترکیب با هر نوع استخلاف در هر موقعیتی را پیش بینی کند بدون اینکه نیاز به سنتز و آزمایش ترکیب مورد نظر باشد.

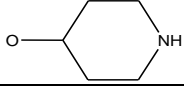
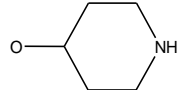
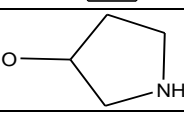
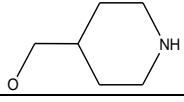
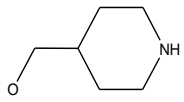
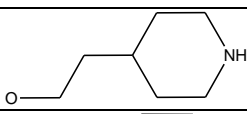
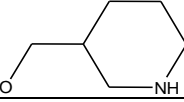
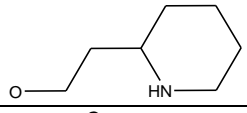
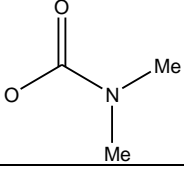
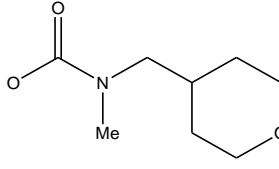


شکل ۴-۱- ساختار کلی ترکیبات پایپرازینیل گلوتامات پیریدین

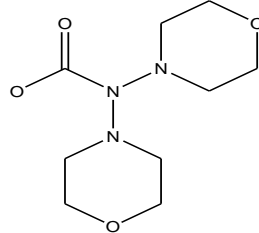
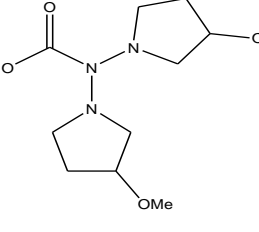
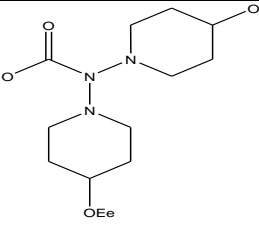
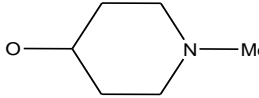
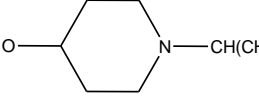
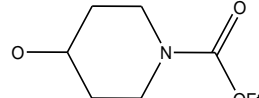
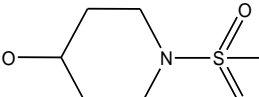
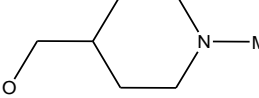
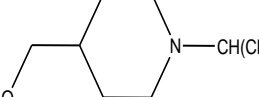
جدول ۴-۱- سری ترکیبات مورد بررسی

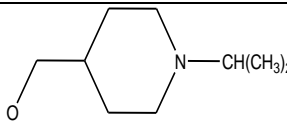
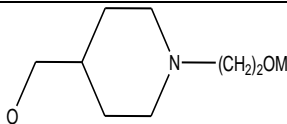
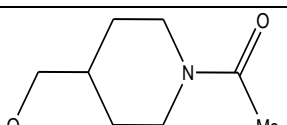
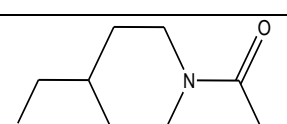
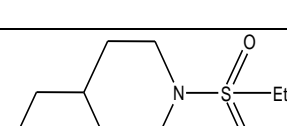
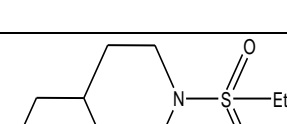
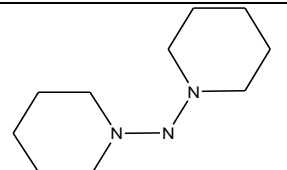
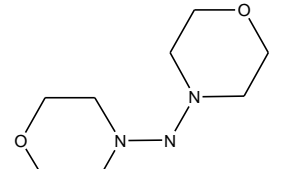
| شماره | R <sub>1</sub>  | X | Y  | Z  | R <sub>2</sub> | pK <sub>i</sub> (M) |
|-------|---|---|----|----|----------------|---------------------|
| ۱*    | EtOCO   | N | N  | CH | H              | ۵/۹۲                |
| ۲     | HeptOCO   | N | CH | CH | H              | ۸/۱۵                |
| ۳     | CH <sub>2</sub> <sup>o</sup> Pent                                 | N | CH | CH | H              | ۷/۳۴                |
| ۴     | (CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub> | N | CH | CH | H              | ۷/۴۸                |
| ۵     | EtOCO   | N | CH | CH | OH             | ۷/۲۱                |
| ۶     | BuOCO   | N | CH | CH | OH             | ۷/۵۱                |
| ۷     | PentOCO   | N | CH | CH | OH             | ۷/۹۶                |
| ۸     | EtOCO   | N | CH | CH | OMe            | ۷/۰۲                |

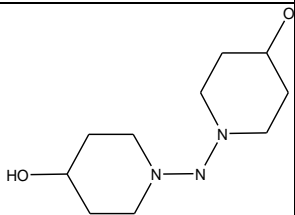
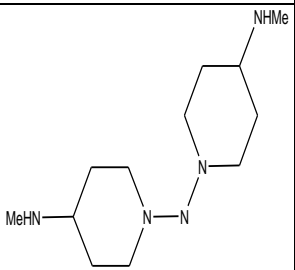
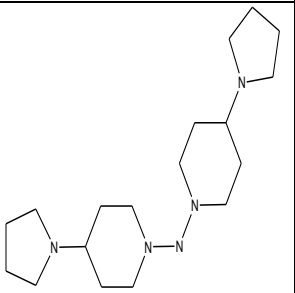
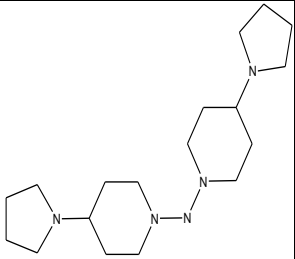
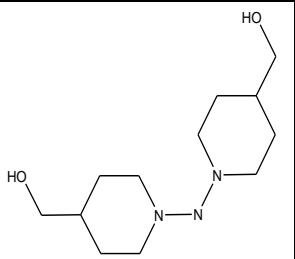
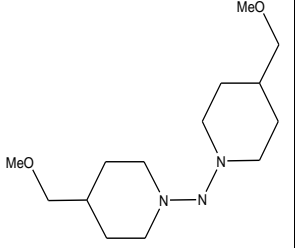
\*این ترکیب از مشتقات پیریدین نمی‌باشد.

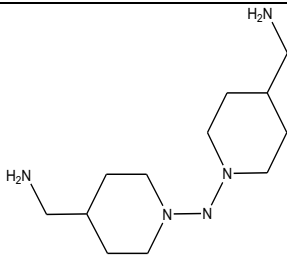
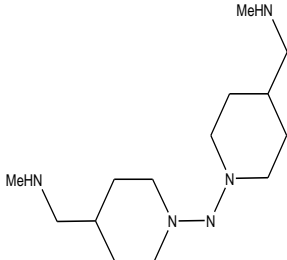
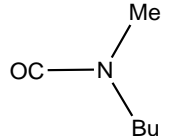
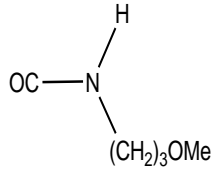
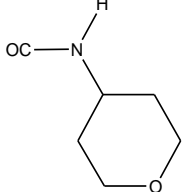
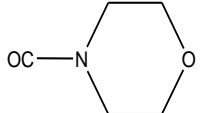
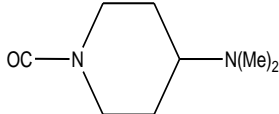
| شماره | R <sub>1</sub> | X | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|----------------|---|----|----|--|---------------------|
| ۹     | BuOCO          | N | CH | CH | OBu  | ۷/۴۶                |
| ۱۰    | EtOCO          | N | CH | CH | O(CH <sub>2</sub> ) <sub>2</sub> OH  | ۶/۸۰                |
| ۱۱    | EtOCO          | N | CH | CH | O(CH <sub>2</sub> ) <sub>3</sub> OH  | ۷/۱۷                |
| ۱۲    | EtOCO          | N | CH | CH | O(CH <sub>2</sub> ) <sub>2</sub> OMe   | ۶/۸۵                |
| ۱۳    | BuOCO          | N | CH | CH | O(CH <sub>2</sub> ) <sub>2</sub> OMe   | ۷/۷۴                |
| ۱۴    | EtOCO          | N | CH | CH | O(CH <sub>2</sub> ) <sub>3</sub> OMe   | ۷/۱۶                |
| ۱۵    | PentOCO        | N | CH | CH |    | ۸/۴۶                |
| ۱۶    | BuOCO          | N | CH | CH |    | ۸/۱۶                |
| ۱۷    | BuOCO          | N | CH | CH |   | ۸/۴۳                |
| ۱۸    | BuOCO          | N | CH | CH |  | ۸/۳۸                |
| ۱۹    | PentOCO        | N | CH | CH |  | ۸/۳۵                |
| ۲۰    | PentOCO        | N | CH | CH |  | ۸/۳۵                |
| ۲۱    | BuOCO          | N | CH | CH |  | ۸/۴۳                |
| ۲۲    | BuOCO          | N | CH | CH |  | ۸/۴۳                |
| ۲۳    | PentOCO        | N | CH | CH |  | ۸/۲۶                |
| ۲۴    | PentOCO        | N | CH | CH |  | ۸/۳۵                |

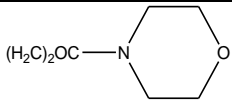
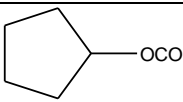
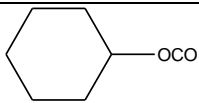
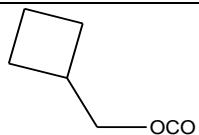
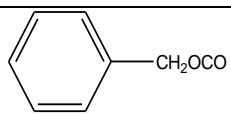
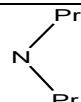
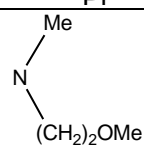


| شماره | R <sub>1</sub> | X | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|----------------|---|----|----|--|---------------------|
| ۲۵    | PentOCO        | N | CH | CH |    | ۸/۳۲                |
| ۲۶    | PentOCO        | N | CH | CH |    | ۸/۳۱                |
| ۲۷    | PentOCO        | N | CH | CH |   | ۸/۲۲                |
| ۲۸    | PentOCO        | N | CH | CH |  | ۸/۳۰                |
| ۲۹    | PentOCO        | N | CH | CH |  | ۸/۱۵                |
| ۳۰    | PentOCO        | N | CH | CH |  | ۸/۱۹                |
| ۳۱    | PentOCO        | N | CH | CH |  | ۸/۲۹                |
| ۳۲    | PentOCO        | N | CH | CH |  | ۸/۲۸                |
| ۳۳    | BuOCO          | N | CH | CH |  | ۸/۳۵                |

| شماره | R <sub>1</sub> | X | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|----------------|---|----|----|--|---------------------|
| ۳۴    | PentOCO        | N | CH | CH |    | ۸/۲۶                |
| ۳۶    | PentOCO        | N | CH | CH |    | ۸/۲۸                |
| ۳۷    | BuOCO          | N | CH | CH |    | ۸/۴۱                |
| ۳۸    | PentOCO        | N | CH | CH |    | ۸/۳۵                |
| ۳۹    | BuOCO          | N | CH | CH |   | ۸/۲۸                |
| ۴۰    | PentOCO        | N | CH | CH |  | ۸/۱۳                |
| ۴۱    | EtOCO          | N | CH | CH | NHMe   | ۷/۴۶                |
| ۴۲    | EtOCO          | N | CH | CH | NHPr   | ۷/۱۹                |
| ۴۳    | PentOCO        | N | CH | CH | NH(CH <sub>2</sub> ) <sub>2</sub> OH   | ۸/۲۴                |
| ۴۴    | EtOCO          | N | CH | CH |  | ۶/۹۵                |
| ۴۵    | EtOCO          | N | CH | CH |  | ۷/۴۱                |

| شماره | R <sub>1</sub> | X | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|----------------|---|----|----|--|---------------------|
| ۴۶    | PentOCO        | N | CH | CH |    | ۷/۹۶                |
| ۴۷    | PentOCO        | N | CH | CH |    | ۸/۱۹                |
| ۴۸    | EtOCO          | N | CH | CH |   | ۷/۴۷                |
| ۴۹    | BuOCO          | N | CH | CH |  | ۸/۱۹                |
| ۵۰    | EtOCO          | N | CH | CH |  | ۷/۲۷                |
| ۵۱    | EtOCO          | N | CH | CH |  | ۷/۷۴                |

| شماره | R <sub>1</sub> | X | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|----------------|---|----|----|--|---------------------|
| ۵۲    | EtOCO          | N | CH | CH |    | ۷/۸۵                |
| ۵۳    | BuOCO          | N | CH | CH |    | ۸/۴۰                |
| ۵۴    | PentOCO        | N | CH | CH | CONH <sub>2</sub>  | ۸/۱۲                |
| ۵۵    | PentOCO        | N | CH | CH |  | ۸/۲۴                |
| ۵۶    | PentOCO        | N | CH | CH |  | ۸/۲۱                |
| ۵۷    | PentOCO        | N | CH | CH |  | ۸/۳۴                |
| ۵۸    | PentOCO        | N | CH | CH |  | ۸/۳۶                |
| ۵۹    | PentOCO        | N | CH | CH |  | ۸/۲۷                |
| ۶۰    | PentOCO        | N | CH | CH | (CH <sub>2</sub> ) <sub>3</sub> OH   | ۸/۴۲                |
| ۶۱    | PentOCO        | N | CH | CH | (CH <sub>2</sub> ) <sub>4</sub> OH   | ۸/۴۴                |

| شماره | R <sub>1</sub>  | X  | Y  | Z  | R <sub>2</sub>   | pK <sub>i</sub> (M) |
|-------|---|----|----|----|--|---------------------|
| ۶۲    | PentOCO   | N  | CH | CH |    | ۸/۴۳                |
| ۶۳*   | EtOCO   | CH | CH | CH | H  | ۵/۴۴                |
| ۶۴*   | EtOCO   | N  | CH | N  | H  | ۶/۴۵                |
| ۶۵    | EtOCO   | N  | CH | CH | H  | ۵/۷۹                |
| ۶۶    | m-tolyl   | N  | CH | CH | Ph   | ۶/۸۲                |
| ۶۷*   | m-tolyl   | CH | CH | CH | Ph   | ۵/۸۴                |
| ۶۸*   | m-tolyl   | CH | CH | CH | H  | ۵/۸۸                |
| ۶۹    | MeOCO   | N  | CH | CH | H  | ۵/۶۱                |
| ۷۰    | OctOCO  | N  | CH | CH | H  | ۷/۸۵                |
| ۷۱    |  | N  | CH | CH | H  | ۷/۳۳                |
| ۷۲    |  | N  | CH | CH | H  | ۷/۳۶                |
| ۷۳    |  | N  | CH | CH | H  | ۷/۵۵                |
| ۷۴    | (CH <sub>3</sub> ) <sub>3</sub> C(CH <sub>2</sub> )OCO                              | N  | CH | CH | H  | ۷/۳۸                |
| ۷۵    |  | N  | CH | CH | H  | ۶/۷۱                |
| ۷۶    | EtOCO   | N  | CH | CH |  | ۷/۳۵                |
| ۷۷    | BuOCO   | N  | CH | CH |  | ۸/۴۰                |

\*این ترکیبات از مشتقات پیریدین نمی باشند

#### ۲-۴- رسم و بهینه‌سازی ساختار مولکول‌ها

اولین مرحله پس از انتخاب سری داده‌ها رسم مولکول‌های مورد بررسی و بهینه‌سازی ساختار آن‌ها می‌باشد. بنابراین ساختار مولکول‌ها به وسیله نرم افزار Hayperchem رسم و ساختار هندسی آنها با احتساب اتم‌های هیدروژن و با استفاده از روش AM1 بهینه شد. بهینه‌سازی تا زمانی که گرادیان به ۰/۰۰۱ کیلو کالری بر مول برسد ادامه یافت.

#### ۳-۴- محاسبه توصیف‌کننده‌های مولکولی

توصیف‌کننده‌ها مقادیر عددی هستند که بیانگر ویژگی خاصی از مولکول می‌باشند که با فعالیت شیمیایی آن ارتباط دارند. ساختارهای بهینه شده به عنوان ورودی به نرم افزار Dragon وارد شده و ۱۴۸۱ توصیف‌کننده از هیجده گروه مختلف برای هر ترکیب توسط این نرم افزار محاسبه گردید.

#### ۴-۴- انتخاب توصیف‌کننده‌های مناسب با استفاده از روش رگرسیون خطی چند گانه

توصیف‌کننده‌های به دست آمده از مرحله قبل به عنوان متغیر مستقل و ثابت بازداري فعالیت  $pK_i$  به عنوان متغیر وابسته وارد نرم افزار SPSS شدند.

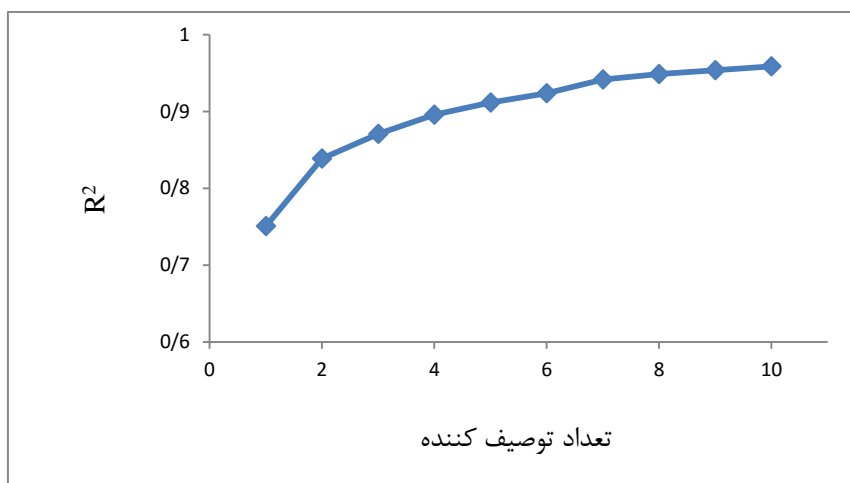
به منظور انتخاب توصیف‌کننده‌های مناسب و جلوگیری از پیچیده شدن محاسبات، توصیف‌کننده‌هایی که با متغیر وابسته همبستگی بیشتری دارند از بین سایر آنها انتخاب شدند. بنابراین توصیف‌کننده‌هایی که دارای بیش از ۹۰٪ مقادیر یکسان بودند حذف شدند. هم چنین از بین توصیف‌کننده‌هایی که دارای هم بستگی بیش از ۰/۹ بودند توصیف‌کننده‌ای که هم بستگی کمتری با متغیر وابسته داشت از بین داده‌ها حذف گردید. سپس با استفاده از فهرست آنالیز، گزینه رگرسیون خطی و انتخاب روش مرحله ای چندین مدل مختلف به دست آمد.

در بین مدل‌های ایجاد شده مدلی که تعداد توصیف‌کننده کمتر،  $R^2$  و آماره  $F$  آن بیشتر باشد، به عنوان مدل برتر انتخاب می‌شود. به منظور ارزیابی مدل انتخاب شده، داده‌ها را به سه دسته‌ی آموزش (شامل ۴۵ ترکیب)، ارزیابی (۱۶ ترکیب) و تست (۱۶ ترکیب) تقسیم می‌شود. سپس ضریب توصیف‌کننده‌های وارد شده در معادله با استفاده از سری آموزش و انتخاب روش Enter بدست می‌آید. توصیف‌کننده‌های انتخاب شده جهت ایجاد مدل MLR به همراه پارامترهای آماری در جدول (۴-۲) آمده است.

جدول ۴-۲- مقایسه آماره‌های مدل‌های به دست آمده از روش MLR

| مدل | توصیف‌کننده‌ها   | $R^2$ | SE    | F       |
|-----|--|-------|-------|---------|
| ۱   | IC3  | ۰/۷۵۱ | ۰/۳۹۱ | ۲۲۷/۴۵۵ |
| ۲   | IC3, Mor11u  | ۰/۸۳۹ | ۰/۳۱۸ | ۱۹۲/۲۸۲ |
| ۳   | IC3, Mor11u, R1u   | ۰/۸۷۱ | ۰/۲۸۶ | ۱۶۴/۱۳۸ |
| ۴   | IC3, Mor11u, R1u, BEHe2  | ۰/۸۹۶ | ۰/۲۵۸ | ۱۵۵/۱۸۹ |
| ۵   | IC3, Mor11u, R1u, BEHe2, R4e                                     | ۰/۹۱۲ | ۰/۲۴۰ | ۱۴۶/۹۰۵ |
| ۶   | IC3, Mor11u, R1u, BEHe2, R4e, MWC08                              | ۰/۹۲۴ | ۰/۲۲۴ | ۱۴۱/۴۷۹ |
| ۷   | IC3, Mor11u, R1u, BEHe2, R4e, MWC08, SRW10                       | ۰/۹۴۲ | ۰/۱۹۶ | ۱۶۱/۴۲۸ |
| ۸   | IC3, Mor11u, R1u, BEHe2, R4e, MWC08, SRW10, HATS0m               | ۰/۹۴۹ | ۰/۱۸۶ | ۱۵۸/۵۵۶ |
| ۹   | IC3, Mor11u, R1u, BEHe2, R4e, MWC08, SRW10, HATS0m, X0AV         | ۰/۹۵۴ | ۰/۱۷۷ | ۱۵۵/۶۲۵ |
| ۱۰  | IC3, Mor11u, R1u, BEHe2, R4e, MWC08, SRW10, HATS0m, X0AV, DF040m | ۰/۹۵۹ | ۰/۱۶۹ | ۱۵۴/۵۴۸ |

در شکل (۴-۲) تأثیر تعداد توصیف‌کننده بر مقدار  $R^2$  نشان داده شده است. همانطور که مشاهده می‌شود از مدل هفتم به بعد تغییر چشمگیری در  $R^2$  مشاهده نمی‌شود و آماره  $F$  نیز کاهش می‌یابد. بنابراین، مدل هفتم به خاطر داشتن تعداد توصیف‌کننده کمتر،  $R^2$  و  $F$  مناسب انتخاب شد.



شکل ۴-۲- تأثیر تعداد توصیف کننده‌ها بر مقادیر R<sup>2</sup> مدل‌های MLR

اثر متوسط یک متغیر مستقل با استفاده از معادله ۴-۱ به دست می‌آید:

$$\text{Mean Effect} = \frac{\beta_x \times \sum x_n}{\sum pK_i} \quad (4-1)$$

$\beta_x$  ضریب متغیر مستقل  $x$  در مدل MLR،  $x_n$  مقدار متغیر مستقل  $x$  مورد نظر برای ترکیب  $n$  ام و  $pK_i$  نیز ثابت بازدارنده فعالیت می‌باشد.

جدول (۴-۳) ضریب رگرسیون، گروه، نام کامل توصیف کننده‌ها و اثر متوسط آنها در مدل منتخب را نشان می‌دهد.

جدول ۴-۳- توصیف کننده‌های وارد شده در مدل منتخب MLR

| شماره | علامت  | ضریب رگرسیون | اثر متوسط | گروه                  | نام کامل  |
|-------|--------|--------------|-----------|-----------------------|---|
| ۱     | IC3    | ۲/۱۰۹        | ۱/۴۶۲     | Topological           | Information content index(neighborhood symmetry of 3-order)   |
| ۲     | Mor11u | -۰/۱۳۲       | ۰/۰۲۶     | 3D-MoRSE              | 3D-MoRSE -signal 11/unweighted  |
| ۳     | R1u    | -۶/۶۰۷       | -۰/۱۹۰    | GETAWAY               | R maximal autocorrelation of lag /unweighted highest eigenvalue n.2 of Burden matrix/weighted by atomic sanderson electronegativities |
| ۴     | BEHe2  | -۷/۴۲۲       | -۳/۶۳     | BCUT                  | R autocorrelation of lag 4/weighted by atomic sanderson electronegativities   |
| ۵     | R4e    | -۵/۳۲۵       | -۰/۰۷۷    | GETAWAY               | R autocorrelation of lag 4/weighted by atomic sanderson electronegativities   |
| ۶     | MWC08  | -۲/۲۵۱       | ۰/۳۷۵     | Molecular walk counts | molecular walk count of order 08  |
| ۷     | SRW10  | ۷/۵۴۴E-۵     | ۰/۳۴۷     | Molecular walk counts | Self-returning walk count of order 10   |



جدول (۴-۴) ماتریس همبستگی توصیف کننده‌های وارد شده در مدل منتخب را نشان می‌دهد. همانطور که ملاحظه می‌شود مقدار همبستگی بین پارامترها قابل قبول (کمتر از ۰/۹) می‌باشد. بنابراین این ماتریس عدم همبستگی بین توصیف کننده‌ها را نشان می‌دهد.

جدول ۴-۴- ماتریس همبستگی توصیف کننده‌های انتخاب شده در مدل نهایی MLR

| توصیف کننده | IC3    | Mor11u | R1u    | BEHe2  | R4e    | MWC08 | SRW10 |
|-------------|--------|--------|--------|--------|--------|-------|-------|
| IC3         | ۱      |        |        |        |        |       |       |
| Mor11u      | -۰/۵۴۲ | ۱      |        |        |        |       |       |
| R1u         | -۰/۲۶۵ | ۰/۴۳۷  | ۱      |        |        |       |       |
| BEHe2       | -۰/۱۷۰ | ۰/۱۱۲  | ۰/۳۲۸  | ۱      |        |       |       |
| R4e         | -۰/۳۹۸ | ۰/۵۸۴  | ۰/۵۴۴  | -۰/۱۴۶ | ۱      |       |       |
| MWC08       | ۰/۳۹۱  | -۰/۳۹۷ | -۰/۵۱۲ | ۰/۳۶۱  | -۰/۵۳۶ | ۱     |       |
| SRW10       | ۰/۴۳۲  | -۰/۴۲۲ | -۰/۵۰۴ | ۰/۳۵۶  | -۰/۵۲۸ | ۰/۱۸۹ | ۱     |

به این ترتیب معادله خطی به صورت زیر به دست آمد:

$$pK_i = 29/926 + 2/109 IC3 - 0/132 Mor11u - 6/607 R1u - 7/422 BEHe2 - 5/325 R4e - 3/251 MWC08 + 7/544 E-5 SRW10 \quad (\text{معادله ۴-۲})$$

#### ۵-۴- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN)

از آنجایی که ممکن است مدل خطی بهترین مدل برای توصیف رفتار سری داده‌ها نباشد، شبکه عصبی مصنوعی برای مدل سازی غیرخطی داده‌ها مورد استفاده قرار گرفت. شبکه عصبی انتخاب شده در این تحقیق یک شبکه پیش خور با الگوریتم آموزشی پس انتشار می‌باشد. مقادیر عددی توصیف کننده‌های وارد شده در مدل خطی به عنوان ورودی و مقادیر تجربی ثابت بازداري فعالیت ( $pK_i$ ) نیز به عنوان هدف به شبکه عصبی داده شدند. همانند روش MLR، داده‌ها به صورت تصادفی به سه سری آموزش (شامل ۴۵ ترکیب)، سری ارزیابی (۱۶ ترکیب) و سری تست (۱۶ ترکیب) تقسیم شدند. سری آموزش برای آموزش شبکه و سری ارزیابی برای بررسی قدرت تعمیم پذیری مدل به دست آمده از سری آموزش به کار گرفته شد. سری تست نیز به عنوان داده‌های خارجی برای ارزیابی نهایی شبکه

در نظر گرفته شد. برای آموزش موفق و در نتیجه قدرت پیش‌بینی شبکه باید عوامل موثر در آموزش شبکه عصبی مصنوعی بررسی و بهینه شود. این پارامترها شامل تعداد متغیرهای ورودی شبکه (تعداد توصیف‌کننده‌ها)، نوع تابع آموزش، نوع تابع انتقال، تعداد گره‌ها در لایه پنهان، پارامتر  $\mu$  (مقدار اولیه برای  $\mu$ ) و تعداد دوره‌های آموزش می‌باشد. در این روش تابع کارایی شبکه، میانگین مربع خطای استاندارد سری ارزیابی می‌باشد.

#### ۴-۵-۱- بهینه‌سازی تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه و تعداد

##### نرون‌های لایه مخفی

هر شبکه سه لایه شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است که تعداد ورودی‌ها (توصیف‌کننده‌ها) با تعداد نرون‌های لایه ورودی برابر است و لایه خروجی دارای یک نرون است که نشان دهنده‌ی  $pK_i$  متناظر با هر ترکیب می‌باشد. اما هیچ راهنمای مناسبی برای انتخاب تعداد لایه‌های پنهان وجود ندارد و ساختار شبکه‌های عصبی، اغلب به روش سعی و خطا ایجاد می‌شود. با این وجود در بیشتر موارد به نظر می‌رسد که تنها یک لایه پنهان مناسب باشد [۵۰]. بنابراین در این تحقیق نیز یک لایه پنهان مورد استفاده قرار گرفت، ولی تعداد نرون‌های لایه پنهان باید بهینه شود.

برای بهینه‌سازی و شناسایی پارامترهای موثر شبکه، باید شبکه به طور مناسب آموزش دیده باشد. بر این اساس شبکه با ورودی‌های از ۲ تا ۱۰ توصیف‌کننده توسط دو الگوریتم آموزشی لونیگ-مارکوارت و تنظیم بایزین با تعداد گره‌های از ۲ تا ۱۰ در لایه پنهان و همچنین توابع لگاریتم سیگموئیدی ( $\text{logsig}$ )، تانژانت سیگموئیدی ( $\text{tansig}$ ) و تابع انتقال خطی ( $\text{purelin}$ )، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمام شبکه‌ها تعداد دوره‌های آموزشی یکسان و برابر ۲۰ در نظر گرفته شد. همچنین از تابع انتقال خطی ( $\text{purelin}$ ) در لایه خروجی استفاده شد.

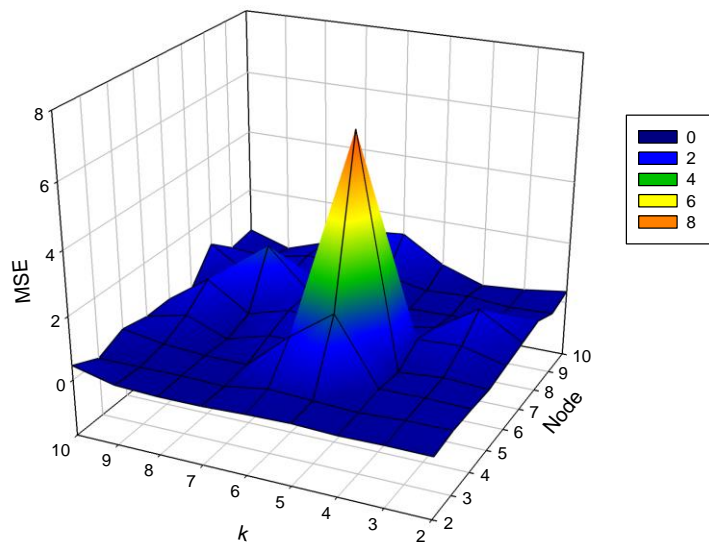
در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطای سری ارزیابی به عنوان معیار انتخاب گردید. نتایج این محاسبات در جداول (۴-۵)، (۴-۶)، (۴-۷)، (۴-۸)، (۴-۹)، (۴-۱۰) آورده شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۴-۳)، (۴-۴)، (۴-۵)، (۴-۶)، (۴-۷)، (۴-۸) ارائه شده است.

با توجه به نتایج به دست آمده، الگوریتم آموزشی لونیبرگ مارکوارت نسبت به الگوریتم آموزشی بایزین دارای میانگین مربع خطاهای کمتری می‌باشد. این الگوریتم علاوه بر سرعت آموزش بسیار زیاد دارای قدرت تعمیم‌پذیری بالایی نیز می‌باشد، بدین معنی که قادر است میان ترکیباتی که در آموزش به کار گرفته شده و آنهایی که در آموزش به کار گرفته نشده، تطابق خوبی برقرار نماید.

همچنین تابع تانژانت سیگموئیدی دارای خطای کمتری نسبت به لگاریتم سیگموئیدی است، اما به دلیل اینکه افزایش تعداد گره‌ها باعث پیچیده شدن شبکه می‌گردد، بنابراین تابع انتقال لگاریتم سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. بنابراین شبکه با تابع آموزش لونیبرگ مارکوارت و تابع انتقال لگاریتم سیگموئید، ۷ متغیر ورودی و تعداد ۳ نرون در لایه پنهان، انتخاب می‌شود. مقدار  $\mu$  نیز در این شرایط برابر با مقدار پیش فرض آن در تابع آموزشی لونیبرگ مارکوارت یعنی ۰/۰۰۱ است.

جدول ۴-۵- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع انتقال لگاریتم سیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

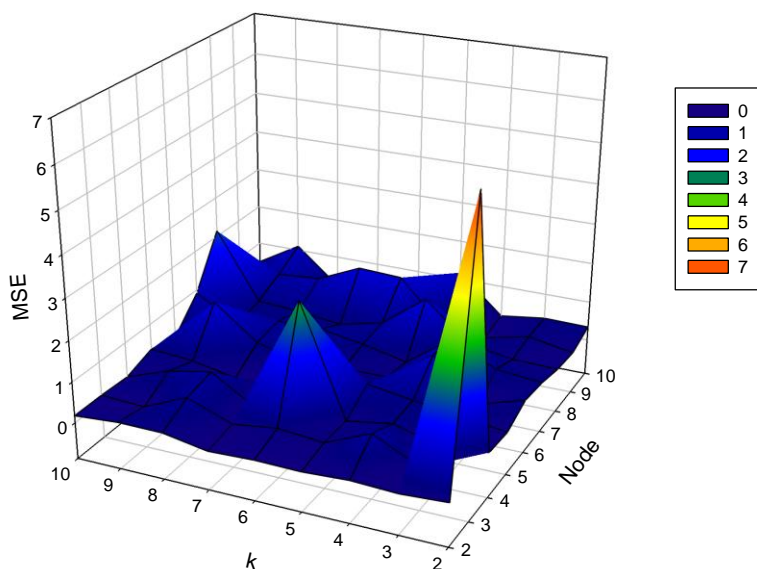
|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |        |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|--------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹      | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۳۰۷   | ۰/۱۳۹۶ | ۰/۱۱۴۷ | ۰/۱۶۷۲ | ۰/۱۳۳۹ | ۰/۰۷۶۷ | ۰/۰۷۱۸ | ۰/۱۲۹۲ | ۰/۴۵۸۲ |
|              | ۳  | ۰/۲۲۲۳   | ۰/۲۱۰۶ | ۰/۱۳۸۶ | ۰/۱۹۹۲ | ۰/۲۶۵۹ | ۰/۰۴۷۸ | ۰/۰۸۴۰ | ۰/۰۷۳۲ | ۰/۰۷۱۹ |
|              | ۴  | ۰/۲۰۶۷   | ۰/۲۴۱۰ | ۰/۲۲۲۰ | ۲/۲۶۹۸ | ۱/۲۸۹۸ | ۰/۱۱۲۳ | ۰/۱۱۳۲ | ۰/۲۱۰۹ | ۰/۴۲۷۷ |
|              | ۵  | ۰/۱۵۷۹   | ۰/۲۳۶۴ | ۰/۱۱۵۰ | ۷/۱۹۸۴ | ۰/۴۸۱۶ | ۰/۴۳۸۷ | ۰/۰۴۹۳ | ۰/۵۶۷۴ | ۰/۲۵۲۸ |
|              | ۶  | ۰/۲۸۴۲   | ۰/۶۲۵۲ | ۰/۵۶۰۱ | ۰/۱۷۳۵ | ۰/۵۲۱۷ | ۰/۶۸۸۳ | ۰/۳۰۵۵ | ۱/۱۸۷۸ | ۰/۳۷۰۳ |
|              | ۷  | ۰/۴۰۴۳   | ۱/۲۱۵۹ | ۰/۳۴۲۱ | ۰/۱۳۴۳ | ۰/۶۰۱۳ | ۰/۷۳۳۰ | ۲/۰۱۱۶ | ۰/۲۰۴۰ | ۰/۰۵۵۱ |
|              | ۸  | ۰/۵۲۴۴   | ۰/۱۶۷۷ | ۰/۱۵۹۵ | ۰/۱۱۹۷ | ۰/۱۲۵۰ | ۰/۳۷۲۱ | ۰/۴۸۸۸ | ۰/۲۳۶۲ | ۱/۰۷۸۶ |
|              | ۹  | ۰/۱۹۱۰   | ۰/۱۳۶۰ | ۰/۲۵۴۷ | ۰/۱۸۰۳ | ۰/۴۳۵۲ | ۰/۵۱۸۵ | ۰/۶۳۹۸ | ۰/۳۳۱۷ | ۰/۴۲۸۴ |
|              | ۱۰ | ۰/۳۶۸۶   | ۰/۱۷۴۰ | ۰/۰۶۳۸ | ۰/۵۲۶۳ | ۱/۳۵۰۲ | ۱/۲۲۷۶ | ۰/۵۷۵۷ | ۰/۱۲۹۰ | ۰/۵۵۰۸ |



شکل ۴-۳- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول ۴-۶- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع انتقال تانژانت سیگموئید (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

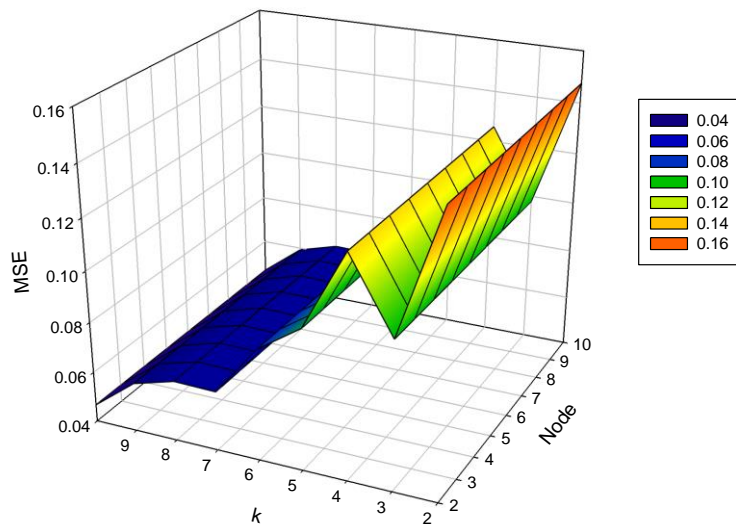
|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |        |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|--------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹      | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۵۶۴   | ۰/۰۹۰۲ | ۰/۱۳۸۸ | ۰/۰۷۸۴ | ۰/۰۹۷۳ | ۰/۰۵۲۸ | ۰/۲۴۷۱ | ۰/۲۵۶۵ | ۰/۱۷۴۲ |
|              | ۳  | ۶/۴۹۲۲   | ۰/۱۴۰۳ | ۰/۶۹۶۳ | ۰/۲۲۴۱ | ۰/۳۲۰۷ | ۰/۱۷۴۷ | ۰/۵۸۷۴ | ۰/۱۱۶۷ | ۰/۱۶۹۹ |
|              | ۴  | ۰/۲۰۴۵   | ۰/۰۸۳۵ | ۰/۴۰۱۱ | ۰/۲۱۶۸ | ۲/۸۸۹۰ | ۰/۰۴۷۴ | ۰/۷۱۶۸ | ۰/۵۰۹۹ | ۰/۱۳۳۶ |
|              | ۵  | ۰/۱۵۲۲   | ۰/۳۰۱۸ | ۰/۴۵۱۱ | ۰/۶۱۳۳ | ۰/۲۸۸۰ | ۰/۱۰۶۷ | ۰/۰۶۸۱ | ۰/۲۷۸۸ | ۰/۳۶۲۰ |
|              | ۶  | ۰/۴۲۰۱   | ۰/۰۷۱۳ | ۱/۰۷۲۲ | ۰/۲۹۷۰ | ۰/۲۷۸۵ | ۰/۲۳۵۴ | ۰/۴۹۵۹ | ۱/۳۲۸۱ | ۰/۲۰۴۱ |
|              | ۷  | ۰/۳۴۸۷   | ۰/۴۵۵۴ | ۰/۹۹۳۰ | ۰/۱۸۵۵ | ۰/۴۲۳۶ | ۰/۶۰۳۱ | ۰/۵۹۷۰ | ۰/۲۹۸۷ | ۰/۹۱۱۲ |
|              | ۸  | ۰/۱۹۵۵   | ۰/۲۲۴۰ | ۰/۴۰۳۹ | ۱/۳۴۶۰ | ۰/۵۱۴۶ | ۰/۲۶۲۳ | ۰/۳۹۳۷ | ۰/۴۶۸۲ | ۲/۱۴۳۴ |
|              | ۹  | ۰/۱۳۹۱   | ۰/۳۳۲۶ | ۰/۰۸۸۶ | ۰/۱۲۱۱ | ۱/۲۶۸۳ | ۱/۳۱۹۶ | ۰/۸۳۷۹ | ۰/۷۵۶۷ | ۰/۲۱۰۶ |
|              | ۱۰ | ۰/۳۴۵۲   | ۰/۳۶۵۷ | ۰/۲۱۵۸ | ۱/۱۷۰۶ | ۰/۳۴۲۹ | ۰/۳۹۰۵ | ۰/۲۰۹۱ | ۱/۱۰۶۲ | ۰/۴۸۹۰ |



شکل ۴-۴- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیبرگ- مارکوارت و تابع انتقال تانژانت سیگموئیدی

جدول ۴-۷- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت (trainlm) و تابع انتقال خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

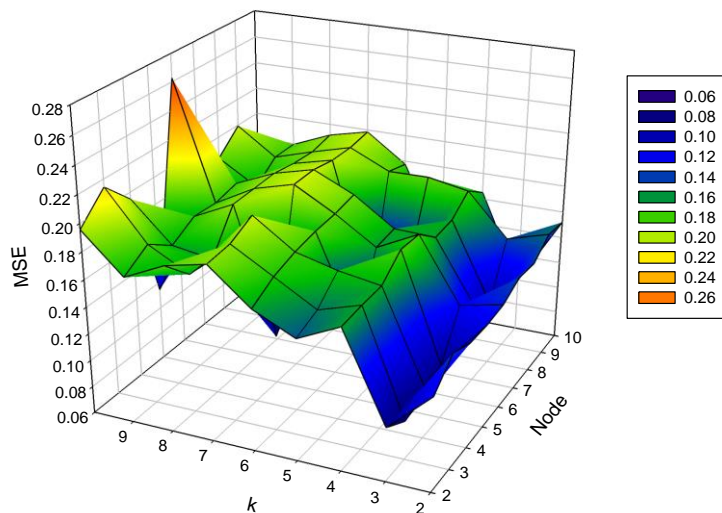
|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |        |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|--------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹      | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۳  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۴  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۵  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۶  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۷  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۸  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۹  | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |
|              | ۱۰ | ۰/۱۴۷۶   | ۰/۰۹۷۹ | ۰/۱۲۵۹ | ۰/۰۹۴۹ | ۰/۰۸۳۷ | ۰/۰۶۳۴ | ۰/۰۶۳۸ | ۰/۰۵۹۲ | ۰/۰۴۶۷ |



شکل ۴-۵- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیبرگ-مارکوارت و تابع انتقال خطی

جدول ۴-۸- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بایزین (trainbr) و تابع انتقال لگاریتم سیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

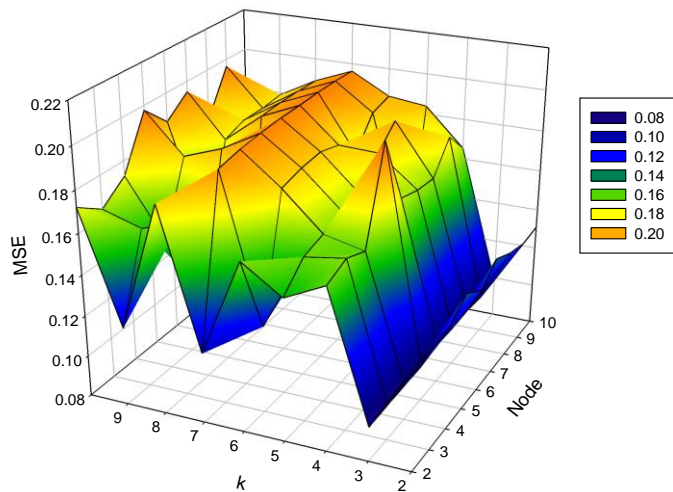
|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |        |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|--------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹      | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۲۸۸   | ۰/۰۹۹۶ | ۰/۱۶۳۶ | ۰/۱۴۹۰ | ۰/۱۶۳۴ | ۰/۱۸۸۴ | ۰/۱۸۱۷ | ۰/۱۶۷۷ | ۰/۱۹۵۷ |
|              | ۳  | ۰/۱۴۴۵   | ۰/۰۸۸۵ | ۰/۱۶۵۲ | ۰/۱۵۵۶ | ۰/۱۷۳۱ | ۰/۱۹۰۴ | ۰/۱۶۳۱ | ۰/۱۷۹۰ | ۰/۲۱۴۵ |
|              | ۴  | ۰/۱۳۷۶   | ۰/۰۹۹۸ | ۰/۱۶۵۸ | ۰/۱۷۰۷ | ۰/۱۷۸۴ | ۰/۱۹۹۰ | ۰/۱۶۱۹ | ۰/۱۶۶۳ | ۰/۱۹۵۰ |
|              | ۵  | ۰/۱۳۶۸   | ۰/۱۰۰۱ | ۰/۱۷۹۴ | ۰/۱۱۴۱ | ۰/۱۸۱۴ | ۰/۰۹۶۱ | ۰/۱۹۰۴ | ۰/۱۷۶۲ | ۰/۱۱۴۴ |
|              | ۶  | ۰/۱۳۸۳   | ۰/۱۳۱۵ | ۰/۱۷۸۴ | ۰/۱۶۳۲ | ۰/۱۸۲۷ | ۰/۱۹۹۲ | ۰/۱۹۱۱ | ۰/۱۷۵۲ | ۰/۲۶۲۷ |
|              | ۷  | ۰/۱۴۵۵   | ۰/۱۱۷۱ | ۰/۱۶۴۴ | ۰/۱۶۵۵ | ۰/۱۸۵۸ | ۰/۱۹۹۶ | ۰/۱۹۲۰ | ۰/۱۷۹۳ | ۰/۱۶۱۱ |
|              | ۸  | ۰/۱۴۳۷   | ۰/۱۰۷۱ | ۰/۱۶۶۰ | ۰/۱۲۹۵ | ۰/۰۹۴۸ | ۰/۱۹۶۴ | ۰/۱۸۹۱ | ۰/۱۶۶۰ | ۰/۱۸۳۷ |
|              | ۹  | ۰/۱۵۴۲   | ۰/۱۴۵۲ | ۰/۱۷۷۰ | ۰/۱۸۲۷ | ۰/۱۸۲۳ | ۰/۱۹۶۰ | ۰/۱۸۸۰ | ۰/۱۷۹۱ | ۰/۱۹۹۳ |
|              | ۱۰ | ۰/۱۵۲۰   | ۰/۰۹۹۹ | ۰/۱۶۵۱ | ۰/۱۱۶۳ | ۰/۱۷۷۵ | ۰/۱۹۸۴ | ۰/۱۹۱۴ | ۰/۱۷۹۳ | ۰/۱۷۲۴ |



شکل ۴-۶- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول ۴-۹- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بایزین (trainbr) و تابع انتقال تانژانت سیگموئید (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |       |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|-------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹     | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۲۹۹   | ۰/۰۹۶۷ | ۰/۱۵۸۲ | ۰/۱۴۸۷ | ۰/۱۶۱۶ | ۰/۱۱۴۲ | ۰/۱۸۰۴ | /۱۱۸۱ | ۰/۱۷۲۰ |
|              | ۳  | ۰/۱۳۰۲   | ۰/۰۹۷۱ | ۰/۱۶۲۴ | ۰/۱۵۹۳ | ۰/۱۲۲۵ | ۰/۱۹۲۳ | ۰/۱۶۴۲ | /۱۵۸۳ | ۰/۱۶۴۰ |
|              | ۴  | ۰/۱۲۹۹   | ۰/۰۹۷۲ | ۰/۱۵۷۸ | ۰/۱۶۲۲ | ۰/۱۷۹۹ | ۰/۱۹۴۵ | ۰/۱۲۲۹ | /۱۴۵۱ | ۰/۱۷۲۸ |
|              | ۵  | ۰/۱۳۷۴   | ۰/۰۹۷۲ | ۰/۲۰۱۱ | ۰/۱۶۵۲ | ۰/۱۸۲۴ | ۰/۱۹۵۶ | ۰/۱۸۵۴ | /۱۷۸۷ | ۰/۱۹۷۷ |
|              | ۶  | ۰/۱۳۰۳   | ۰/۰۹۹۰ | ۰/۱۶۷۸ | ۰/۱۶۶۲ | ۰/۱۸۳۴ | ۰/۱۹۶۴ | ۰/۱۰۹۲ | /۱۲۹۷ | ۰/۱۸۶۲ |
|              | ۷  | ۰/۱۳۷۹   | ۰/۰۹۸۶ | ۰/۱۶۴۶ | ۰/۱۷۹۴ | ۰/۱۸۰۲ | ۰/۱۹۶۵ | ۰/۱۸۸۵ | /۱۷۴۰ | ۰/۱۹۵۰ |
|              | ۸  | ۰/۱۳۲۹   | ۰/۰۹۹۰ | ۰/۱۶۳۴ | ۰/۱۸۸۶ | ۰/۱۶۹۷ | ۰/۱۹۶۹ | ۰/۱۸۹۲ | /۱۷۷۷ | ۰/۱۷۵۱ |
|              | ۹  | ۰/۱۲۹۹   | ۰/۰۹۷۹ | ۰/۱۷۷۳ | ۰/۱۱۵۵ | ۰/۱۸۳۷ | ۰/۱۹۶۸ | ۰/۱۸۹۱ | /۱۷۷۷ | ۰/۱۹۶۵ |
|              | ۱۰ | ۰/۱۳۰۹   | ۰/۰۹۸۶ | ۰/۱۶۵۴ | ۰/۱۸۳۹ | ۰/۱۸۶۲ | ۰/۱۹۶۲ | ۰/۱۸۸۸ | /۱۷۸۹ | ۰/۱۴۷۷ |

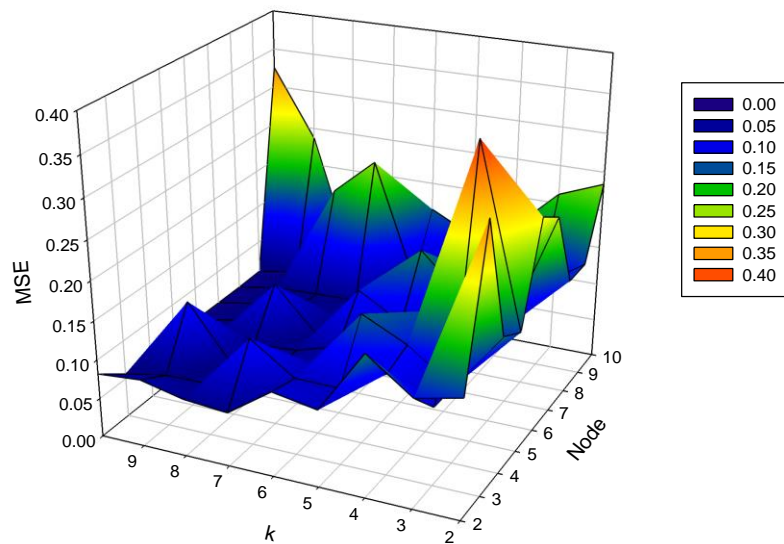


شکل ۴-۷- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی



جدول ۴-۱۰- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بایزین (trainbr) و تابع انتقال خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

|              |    | تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |        |        |        |        |        |        |        |        |
|--------------|----|--|--------|--------|--------|--------|--------|--------|--------|--------|
|              |    | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      | ۹      | ۱۰     |
| تعداد گره‌ها | ۲  | ۰/۱۵۰۴   | ۰/۱۳۷۶ | ۰/۱۷۹۱ | ۰/۰۹۸۱ | ۰/۱۰۸۱ | ۰/۰۶۹۱ | ۰/۰۷۳۸ | ۰/۰۸۶۶ | ۰/۰۸۲۱ |
|              | ۳  | ۰/۳۳۱۸   | ۰/۰۹۹۷ | ۰/۱۲۸۵ | ۰/۰۹۸۳ | ۰/۰۹۹۰ | ۰/۱۳۷۷ | ۰/۰۶۷۶ | ۰/۰۶۶۵ | ۰/۰۵۶۳ |
|              | ۴  | ۰/۱۷۲۵   | ۰/۰۹۸۷ | ۰/۱۳۹۴ | ۰/۱۶۶۹ | ۰/۰۸۹۸ | ۰/۰۶۷۴ | ۰/۰۶۷۶ | ۰/۱۳۸۸ | ۰/۰۵۶۵ |
|              | ۵  | ۰/۱۵۳۴   | ۰/۳۷۳۴ | ۰/۱۵۷۸ | ۰/۱۰۰۰ | ۰/۰۸۹۷ | ۰/۰۷۲۵ | ۰/۰۶۹۷ | ۰/۰۶۶۴ | ۰/۰۵۶۳ |
|              | ۶  | ۰/۲۷۱۹   | ۰/۱۳۰۳ | ۰/۱۳۲۵ | ۰/۱۰۷۷ | ۰/۱۳۹۸ | ۰/۰۶۹۲ | ۰/۱۲۶۷ | ۰/۰۶۷۰ | ۰/۰۵۶۹ |
|              | ۷  | ۰/۲۵۱۶   | ۰/۱۰۰۳ | ۰/۱۲۸۰ | ۰/۱۸۰۴ | ۰/۱۰۵۸ | ۰/۰۷۰۰ | ۰/۰۶۸۴ | ۰/۰۶۶۶ | ۰/۰۵۶۷ |
|              | ۸  | ۰/۱۴۹۱   | ۰/۱۱۴۶ | ۰/۱۳۶۲ | ۰/۰۹۹۳ | ۰/۱۰۵۸ | ۰/۰۶۹۴ | ۰/۰۶۷۷ | ۰/۰۶۶۰ | ۰/۰۵۶۵ |
|              | ۹  | ۰/۱۴۸۹   | ۰/۱۱۰۱ | ۰/۱۳۵۹ | ۰/۱۳۷۷ | ۰/۰۸۶۱ | ۰/۲۳۸۵ | ۰/۱۹۱۰ | ۰/۰۶۶۷ | ۰/۰۵۷۴ |
|              | ۱۰ | ۰/۲۳۴۳   | ۰/۲۱۲۲ | ۰/۱۵۸۹ | ۰/۱۳۶۴ | ۰/۱۶۴۲ | ۰/۰۶۸۳ | ۰/۰۶۸۶ | ۰/۲۳۷۰ | ۰/۳۲۵۰ |



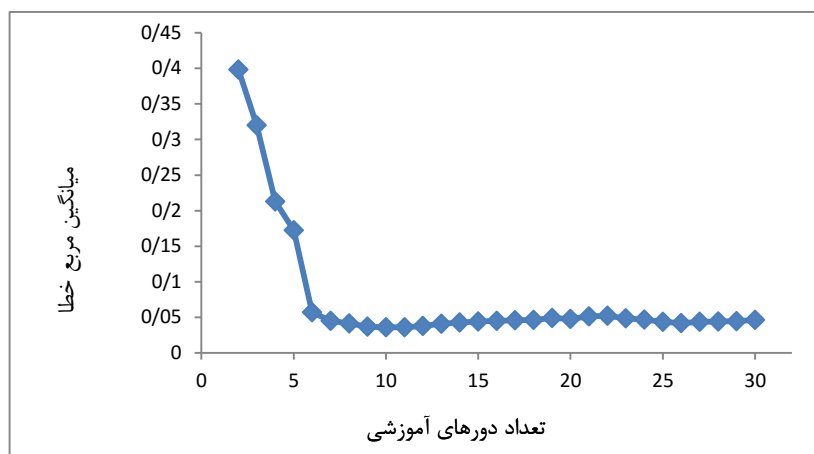
شکل ۴-۸- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال خطی

#### ۴-۵-۲- بهینه سازی تعداد دوره‌های آموزشی

تعداد دوره‌های آموزشی پارامتری است که برای جلوگیری از آموزش بیش از حد شبکه بهینه‌سازی می‌شود. برای این منظور، شبکه با الگوریتم آموزشی لوببرگ مارکوات، ۷ توصیف کننده، تابع انتقال لگاریتم سیگموئیدی و مقدار بهینه تعداد نرون‌ها (۳) در تعداد دوره‌های متفاوت آموزش داده شد و میانگین مربع خطاها برای سری ارزیابی به دست آمد. جدول (۴-۱۱) و شکل (۴-۹) نشان می‌دهد که با افزایش تعداد دوره‌های آموزش میانگین مربع خطا (MSE) کاهش می‌یابد تا جایی که در ۱۰ دور تکرار به حداقل خطا می‌رسد و سپس با افزایش بیشتر تعداد دوره‌های یادگیری، خطاها اندکی افزایش می‌یابد بنابراین ۱۰ بار تکرار به عنوان تعداد دوره‌های بهینه برای یادگیری شبکه انتخاب شد.

جدول ۴-۱۱- مقادیر میانگین مربع خطای سری ارزیابی با تغییر تعداد دوره‌های آموزشی

| تعداد دوره‌های آموزشی | میانگین مربع خطا | تعداد دوره‌های آموزشی | میانگین مربع خطا |
|-----------------------|------------------|-----------------------|------------------|
| ۲                     | ۰/۳۹۸۵           | ۱۷                    | ۰/۰۴۵۹           |
| ۳                     | ۰/۳۲۰۰           | ۱۸                    | ۰/۰۴۶۴           |
| ۴                     | ۰/۲۱۳۰           | ۱۹                    | ۰/۰۴۹۰           |
| ۵                     | ۰/۱۷۲۲           | ۲۰                    | ۰/۰۴۷۸           |
| ۶                     | ۰/۰۵۷۳           | ۲۱                    | ۰/۰۵۱۶           |
| ۷                     | ۰/۰۴۴۹           | ۲۲                    | ۰/۰۵۲۱           |
| ۸                     | ۰/۰۴۱۳           | ۲۳                    | ۰/۰۴۸۶           |
| ۹                     | ۰/۰۳۶۸           | ۲۴                    | ۰/۰۴۷۱           |
| ۱۰                    | ۰/۰۳۵۹           | ۲۵                    | ۰/۰۴۳۹           |
| ۱۱                    | ۰/۰۳۶۱           | ۲۶                    | ۰/۰۴۲۰           |
| ۱۲                    | ۰/۰۳۸۰           | ۲۷                    | ۰/۰۴۳۹           |
| ۱۳                    | ۰/۰۴۱۱           | ۲۸                    | ۰/۰۴۴۱           |
| ۱۴                    | ۰/۰۴۳۰           | ۲۹                    | ۰/۰۴۴۸           |
| ۱۵                    | ۰/۰۴۴۰           | ۳۰                    | ۰/۰۴۶۴           |
| ۱۶                    | ۰/۰۴۵۲           |                       |                  |



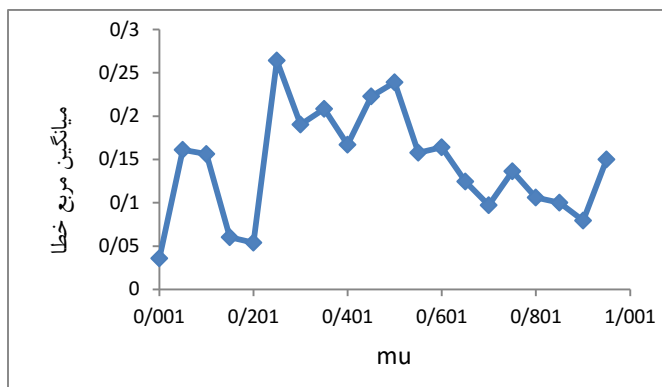
شکل ۴-۹- منحنی مقدار میانگین مربع خطای سری ارزیابی در برابر تعداد دوره‌های آموزشی متفاوت

#### ۴-۵-۳- بهینه سازی پارامتر $\mu$

برای بهینه کردن مقدار  $\mu$ ، ساختار شبکه با ۷ متغیر ورودی، ۳ گره در لایه پنهان و الگوریتم آموزشی لوبنبرگ مارکوارت، تابع انتقال لایه پنهان لگاریتم سیگموئید و تعداد دوره‌های آموزشی نیز برابر با ۱۰ در نظر گرفته شد. سپس مقدار  $\mu$  از ۰/۰۰۱ تا ۱ با گام‌های ۰/۰۵ تغییر داده شد و آن‌گاه برای هر مورد مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. سپس مقدار میانگین مربع خطا بر حسب  $\mu$  رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۴-۱۲) و شکل (۴-۱۰) مقدار بهینه ۰/۰۰۱ به دست آمد.

جدول ۴-۱۲- میانگین مربع خطای حاصل از سری ارزیابی در مقادیر متفاوت  $\mu$

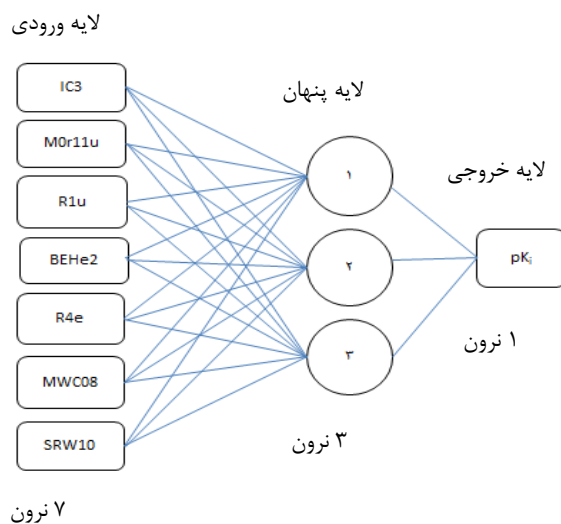
| مقادیر $\mu$ | میانگین مربع خطا | مقادیر $\mu$ | میانگین مربع خطا |
|--------------|------------------|--------------|------------------|
| ۰/۰۰۱        | ۰/۰۳۵۹           | ۰/۵۰۱        | ۰/۲۳۹۰           |
| ۰/۰۵۱        | ۰/۱۶۱۲           | ۰/۵۵۱        | ۰/۱۵۷۹           |
| ۰/۱۰۱        | ۰/۱۵۶۲           | ۰/۶۰۱        | ۰/۱۶۴۱           |
| ۰/۱۵۱        | ۰/۰۶۰۱           | ۰/۶۵۱        | ۰/۱۲۴۶           |
| ۰/۲۰۱        | ۰/۰۵۳۹           | ۰/۷۰۱        | ۰/۰۹۷۲           |
| ۰/۲۵۱        | ۰/۲۶۴۰           | ۰/۷۵۱        | ۰/۱۳۶۲           |
| ۰/۳۰۱        | ۰/۱۹۰۲           | ۰/۸۰۱        | ۰/۱۰۶۱           |
| ۰/۳۵۱        | ۰/۲۰۸۴           | ۰/۸۵۱        | ۰/۱۰۰۲           |
| ۰/۴۰۱        | ۰/۱۶۶۹           | ۰/۹۰۱        | ۰/۰۷۹۵           |
| ۰/۴۵۱        | ۰/۲۲۲۹           | ۰/۹۵۱        | ۰/۱۵۰۰           |



شکل ۴-۱۰ - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر  $\mu$

#### ۴-۵-۴- ساختار شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج به دست آمده در بهینه‌سازی شبکه، شبکه عصبی سه لایه پس انتشار با الگوریتم آموزشی لونیگ مارکوارت و تابع انتقال لگاریتم سیگموئیدی، به عنوان شبکه بهینه انتخاب شد که ساختار آن در شکل (۴-۱۱) نشان داده شده است.



شکل ۴-۱۱ - ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۴-۱۳) ارائه شده است.

جدول ۴-۱۳- توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

|          |                        |
|----------|------------------------|
| trainlm  | تابع آموزش             |
| logsig   | تابع انتقال لایه پنهان |
| pureline | تابع انتقال لایه خروجی |
| ۳        | تعداد نرون لایه پنهان  |
| ۷        | تعداد متغیرهای ورودی   |
| ۱۰       | تعداد دوره‌های آموزش   |
| ۰/۰۰۱    | mu                     |

#### ۴-۶- ارزیابی مدل‌های خطی و غیرخطی

##### ۴-۶-۱- ارزیابی مدل‌های MLR و ANN با استفاده از سری تست

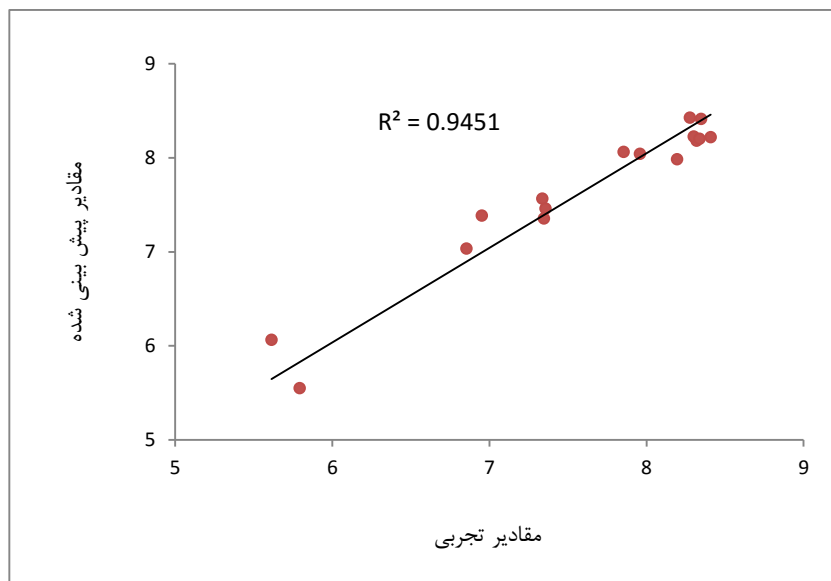
مهمترین مرحله در مطالعات رابطه کمی ساختار- فعالیت ارزیابی قدرت پیش بینی مدل‌ها می باشد، که به صورت پیش بینی رضایت بخش فعالیت مولکول‌هایی که در سری آموزش وجود نداشته اند، مشخص می‌گردد. بدین منظور شبکه عصبی با ۴۵ ترکیب آموزش داده شد. معیار بهینه سازی شبکه، کمترین میزان میانگین مربعات خطا در سری ارزیابی که شامل ۱۶ ترکیب که به صورت تصادفی انتخاب شده اند می باشد. پس از انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی قدرت پیش بینی شبکه، به وسیله ۱۶ ترکیب که در آموزش استفاده نشده‌اند مورد بررسی قرار گرفت. بدین ترتیب که در مدل خطی با استفاده از ضرایب مدل برتر سری آموزش و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر  $pK_i$  این ترکیبات پیش‌بینی شد. نتایج حاصله در جداول ۴-۱۴، ۴-۱۵، ۴-۱۶، ۴-۱۷ به ترتیب برای سری ارزیابی و تست آورده شده‌اند. همچنین نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی برای هر دو سری در شکل‌های ۴-۱۲، ۴-۱۳ نشان داده شده است که  $R^2$  حاصله بیان می‌کند که هر دو روش خطی و غیرخطی در پیش‌بینی ثابت بازدارندگی فعالیت ترکیبات مورد بررسی، توانمند می‌باشند.

جدول ۴-۱۴ - مقادیر تجربی و پیش‌بینی شده‌ی  $pK_i$  به وسیله‌ی روش MLR برای سری ارزیابی

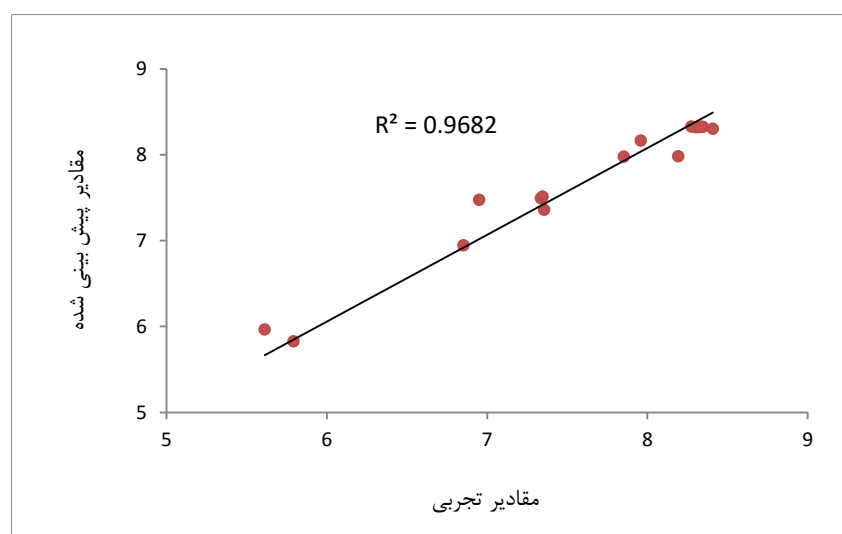
| شماره ترکیب | $pK_i(M)$   |                    | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
|             | مقدار تجربی | مقدار پیش‌بینی شده |                |
| ۳           | ۷/۳۴        | ۷/۵۶               | ۳/۰۰           |
| ۷           | ۷/۹۶        | ۸/۰۴               | ۱/۰۱           |
| ۱۲          | ۶/۸۵        | ۷/۰۳               | ۲/۶۳           |
| ۱۹          | ۸/۳۵        | ۸/۴۱               | ۰/۷۲           |
| ۲۵          | ۸/۳۲        | ۸/۱۸               | -۱/۶۸          |
| ۲۸          | ۸/۳۰        | ۸/۲۲               | -۰/۹۶          |
| ۳۲          | ۸/۲۸        | ۸/۴۳               | ۱/۸۱           |
| ۳۷          | ۸/۴۱        | ۸/۲۲               | -۲/۲۶          |
| ۴۴          | ۶/۹۵        | ۷/۳۸               | ۶/۱۹           |
| ۴۹          | ۸/۱۹        | ۷/۹۸               | -۲/۵۶          |
| ۵۲          | ۷/۸۵        | ۸/۰۶               | ۲/۶۸           |
| ۵۷          | ۸/۳۴        | ۸/۲۰               | -۱/۶۸          |
| ۶۵          | ۵/۷۹        | ۵/۵۵               | -۴/۱۵          |
| ۶۹          | ۵/۶۱        | ۶/۰۶               | ۸/۰۲           |
| ۷۲          | ۷/۳۶        | ۷/۴۶               | ۱/۳۶           |
| ۷۶          | ۷/۳۵        | ۷/۳۵               | ۰              |

جدول ۴-۱۴ - مقادیر تجربی و پیش‌بینی شده‌ی  $pK_i$  به وسیله‌ی روش ANN برای سری ارزیابی

| شماره ترکیب | $pK_i(M)$   |                    | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
|             | مقدار تجربی | مقدار پیش‌بینی شده |                |
| ۳           | ۷/۳۴        | ۷/۴۹               | ۲/۰۴           |
| ۷           | ۷/۹۶        | ۸/۱۷               | ۲/۶۴           |
| ۱۲          | ۶/۸۵        | ۶/۹۴               | ۱/۳۱           |
| ۱۹          | ۸/۳۵        | ۸/۳۳               | -۰/۲۴          |
| ۲۵          | ۸/۳۲        | ۸/۳۲               | ۰              |
| ۲۸          | ۸/۳۰        | ۸/۳۲               | ۰/۲۴           |
| ۳۲          | ۸/۲۸        | ۸/۳۳               | ۰/۶۰           |
| ۳۷          | ۸/۴۱        | ۸/۳۰               | -۱/۳۱          |
| ۴۴          | ۶/۹۵        | ۷/۴۷               | ۷/۴۸           |
| ۴۹          | ۸/۱۹        | ۷/۹۸               | -۲/۵۶          |
| ۵۲          | ۷/۸۵        | ۷/۹۸               | ۱/۶۶           |
| ۵۷          | ۸/۳۴        | ۸/۳۲               | -۰/۲۴          |
| ۶۵          | ۵/۷۹        | ۵/۸۲               | ۰/۵۲           |
| ۶۹          | ۵/۶۱        | ۵/۹۶               | ۶/۲۴           |
| ۷۲          | ۷/۳۶        | ۷/۳۶               | ۰              |
| ۷۶          | ۷/۳۵        | ۷/۵۱               | ۲/۱۸           |



(الف)



(ب)

شکل ۴-۱۲- نمودار مقادیر پیش‌بینی شده  $pK_a$  بر حسب مقادیر تجربی برای سری ارزیابی توسط الف) روش MLR،  
ب) روش ANN

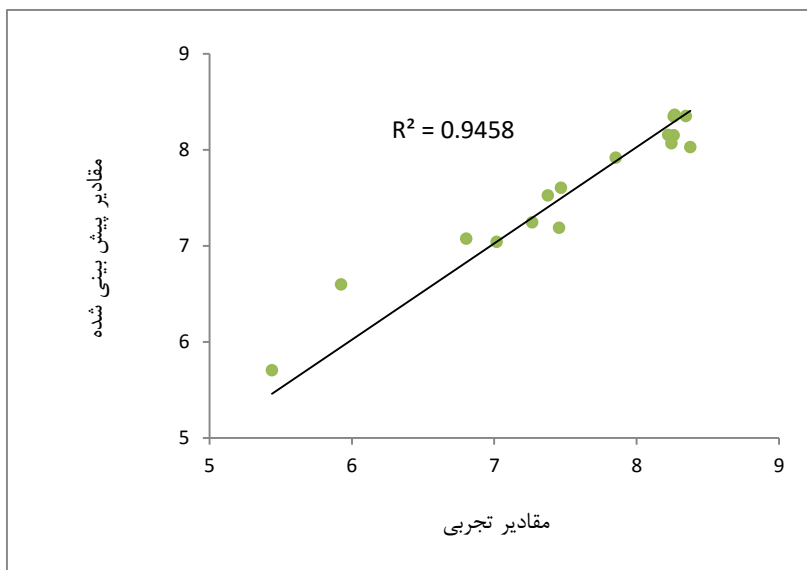
جدول ۴-۱۶ - مقادیر تجربی و پیش‌بینی شده‌ی  $pK_i$  به وسیله‌ی روش MLR برای سری تست

| شماره ترکیب | $pK_i(M)$   |                    | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
|             | مقدار تجربی | مقدار پیش‌بینی شده |                |
| ۱           | ۵/۹۲        | ۶/۶۰               | ۱۱/۴۹          |
| ۸           | ۷/۰۲        | ۷/۰۴               | ۰/۲۸           |
| ۱۰          | ۶/۸۰        | ۷/۰۷               | ۳/۹۷           |
| ۱۸          | ۸/۳۸        | ۸/۰۲               | -۴/۳۰          |
| ۲۳          | ۸/۲۶        | ۸/۱۵               | -۱/۳۳          |
| ۲۷          | ۸/۲۲        | ۸/۱۵               | -۰/۸۵          |
| ۳۴          | ۸/۲۶        | ۸/۳۵               | ۱/۰۹           |
| ۳۸          | ۸/۳۵        | ۸/۳۵               | ۰              |
| ۴۱          | ۷/۴۶        | ۷/۱۹               | -۳/۶۲          |
| ۴۳          | ۸/۲۴        | ۸/۰۷               | -۲/۰۶          |
| ۴۸          | ۷/۴۷        | ۷/۶۰               | ۱/۷۴           |
| ۵۰          | ۷/۲۷        | ۷/۲۴               | -۰/۴۱          |
| ۵۹          | ۸/۲۷        | ۸/۳۶               | ۱/۰۹           |
| ۶۳          | ۵/۴۴        | ۵/۷۰               | ۴/۷۸           |
| ۷۰          | ۷/۸۵        | ۷/۹۲               | ۰/۸۹           |
| ۷۴          | ۷/۳۸        | ۷/۵۲               | ۱/۹۰           |

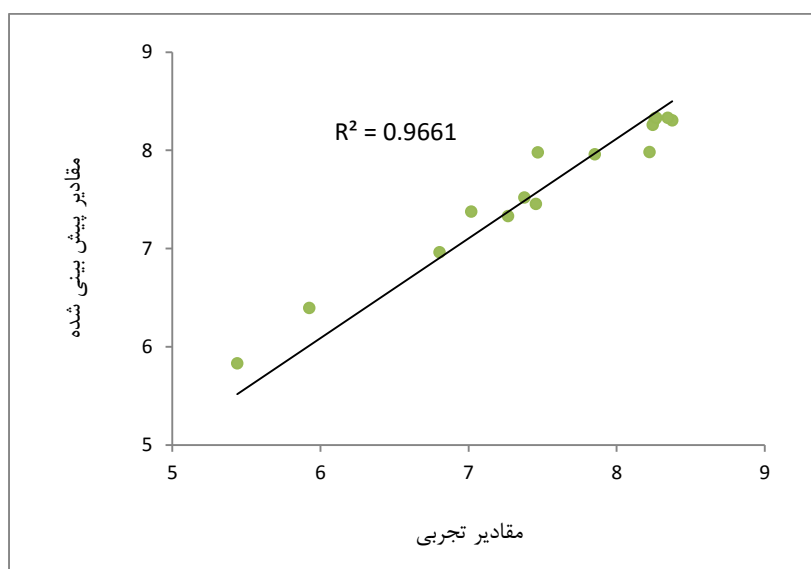
جدول ۴-۱۷ - مقادیر تجربی و پیش‌بینی شده‌ی  $pK_i$  به وسیله‌ی روش ANN برای سری تست

| شماره ترکیب | $pK_i(M)$   |                    | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
|             | مقدار تجربی | مقدار پیش‌بینی شده |                |
| ۱           | ۵/۹۲        | ۶/۳۹               | ۷/۹۴           |
| ۸           | ۷/۰۲        | ۷/۳۸               | ۵/۱۳           |
| ۱۰          | ۶/۸۰        | ۶/۹۶               | ۲/۳۵           |
| ۱۸          | ۸/۳۸        | ۸/۳۱               | -۰/۸۴          |
| ۲۳          | ۸/۲۶        | ۸/۳۱               | ۰/۶۱           |
| ۲۷          | ۸/۲۲        | ۷/۹۸               | -۲/۹۲          |
| ۳۴          | ۸/۲۶        | ۸/۳۲               | ۰/۷۳           |
| ۳۸          | ۸/۳۵        | ۸/۳۳               | -۰/۲۴          |
| ۴۱          | ۷/۴۶        | ۷/۴۵               | -۰/۱۳          |
| ۴۳          | ۸/۲۴        | ۸/۲۶               | ۰/۲۴           |
| ۴۸          | ۷/۴۷        | ۷/۹۸               | ۶/۸۳           |
| ۵۰          | ۷/۲۷        | ۷/۳۳               | ۰/۸۳           |
| ۵۹          | ۸/۲۷        | ۸/۳۳               | ۰/۷۳           |
| ۶۳          | ۵/۴۴        | ۵/۸۳               | ۷/۱۷           |
| ۷۰          | ۷/۸۵        | ۷/۹۶               | ۱/۴۰           |
| ۷۴          | ۷/۳۸        | ۷/۵۲               | ۱/۹۰           |





(الف)



(ب)

شکل ۴-۱۳- نمودار مقادیر پیش بینی شده ی  $pK_i$  بر حسب مقادیر تجربی برای سری تست توسط الف) روش MLR ،  
 ب) روش ANN

#### ۴-۶-۲- ارزیابی مدل‌های ANN و MLR به روش حذف مرحله‌ای تک تک داده‌ها

برای ارزیابی مدل به دست آمده، روش حذف مرحله‌ای تک تک داده‌ها به کار گرفته شد. در این روش یکی از ترکیبات از سری داده‌ها کنار گذاشته شد و سپس مدل‌سازی با ۷۶ ترکیب باقیمانده صورت گرفت. این مدل برای پیش‌بینی  $pK_i$  ترکیب کنار گذاشته شده به کار گرفته شد، سپس این فرایند برای تمام ترکیبات سری داده‌ها تکرار شد. نتایج حاصله برای دو روش ANN و MLR در جدول‌های ۴-۱۸، ۴-۱۹ و منحنی مقادیر پیش‌بینی شده  $pK_i$  بر حسب مقادیر تجربی در شکل ۴-۱۴ آورده شده است. نتایج این ارزیابی توانایی هر دو روش را در پیش‌بینی  $pK_i$  نشان می‌دهد.

جدول ۴-۱۸- ارزیابی مدل MLR به روش حذف مرحله‌ای تک تک داده‌ها

| شماره ترکیب | مقدار واقعی | مقدار پیش‌بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۱           | ۵/۹۲        | ۶/۵۹               | ۱۱/۳۲          |
| ۲           | ۱۵/۸        | ۷/۹۶               | -۲/۳۳          |
| ۳           | ۳۴/۷        | ۷/۵۲               | ۲/۴۵           |
| ۴           | ۴۸/۷        | ۷/۴۹               | ۰/۱۳           |
| ۵           | ۲۱/۷        | ۶/۸۷               | -۴/۷۲          |
| ۶           | ۵۱/۷        | ۷/۷۰               | ۲/۵۳           |
| ۷           | ۹۶/۷        | ۸/۰۸               | ۱/۵۱           |
| ۸           | ۰۲/۷        | ۷/۰۵               | ۰/۴۳           |
| ۹           | ۴۶/۷        | ۷/۴۴               | -۰/۲۷          |
| ۱۰          | ۸۰/۶        | ۷/۰۵               | ۳/۶۸           |
| ۱۱          | ۱۷/۷        | ۲۲/۷               | ۰/۷۰           |
| ۱۲          | ۸۵/۶        | ۷/۰۰               | ۲/۱۹           |
| ۱۳          | ۷۴/۷        | ۸۳/۷               | ۱/۱۶           |
| ۱۴          | ۱۶/۷        | ۲۲/۷               | ۰/۸۴           |
| ۱۵          | ۴۶/۸        | ۲۰/۸               | -۳/۰۷          |
| ۱۶          | ۱۶/۸        | ۲۳/۸               | ۰/۸۶           |
| ۱۷          | ۸/۴۳        | ۹۸/۷               | -۵/۳۴          |
| ۱۸          | ۸/۳۸        | ۰۲/۸               | -۴/۳۰          |
| ۱۹          | ۸/۳۵        | ۴۳/۸               | ۰/۹۶           |
| ۲۰          | ۸/۳۵        | ۵۲/۸               | ۲/۰۴           |
| ۲۱          | ۸/۴۳        | ۳۶/۸               | -۰/۸۳          |
| ۲۲          | ۸/۴۳        | ۴۱/۸               | -۰/۲۴          |
| ۲۳          | ۸/۲۶        | ۱۶/۸               | -۱/۲۱          |

| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۲۴          | ۸/۳۵        | ۳۳/۸               | -۰/۲۴          |
| ۲۵          | ۸/۳۲        | ۱۹/۸               | -۱/۵۶          |
| ۲۶          | ۸/۳۱        | ۹۲/۸               | ۷/۳۴           |
| ۲۷          | ۸/۲۲        | ۱۸/۸               | -۰/۴۹          |
| ۲۸          | ۸/۳۰        | ۱۷/۸               | -۱/۵۷          |
| ۲۹          | ۸/۱۵        | ۱۰/۸               | -۰/۶۱          |
| ۳۰          | ۸/۱۹        | ۲۸/۸               | ۱/۱۰           |
| ۳۱          | ۸/۲۹        | ۵۳/۸               | ۲/۹۰           |
| ۳۲          | ۸/۲۸        | ۴۸/۸               | ۲/۴۲           |
| ۳۳          | ۸/۳۵        | ۳۶/۸               | ۰/۱۲           |
| ۳۴          | ۸/۲۶        | ۲۸/۸               | ۰/۲۴           |
| ۳۵          | ۸/۲۸        | ۱۷/۸               | -۱/۳۳          |
| ۳۶          | ۸/۲۸        | ۴۸/۸               | ۲/۴۲           |
| ۳۷          | ۸/۴۱        | ۱۲/۸               | -۳/۴۵          |
| ۳۸          | ۸/۳۵        | ۳۷/۸               | ۰/۲۴           |
| ۳۹          | ۸/۲۸        | ۱۲/۸               | -۱/۹۳          |
| ۴۰          | ۸/۱۳        | ۵۲/۸               | ۴/۸۰           |
| ۴۱          | ۷/۴۶        | ۱۸/۷               | -۳/۷۵          |
| ۴۲          | ۷/۱۹        | ۲۵/۷               | ۰/۸۳           |
| ۴۳          | ۸/۲۴        | ۰/۱/۸              | -۲/۷۹          |
| ۴۴          | ۶/۹۵        | ۳۱/۷               | ۵/۱۸           |
| ۴۵          | ۷/۴۱        | ۸۹/۶               | -۷/۰۲          |
| ۴۶          | ۷/۹۶        | ۹۷/۷               | ۰/۱۳           |
| ۴۷          | ۸/۱۹        | ۳۷/۸               | ۲/۲۰           |
| ۴۸          | ۷/۴۷        | ۶۱/۷               | ۱/۸۷           |
| ۴۹          | ۸/۱۹        | ۹۷/۷               | -۲/۶۹          |
| ۵۰          | ۷/۲۷        | ۲۰/۷               | -۰/۹۶          |
| ۵۱          | ۷/۷۴        | ۷۰/۷               | -۰/۵۲          |
| ۵۲          | ۷/۸۵        | ۰۲/۸               | ۲/۱۷           |
| ۵۳          | ۸/۴۰        | ۱۵/۸               | -۲/۹۸          |
| ۵۴          | ۸/۱۲        | ۹۶/۷               | -۱/۹۷          |
| ۵۵          | ۸/۲۴        | ۲۳/۸               | -۰/۱۲          |
| ۵۶          | ۸/۲۱        | ۲۵/۸               | ۰/۴۹           |
| ۵۷          | ۸/۳۴        | ۲۴/۸               | -۱/۲۰          |
| ۵۸          | ۸/۳۶        | ۱۲/۸               | -۲/۸۷          |
| ۵۹          | ۸/۲۷        | ۳۹/۸               | ۱/۴۵           |
| ۶۰          | ۸/۴۲        | ۳۰/۸               | -۱/۴۳          |
| ۶۱          | ۸/۴۴        | ۳۹/۸               | -۰/۵۹          |
| ۶۲          | ۸/۴۳        | ۴۳/۸               | ۰              |

ادامه جدول ۴-۱۸

| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۶۳          | ۵/۴۴        | ۵۹/۵               | ۲/۷۶           |
| ۶۴          | ۶/۴۵        | ۶۳/۶               | ۲/۷۹           |
| ۶۵          | ۵/۷۹        | ۳۷/۵               | -۷/۲۵          |
| ۶۶          | ۶/۸۲        | ۷۵/۶               | -۱/۰۳          |
| ۶۷          | ۵/۸۴        | ۹۲/۵               | ۱/۳۷           |
| ۶۸          | ۵/۸۸        | ۳۷/۵               | -۸/۶۷          |
| ۶۹          | ۵/۶۱        | ۰۲/۶               | ۷/۳۱           |
| ۷۰          | ۷/۸۵        | ۹۲/۷               | ۰/۸۹           |
| ۷۱          | ۷/۳۳        | ۳۳/۷               | ۰              |
| ۷۲          | ۷/۳۶        | ۴۰/۷               | ۰/۵۴           |
| ۷۳          | ۷/۵۵        | ۵۵/۷               | ۰              |
| ۷۴          | ۷/۳۸        | ۵۱/۷               | ۱/۷۶           |
| ۷۵          | ۶/۷۱        | ۷۲/۶               | ۰/۱۵           |
| ۷۶          | ۷/۳۵        | ۳۷/۷               | ۰/۲۷           |
| ۷۷          | ۸/۴۰        | ۲۷/۸               | -۱/۵۵          |

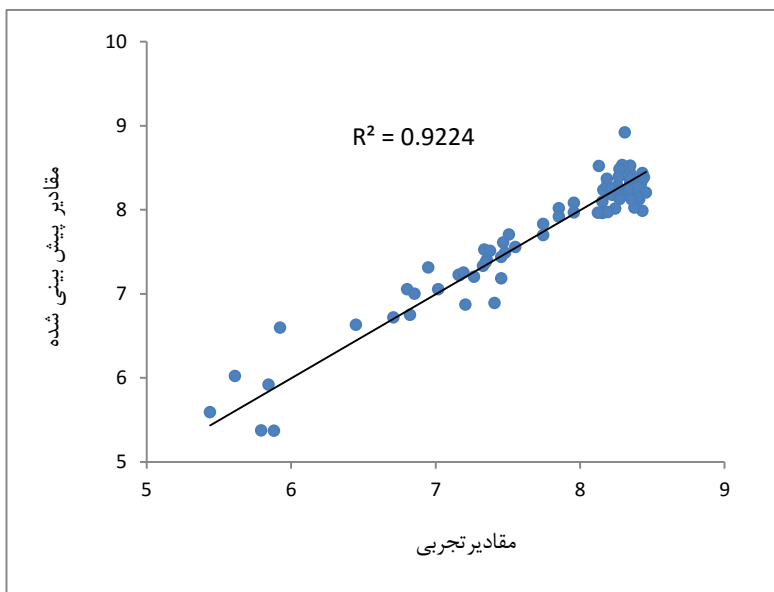
جدول ۴-۱۹- ارزیابی مدل ANN به روش حذف مرحله‌ای تک تک داده‌ها

| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۱           | ۵/۹۲        | ۵۰/۶               | ۹/۸۰           |
| ۲           | ۱۵/۸        | ۸۸/۷               | -۳/۳۱          |
| ۳           | ۳۴/۷        | ۵۳/۷               | ۲/۵۹           |
| ۴           | ۴۸/۷        | ۳۰/۷               | -۲/۴۱          |
| ۵           | ۲۱/۷        | ۷۹/۶               | -۵/۸۳          |
| ۶           | ۵۱/۷        | ۰۳/۷               | -۶/۳۹          |
| ۷           | ۹۶/۷        | ۱۴/۸               | ۲/۲۶           |
| ۸           | ۰۲/۷        | ۱۸/۷               | ۲/۲۸           |
| ۹           | ۴۶/۷        | ۴۷/۷               | ۰/۱۳           |
| ۱۰          | ۸۰/۶        | ۹۸/۶               | ۲/۶۵           |
| ۱۱          | ۱۷/۷        | ۱۰/۷               | -۰/۹۸          |
| ۱۲          | ۸۵/۶        | ۹۱/۶               | ۰/۸۸           |
| ۱۳          | ۷۴/۷        | ۸۸/۷               | ۱/۸۱           |
| ۱۴          | ۱۶/۷        | ۰۸/۷               | -۱/۱۲          |
| ۱۵          | ۴۶/۸        | ۲۵/۸               | -۲/۴۸          |
| ۱۶          | ۱۶/۸        | ۲۰/۸               | ۰/۴۹           |
| ۱۷          | ۸/۴۳        | ۲۸/۸               | -۱/۷۸          |
| ۱۸          | ۸/۳۸        | ۱۶/۸               | -۲/۶۳          |
| ۱۹          | ۸/۳۵        | ۳۵/۸               | ۰              |
| ۲۰          | ۸/۳۵        | ۴۲/۸               | ۰/۸۴           |

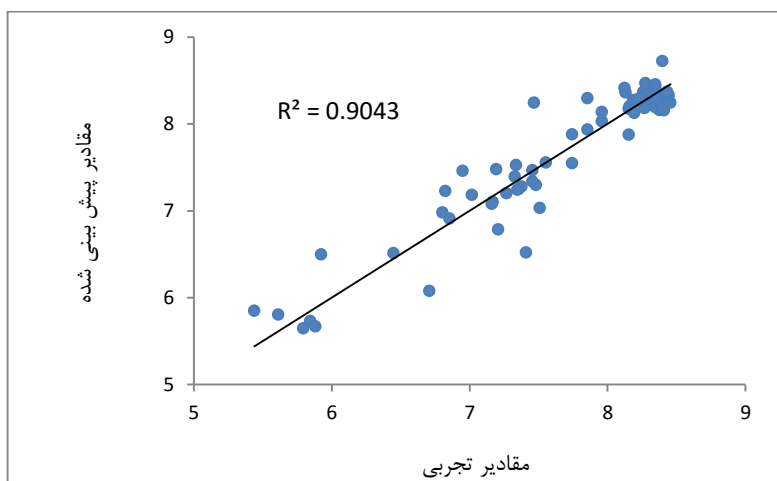
| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۲۱          | ۸/۴۳        | ۲۹/۸               | -۱/۶۶          |
| ۲۲          | ۸/۴۳        | ۲۶/۸               | -۲/۰۲          |
| ۲۳          | ۸/۲۶        | ۳۶/۸               | ۱/۲۱           |
| ۲۴          | ۸/۳۵        | ۱۹/۸               | -۱/۹۲          |
| ۲۵          | ۸/۳۲        | ۲۵/۸               | -۰/۸۴          |
| ۲۶          | ۸/۳۱        | ۳۴/۸               | ۰/۳۶           |
| ۲۷          | ۸/۲۲        | ۲۸/۸               | ۰/۷۳           |
| ۲۸          | ۸/۳۰        | ۲۹/۸               | -۰/۱۲          |
| ۲۹          | ۸/۱۵        | ۱۸/۸               | ۰/۳۷           |
| ۳۰          | ۸/۱۹        | ۲۵/۸               | ۰/۷۳           |
| ۳۱          | ۸/۲۹        | ۳۶/۸               | ۰/۸۴           |
| ۳۲          | ۸/۲۸        | ۴۷/۸               | ۲/۲۹           |
| ۳۳          | ۸/۳۵        | ۴۵/۸               | ۱/۲۰           |
| ۳۴          | ۸/۲۶        | ۲۹/۸               | ۰/۳۶           |
| ۳۵          | ۸/۲۸        | ۳۳/۸               | ۰/۶۰           |
| ۳۶          | ۸/۲۸        | ۳۴/۸               | ۰/۷۲           |
| ۳۷          | ۸/۴۱        | ۱۶/۸               | -۲/۹۷          |
| ۳۸          | ۸/۳۵        | ۲۹/۸               | -۰/۷۲          |
| ۳۹          | ۸/۲۸        | ۲۳/۸               | -۰/۶۰          |
| ۴۰          | ۸/۱۳        | ۳۶/۸               | ۲/۸۳           |
| ۴۱          | ۷/۴۶        | ۳۴/۷               | -۱/۶۱          |
| ۴۲          | ۷/۱۹        | ۴۸/۷               | ۴/۰۳           |
| ۴۳          | ۸/۲۴        | ۲۹/۸               | ۰/۶۱           |
| ۴۴          | ۶/۹۵        | ۴۶/۷               | ۷/۳۴           |
| ۴۵          | ۷/۴۱        | ۵۲/۶               | -۱۲/۰۱         |
| ۴۶          | ۷/۹۶        | ۰۳/۸               | ۰/۸۸           |
| ۴۷          | ۸/۱۹        | ۲۷/۸               | ۰/۹۸           |
| ۴۸          | ۷/۴۷        | ۲۵/۸               | ۱۰/۴۴          |
| ۴۹          | ۸/۱۹        | ۱۳/۸               | -۰/۷۳          |
| ۵۰          | ۷/۲۷        | ۲۰/۷               | -۰/۹۶          |
| ۵۱          | ۷/۷۴        | ۵۵/۷               | -۲/۴۵          |
| ۵۲          | ۷/۸۵        | ۳۰/۸               | ۵/۷۳           |
| ۵۳          | ۸/۴۰        | ۱۷/۸               | -۲/۷۴          |
| ۵۴          | ۸/۱۲        | ۴۱/۸               | ۳/۵۷           |
| ۵۵          | ۸/۲۴        | ۲۹/۸               | ۰/۶۱           |
| ۵۶          | ۸/۲۱        | ۲۶/۸               | ۰/۶۱           |
| ۵۷          | ۸/۳۴        | ۳۲/۸               | -۰/۲۴          |
| ۵۸          | ۸/۳۶        | ۲۹/۸               | -۰/۸۴          |
| ۵۹          | ۸/۲۷        | ۱۸/۸               | -۱/۰۹          |

## ادامه جدول ۴-۱۹

| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۶۰          | ۸/۴۲        | ۳۵/۸               | -۰/۸۳          |
| ۶۱          | ۸/۴۴        | ۳۳/۸               | -۱/۳۰          |
| ۶۲          | ۸/۴۳        | ۳۷/۸               | -۰/۷۱          |
| ۶۳          | ۵/۴۴        | ۸۵/۵               | ۷/۵۴           |
| ۶۴          | ۶/۴۵        | ۵۱/۶               | ۰/۹۳           |
| ۶۵          | ۵/۷۹        | ۶۵/۵               | -۲/۴۲          |
| ۶۶          | ۶/۸۲        | ۲۳/۷               | ۶/۰۱           |
| ۶۷          | ۵/۸۴        | ۷۳/۵               | -۱/۸۸          |
| ۶۸          | ۵/۸۸        | ۶۷/۵               | -۳/۵۷          |
| ۶۹          | ۵/۶۱        | ۸۰/۵               | ۳/۳۹           |
| ۷۰          | ۷/۸۵        | ۹۳/۷               | ۱/۰۲           |
| ۷۱          | ۷/۳۳        | ۳۹/۷               | ۰/۸۲           |
| ۷۲          | ۷/۳۶        | ۲۵/۷               | -۱/۴۹          |
| ۷۳          | ۷/۵۵        | ۵۶/۷               | ۰/۱۳           |
| ۷۴          | ۷/۳۸        | ۲۸/۷               | -۱/۳۶          |
| ۷۵          | ۶/۷۱        | ۰۸/۶               | -۹/۳۹          |
| ۷۶          | ۷/۳۵        | ۲۴/۷               | -۱/۵۰          |
| ۷۷          | ۸/۴۰        | ۷۲/۸               | ۳/۸۱           |



(الف)

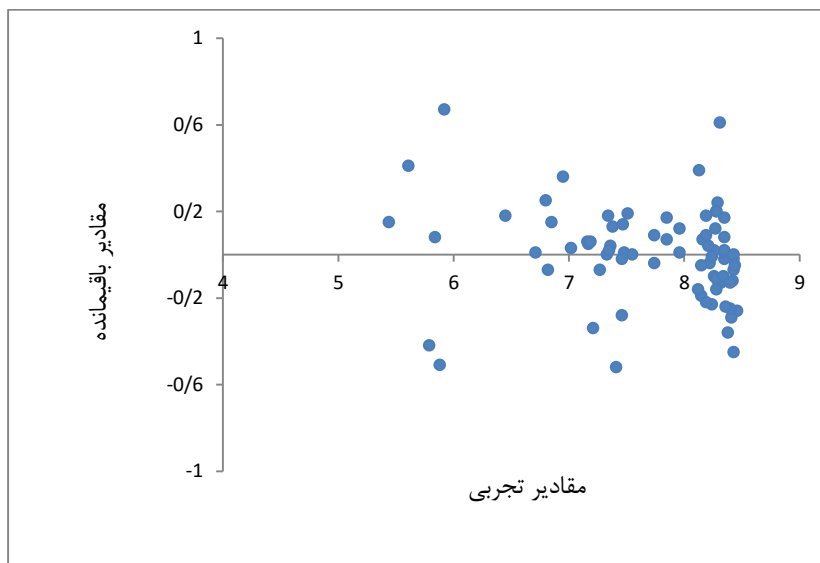


(ب)

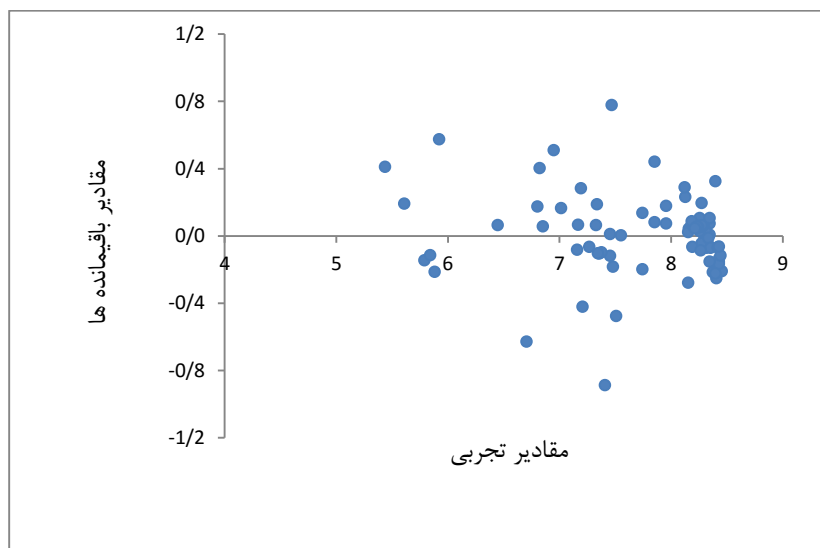
شکل ۴-۱۴- نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی  $pK_i$  به روش حذف مرحله‌ای تک تک داده‌ها الف)

روش ANN ب) روش MLR

نمودار باقیمانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد. اگر باقیمانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند نشان می‌دهد که خطای معین در مدل وجود ندارد. شکل (۴-۱۵) نمودار باقیمانده‌ها را که از تفاضل مقدار تجربی ثابت بازداري فعالیت از مقدار محاسبه شده بر حسب مقدار تجربی توسط مدل‌های خطی و غیرخطی به دست می‌آید، نشان می‌دهد. پراکنندگی نسبتاً متقارن نقاط در دو طرف محور افقی نشان دهنده عدم وجود خطای معین است.



(الف)



(ب)

شکل ۴-۱۵- نمودار باقیمانده‌ها بر حسب مقدار تجربی  $pK_i$  برای کل داده‌ها الف) روش MLR ب) روش ANN



#### ۴-۶-۳- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری

مطابق جدول (۴-۲۰) هفت پارامتر آماری، جهت ارزیابی توانایی پیش‌گویی مدل‌های ساخته شده به روش خطی (MLR) و غیرخطی (ANN) به کار گرفته شد. مقادیر این پارامترها نشان می‌دهد که هر دو مدل ارائه شده از قدرت پیش‌گویی مناسبی برخوردار هستند.

جدول ۴-۲۰- پارامترهای آماری دو روش MLR و ANN

| پارامتر        | سری تست (n=۱۶) |        | سری ارزیابی (n=۱۶) |        | کل داده‌ها (n=۷۷) |        |
|----------------|----------------|--------|--------------------|--------|-------------------|--------|
|                | MLR            | ANN    | MLR                | ANN    | MLR               | ANN    |
| MAE            | ۰/۱۷۲۵         | ۰/۱۷۰۶ | ۰/۱۸۰۰             | ۰/۱۲۹۴ | ۰/۱۵۸۸            | ۰/۱۷۰۹ |
| MSE            | ۰/۰۵۶۹         | ۰/۰۵۶۰ | ۰/۰۴۶۱             | ۰/۰۳۵۷ | ۰/۰۴۷۳            | ۰/۰۵۹۷ |
| PRESS          | ۰/۹۱۰۴         | ۰/۸۹۶۰ | ۰/۷۳۷۶             | ۰/۵۷۱۲ | ۳/۶۴۲۱            | ۴/۵۹۶۹ |
| SEP            | ۰/۲۳۸۵         | ۰/۲۳۶۶ | ۰/۲۱۴۷             | ۰/۱۸۸۹ | ۰/۲۱۷۴            | ۰/۲۴۴۳ |
| R <sup>2</sup> | ۰/۹۴۵۸         | ۰/۹۶۶۱ | ۰/۹۴۵۱             | ۰/۹۶۸۲ | ۰/۹۲۲۴            | ۰/۹۰۴۳ |
| REP(%)         | ۳/۱۶۴۴         | ۳/۱۳۹۲ | ۲/۸۳۳۱             | ۲/۴۹۲۷ | ۲/۸۱۴۹            | ۳/۱۶۳۲ |
| MRE            | ۲/۴۸۷۷         | ۲/۴۹۸۱ | ۲/۵۴۳۵             | ۱/۸۲۹۰ | ۲/۱۴۷۷            | ۲/۳۲۸۸ |

#### ۴-۶-۴- ارزیابی شبکه به روش Y- تصادفی

برای بررسی هر گونه ارتباط تصادفی بین داده‌ها آزمون Y- تصادفی به کار گرفته شد. در این تکنیک، pKi تجربی ترکیبات دارویی به صورت تصادفی جابجا و مدل‌های جدیدی ایجاد شدند. اگر بین مقدار ضریب تعیین مدل اصلی و مدل‌های جدید تفاوت قابل توجهی وجود داشته باشد، نشان می‌دهد که هیچ گونه ارتباط تصادفی در مدل اصلی وجود ندارد. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۴-۲۱) نشان داده شده است. مقادیر پایین R<sup>2</sup> در این آزمون بیان‌گر تصادفی نبودن مدل اصلی می‌باشد.

جدول ۴-۲۱- مقادیر  $R^2$  برای سری تست و ارزیابی با استفاده از آزمون  $Y$ -تصادفی

| شماره تکرار آزمون | $R^2$ برای سری ارزیابی | $R^2$ برای سری تست |
|-------------------|------------------------|--------------------|
| ۱                 | ۰/۰۴۴۵                 | ۰/۰۰۴۹             |
| ۲                 | ۰/۰۷۶۱                 | ۰/۰۱۶۴             |
| ۳                 | ۰/۰۴۴۰                 | ۰/۰۷۳۲             |
| ۴                 | ۰/۰۰۰۵                 | ۰/۰۳۳۹             |
| ۵                 | ۰/۰۱۸۸                 | ۰/۰۷۵۴             |
| ۶                 | ۰/۰۱۹۶                 | ۰/۲۳۳              |
| ۷                 | ۰/۰۱۵۳                 | ۰/۰۶۳۰             |
| ۸                 | ۰/۰۰۹۰                 | ۰/۰۰۳۴۲            |
| ۹                 | ۰/۰۰۳۹                 | ۰/۰۴۸۱             |
| ۱۰                | ۰/۰۱۶۷                 | ۰/۱۱۲۲             |

#### ۴-۷- بررسی ارتباط توصیف‌کننده‌های منتخب با ثابت بازداري فعاليت ترکیبات

در این بخش با توجه به توصیف‌کننده‌های وارد شده در مدل‌های ANN و MLR یک بررسی اجمالی روی اثرات مختلف موجود در ثابت بازداري ترکیبات دارویی صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۷ توصیف‌کننده بود که هرکدام بیانگر خصوصیات متفاوتی از ترکیبات مورد بررسی است. این توصیف‌کننده‌ها شامل IC3, Mor11u, R1u, BEHe2, R4e, MWC08 و SRW10 می‌باشند.

#### ۴-۷-۱- توصیف‌کننده‌های Topological

این توصیف‌کننده‌ها از گراف مولکولی تهی از هیدروژن محاسبه می‌شوند، گراف مولکولی را با علامت  $G = G(V, E)$  تعریف می‌کنند که در این رابطه  $V$  به عنوان رئوس که نشان‌دهنده تعداد اتم-های موجود در ساختار مولکول بوده، و  $E$  به عنوان یال‌ها که نشان‌دهنده پیوندهای متصل‌کننده این رئوس (اتم‌ها) به یکدیگر هستند، معرفی می‌شوند.

تعداد یال‌های متصل به یک رأس را درجه آن رأس می‌نامند. یک گراف مولکولی که بدون در نظر گرفتن اتم‌های هیدروژن به دست بیاید، گراف مولکولی تهی شده از هیدروژن نامیده می‌شود. با استفاده از نظریه گراف و با در نظر گرفتن موقعیت رئوس مولکول نسبت به یکدیگر می‌توان ساختار یک مولکول را به صورت یک ماتریس نشان داد. این ماتریس با افزایش تعداد اتم‌ها در مولکول (رئوس) و نیز با بزرگتر شدن مولکول و ورود حلقه‌ها به ساختار آن و همچنین پرشاخه‌تر شدن مولکول پیچیده‌تر می‌شود [۴۶].

توصیف‌کننده IC3<sup>۱</sup> از این دسته توصیف‌کننده‌ها می‌باشد، این توصیف‌کننده‌ها جزء شاخص‌های اطلاعاتی هستند و در بردارنده اطلاعاتی راجع به ماهیت شیمیایی، پیوند اتم‌ها در فضا، توپولوژی مولکولی و تقارن می‌باشند [۵۶]. بنابراین هر چه مولکول حجیم‌تر و پر شاخه‌تر گردد مقدار عددی این توصیف‌کننده افزایش می‌یابد. علامت مثبت اثر متوسط این توصیف‌کننده در معادله نشان دهنده تأثیر پیوندها و نحوه قرار گرفتن اتم‌ها در فضای مولکولی، روی  $pK_i$  می‌باشد.

#### ۴-۷-۲- توصیف‌کننده‌های 3D - MORSE

توصیف‌کننده های 3D - Morse (نمایش سه بعدی ساختار مولکول براساس تفرق الکترون) از

طریق معادله تبدیلی که در پراش الکترون استفاده می‌شود، محاسبه می‌گردند:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (3-4)$$

I شدت الکترون پراکنده شده،  $A_i$  و  $A_j$  خاصیت اتمی اتم  $i$  و  $j$ ،  $s$  زاویه پراکندگی،  $r_{ij}$  فاصله بین

اتم‌های  $i$  و  $j$ ،  $N$  تعداد کل اتم‌ها را نشان می‌دهد. این روش باعث می‌شود که ساختار سه بعدی مولکول به یک کد ثابت تبدیل شود.

۱ . Information content index(neighborhood symmetry of 3-order)

۲ . 3D- Molecular Representation of Structure based on Electron diffraction

۳ . 3D -MORSE-signal 11 un weighted

یکی از این توصیف‌کننده‌های سه بعدی Mor11u<sup>۳</sup> است که در مدل نهایی دیده می‌شود. چون این توصیف‌کننده آرایش سه بعدی اتم‌ها را بدون اینکه به اندازه مولکول ارتباط داشته باشند، بیان می‌کند بنابراین برای تعداد زیادی مولکول با تفاوت‌های ساختاری زیاد قابل کاربرد است [۵۷]. بنابراین با توجه به علامت مثبت اثر متوسط آن در مدل می‌توان گفت که با افزایش مقدار این توصیف‌کننده مقدار pK<sub>i</sub> نیز افزایش می‌یابد.

#### ۴-۷-۳- توصیف‌کننده‌های GETAWAY<sup>۱</sup>

این توصیف‌کننده‌ها که بیان‌کننده ویژگی‌های توپولوژیکی و هندسی مولکول هستند، از ماتریس تأثیر مولکولی (MIM)<sup>۲</sup> محاسبه می‌شوند. این ماتریس به آسانی از مختصات فضایی اتم‌های مولکول در یک صورتبندی مشخص به دست می‌آید که با رابطه زیر بیان می‌گردد:

$$H = M (M^T M) \cdot M^T \quad (۴-۴)$$

که M ماتریس مختصات اتمی، بالانویس T به معنای ماتریس ترانپوز<sup>۳</sup> و H ماتریس تأثیر مولکولی است. عناصر قطری ( $h_{ii}$ ) ماتریس H لوریج<sup>۴</sup> نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. به عنوان مثال اتم‌های دورتر نسبت به اتم‌های نزدیکتر به مرکز اتم، دارای  $h_{ii}$  بزرگتر هستند.

اولین سری از این توصیف‌کننده‌های مولکولی، GETAWAY H- نامیده می‌شوند که تنها توسط اطلاعات به دست آمده توسط ماتریس نفوذ مولکولی محاسبه می‌شوند در حالی که سری R- GETAWAY، ترکیبی از اطلاعات ماتریس تأثیر – فاصله است که به صورت زیر تعریف می‌شود:

---

۱ . Geometrical, Topology and atom Weighted assembly  
 ۲ . Molecular Influence Matrix  
 ۳ . Transpose matrix  
 ۴ . Leverage

$$[R]_{ij} = \left[ \frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \quad (5-4)$$

که  $h_{ii}$  و  $h_{jj}$  لوریج (قدرت نفوذ) دو اتم  $r_{ij}$  و فاصله آن دو اتم است [58].

$R1u$  (از دسته توصیف کننده‌ی R-GETAWAY) از لوریج‌های دو اتم که در فاصله توپولوژیکی ۱ قرار دارند و با ویژگی خاصی از مولکول وزن دار نشده، محاسبه می‌شود و توصیف کننده  $R4e$  که بر اساس الکترونگاتیویته ساندرسون وزن دار شده است.

این توصیف کننده‌ها از آن جهت مهم هستند که به ساختار سه بعدی مولکول حساس هستند و معمولاً برای مقایسه مولکول‌ها و حتی صورتبندی‌ها از نظر شکل، تقارن، اندازه اتم به کار می‌روند. با توجه به علامت منفی اثر متوسط این توصیف کننده‌ها در مدل می‌توان گفت افزایش مقدار این توصیف کننده‌ها، مقدار  $pK_i$  را کاهش می‌دهد.

#### ۴-۷-۴- توصیف کننده‌های BCUT<sup>۲</sup>

این توصیف کننده‌ها از دسته توصیف کننده‌های دوبعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده<sup>۴</sup> (ماتریس بردن) می‌باشند. ماتریس بردن یک گراف تهی از هیدروژن را ارائه می‌دهد که در آن  $B_{ii}$ ، یعنی عناصر قطری در ارتباط با خاصیت اتمی مانند الکترونگاتیویته، حجم و اندروالس، جرم اتمی و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند [59 و 60].

۱ . R autocorrelation of lag 1 unweighted

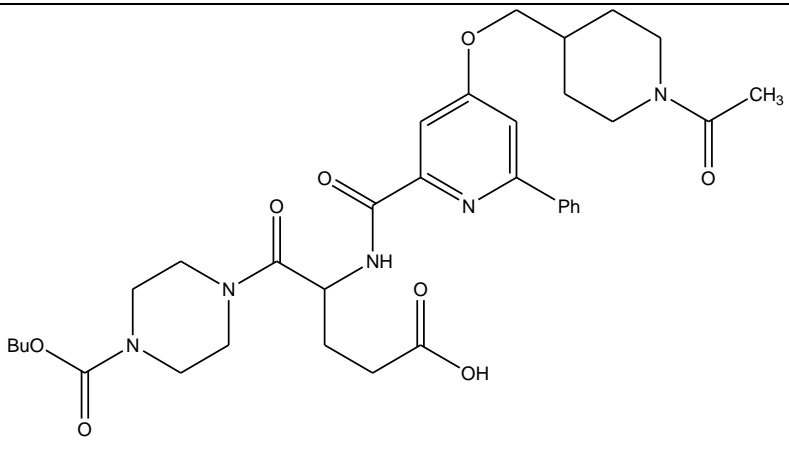
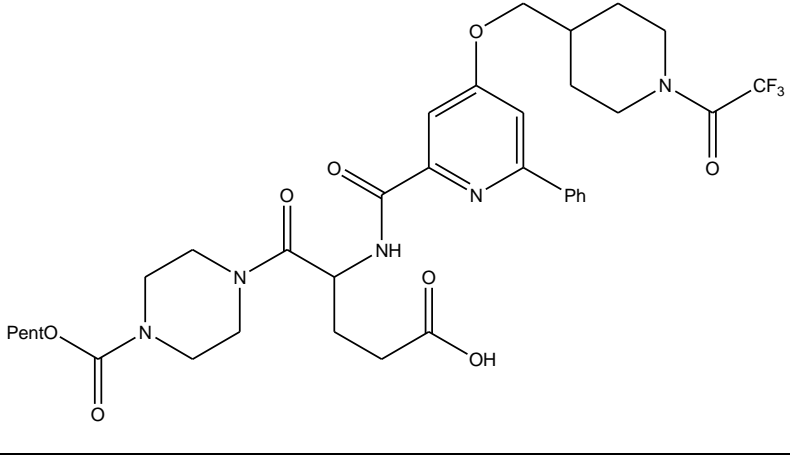
۲ . R autocorrelation of lag 4/weighted by atomic sanderson electronegativities

۳ . Burden-CAS- university of texas eigenvalues

۴ . Modified connectivity matrix

${}^1\text{BEHe2}$  بالاترین مقدار ویژه از ماتریس بودن که با الکترونگاتیویته ساندرسون وزن دار شده است، بزرگترین ضریب را در مدل به خود اختصاص داده است. هر چه الکترونگاتیویته اتم‌های تشکیل دهنده مولکول بیشتر باشد مقدار عددی این توصیف‌کننده افزایش پیدا می‌کند. با توجه به علامت منفی اثر متوسط این توصیف‌کننده در معادله می‌توان گفت که مقادیر بالاتر این توصیف‌کننده باعث کاهش مقدار  $\text{pK}_i$  می‌شود. مثالی از اثر این توصیف‌کننده در جدول (۴-۲۲) ارائه شده است.

جدول ۴-۲۲- مثالی از اثر توصیف‌کننده  $\text{BEHe2}$  بر ثابت بازداري فعالیت

| ترکیب   | $\text{BEHe2}$ | $\text{pK}_i$ |
|---|----------------|---------------|
|   | ۳/۸۲۷          | ۸/۴۱          |
|  | ۳/۸۲۹          | ۸/۳۵          |

۱ . highest eigenvalue n.2 of Burden matrix/weighted by atomic sanderson electronegativities

#### ۴-۷-۵- توصیف‌کننده‌های Molecular walk counts

توصیف‌کننده‌های MWC08<sup>۱</sup> و SRW10<sup>۲</sup> از دسته توصیف‌کننده‌های شمارش‌های مسیر مولکولی هستند. که بیانگر خصوصیات ساختمانی مانند شاخه‌دار شدن و اندازه مولکول می‌باشند. به طور کل می‌توان توصیف‌کننده MWC08 به عنوان پیچیدگی گراف مولکولی در نظر گرفت [۶۱].

توصیف‌کننده SRW10 که به عنوان توصیف‌کننده‌های مولکولی و اتمی شناخته می‌شوند، بر اساس گراف مولکولی تهی از هیدروژن به دست می‌آیند [۶۲]. با شاخه‌دار شدن و حجیم شدن مولکول مقدار عددی این توصیف‌کننده‌ها افزایش می‌یابد. علامت مثبت اثر متوسط این توصیف‌کننده‌ها در مدل نشان می‌دهد که با افزایش مقدار این توصیف‌کننده‌ها مقدار  $pK_i$  نیز افزایش پیدا می‌کند.

#### ۴-۸- نتیجه‌گیری

در این تحقیق از مدل‌های خطی و غیرخطی برای پیش‌بینی ثابت بازدارندگی فعالیت بعضی ترکیبات دارویی استفاده شده است. بنابراین بر اساس این مدل‌ها می‌توان ثابت بازدارندگی فعالیت ترکیبات دارویی، که از لخته شدن خون جلوگیری می‌کنند را برای همین طبقه از ترکیبات که هنوز سنتز نشده‌اند پیش‌بینی کرد. بنابراین با صرف هزینه و وقت کمتر می‌توان پیش از سنتز ترکیب دارویی با استفاده از مدل معتبر فعالیت دارویی ترکیب را پیش‌بینی کرد تا داروهایی با فعالیت دارویی مناسب با اثر بخشی بهتر تهیه شود.

۱ . molecular walk count of order 08

۲ . Self-returning walk count of order 10

## فصل پنجم

مطالعه ارتباط کمی ساختار-فعالیت مشتقات

۵ ، ۶- دی هیدروکسی پیریمیدین-۴-

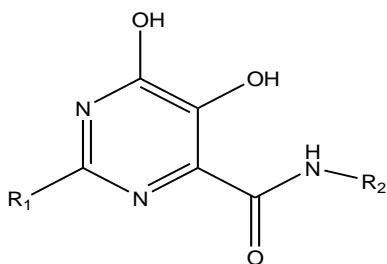
کربوکسامید با استفاده از روش رگرسیون

خطی چندگانه و شبکه عصبی مصنوعی



## ۱-۵- سری داده‌ها

ترکیبات مورد استفاده در این تحقیق، داده‌های تجربی ( $pIC_{50}$ ) ۳۴ ترکیب از مشتقات ۵، ۶-دی هیدروکسی پیریمیدین-۴-کربوکسامید به عنوان سری داده‌ها مورد استفاده قرار گرفت [۶۳]. ساختار کلی این ترکیبات در شکل ۱-۵ آمده است. همچنین مقادیر تجربی  $pIC_{50}$  این ترکیبات بر حسب مولار به همراه استخلافهای آنان در جدول ۱-۵ آمده است.

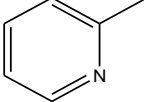
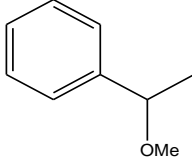
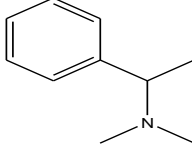
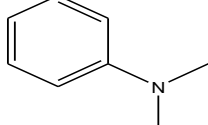
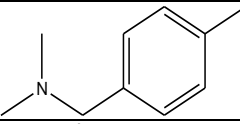
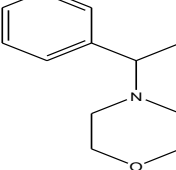
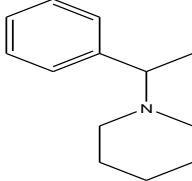
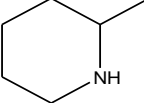
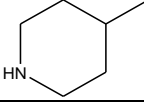
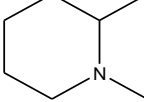
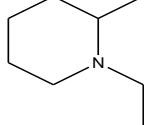
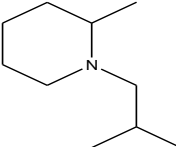


شکل ۱-۵- ساختار کلی ترکیبات ۵، ۶-دی هیدروکسی پیریمیدین-۴-کربوکسامید

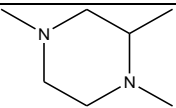
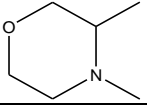
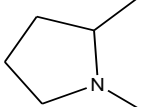
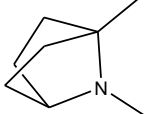
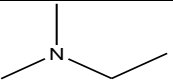
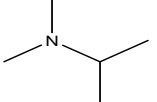
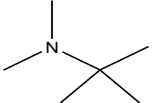
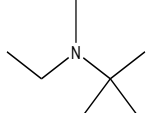
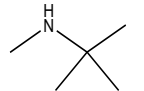
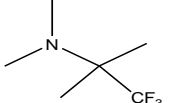
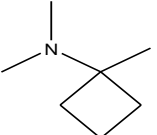
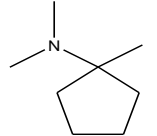
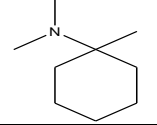
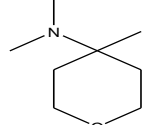
جدول ۱-۵- سری ترکیبات مورد مطالعه

| شماره ترکیب | R <sub>1</sub>  | R <sub>2</sub>                                       | $pIC_{50}(M)$ |
|-------------|-----------------|--|---------------|
| ۱           |                 | Ph   | ۷/۱۶          |
| ۲           |                 | H  | ۶/۰۰          |
| ۳           |                 | -CH <sub>2</sub> <sup>c</sup> hexyl                  | ۴/۳۰          |
| ۴           |                 | -CH <sub>2</sub> CH <sub>2</sub> Ph                  | ۴/۳۰          |
| ۵           |                 | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۷۰          |
| ۶           | CH <sub>3</sub> | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۲۲          |
| ۷           | <sup>i</sup> Pr | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۲۲          |
| ۸           |                 | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۰۰          |

ادامه جدول ۱-۵

| شماره ترکیب | R <sub>1</sub>  | R <sub>2</sub>                                       | pIC <sub>50</sub> (M) |
|-------------|---|--|-----------------------|
| ۹           |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۷۰                  |
| ۱۰          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۳۰                  |
| ۱۱          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۱۶                  |
| ۱۲          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۷۰                  |
| ۱۳          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۸۲                  |
| ۱۴          |   | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۳۰                  |
| ۱۵          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۰۰                  |
| ۱۶          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۰۰                  |
| ۱۷          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۵/۹۱                  |
| ۱۸          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۷۰                  |
| ۱۹          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۰۰                  |
| ۲۰          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۷۰                  |

ادامه جدول ۱-۵

| شماره ترکیب | R <sub>1</sub>  | R <sub>2</sub>                                       | pIC <sub>50</sub> (M) |
|-------------|---|--|-----------------------|
| ۲۱          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۶۴                  |
| ۲۲          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۵۲                  |
| ۲۳          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۹۲                  |
| ۲۴          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۳۰                  |
| ۲۵          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۶/۷۰                  |
| ۲۶          |    | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۳۰                  |
| ۲۷          |   | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۲۲                  |
| ۲۸          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۴۰                  |
| ۲۹          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۸/۱۰                  |
| ۳۰          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۸/۰۰                  |
| ۳۱          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۲۸                  |
| ۳۲          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۰۶                  |
| ۳۳          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۱۰                  |
| ۳۴          |  | -CH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-F) | ۷/۵۹                  |

## ۵-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

ابتدا ساختار مولکول‌ها در نرم افزار HyperChem رسم و ساختار هندسی آنها با احتساب اتم‌های هیدروژن و با استفاده از روش AM1 بهینه شد. بهینه‌سازی تا زمانی که گرادیان به ۰/۰۰۱ کیلو کالری بر مول برسد ادامه یافت.

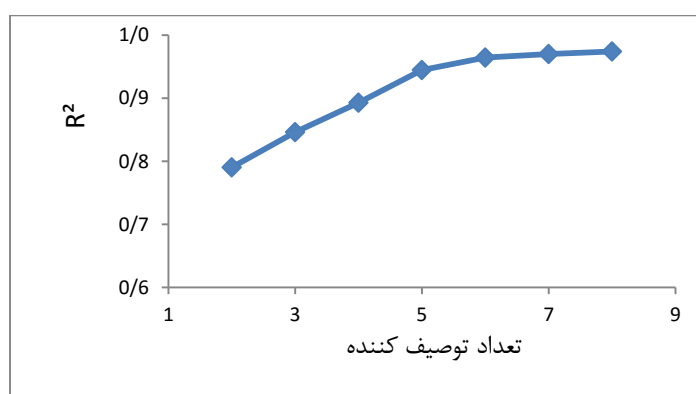
## ۵-۳- محاسبه و انتخاب بهترین توصیف‌کننده‌ها

برای محاسبه توصیف‌کننده‌ها، ساختارهای بهینه شده مولکول‌ها به عنوان ورودی به نرم افزار Dragon وارد شدند. محاسبه توصیف‌کننده‌ها توسط این نرم‌افزار صورت گرفت و ۱۴۸۱ توصیف‌کننده از هیجده گروه مختلف برای هر مولکول محاسبه گردید. سپس توصیف‌کننده‌های به دست آمده به عنوان متغیر مستقل و فعالیت بازداری ترکیبات به عنوان متغیر وابسته وارد نرم افزار SPSS شدند. به منظور انتخاب توصیف‌کننده‌های مناسب و جلوگیری از پیچیده شدن محاسبات، توصیف‌کننده‌هایی که با متغیر وابسته همبستگی بیشتری دارند از بین سایر آنها انتخاب شدند. بنابراین توصیف‌کننده‌هایی که دارای ۹۰٪ مقادیر یکسان برای همه ترکیبات بودند حذف شدند. از میان توصیف‌کننده‌هایی که ضریب همبستگی بزرگتر از ۰/۹ داشتند، توصیف‌کننده‌ای که همبستگی کمتری با متغیر وابسته داشت، حذف شد. سپس با استفاده از فهرست آنالیز، گزینه رگرسیون خطی و انتخاب روش مرحله‌ای چندین مدل مختلف به دست آمد. در بین مدل‌های ایجاد شده مدلی که تعداد توصیف‌کننده کمتر،  $R^2$  و آماره F آن بیشتر باشد، به عنوان مدل برتر انتخاب می‌شود. به منظور ارزیابی مدل انتخاب شده، داده‌ها به دو دسته‌ی آموزش (شامل ۲۶ ترکیب) و تست (۸ ترکیب) تقسیم می‌شود. سپس ضریب توصیف‌کننده‌های وارد شده در معادله با استفاده از سری آموزش و انتخاب روش Enter بدست می‌آید. جدول (۵-۲) کل توصیف‌کننده‌های انتخاب شده جهت ایجاد مدل MLR به همراه پارامترهای آماری را نشان می‌دهد.

جدول ۵-۲- توصیف‌کننده‌های انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله

| مدل | توصیف‌کننده‌ها  | R <sup>2</sup> | SE    | F       |
|-----|---|----------------|-------|---------|
| ۱   | BEHe2   | ۰/۵۴۵          | ۰/۵۵۷ | ۳۸/۳۰   |
| ۲   | BEHe2, ATS3e  | ۰/۷۹۰          | ۰/۳۸۴ | ۵۸/۵۳   |
| ۳   | BEHe2, ATS3e, RDF095v                                     | ۰/۸۴۶          | ۰/۳۳۴ | ۵۵/۲۰   |
| ۴   | BEHe2, ATS3e, RDF095v, G3u                                | ۰/۸۹۳          | ۰/۲۸۳ | ۶۰/۸۵۷  |
| ۵   | BEHe2, ATS3e, RDF095v, G3u, Mor17m                        | ۰/۹۴۵          | ۰/۲۰۷ | ۹۵/۶۴۶  |
| ۶   | BEHe2, ATS3e, RDF095v, G3u, Mor17m, Mor30u                | ۰/۹۶۴          | ۰/۱۷۰ | ۱۲۱/۱۲۰ |
| ۷   | BEHe2, ATS3e, RDF095v, G3u, Mor17m, Mor30u, Mor22u        | ۰/۹۷۰          | ۰/۱۶۰ | ۱۱۸/۴۸۷ |
| ۸   | BEHe2, ATS3e, RDF095v, G3u, Mor17m, Mor30u, Mor22u, BELv5 | ۰/۹۷۴          | ۰/۱۵۰ | ۱۲۰/۵۷۰ |

در شکل (۲-۵) تأثیر تعداد توصیف‌کننده بر مقدار R<sup>2</sup> نشان داده شده است. همانطور که در شکل مشاهده می‌شود از مدل ششم به بعد تغییر چندانی در R<sup>2</sup> مشاهده نمی‌شود و آماره F نیز کاهش می‌یابد. بنابراین مدل ششم به دلیل داشتن تعداد توصیف‌کننده کمتر، R<sup>2</sup> و F مناسب بر سایر مدل‌ها برتری دارد. توصیف‌کننده‌های وارد شده در این مدل به همراه ضریب رگرسیون و اثر متوسط آنها در جدول ۳-۵ آورده شده است.



شکل ۵-۲- تأثیر تعداد توصیف‌کننده‌ها بر مقادیر R<sup>2</sup> مدل‌های MLR

جدول ۵-۳- توصیف‌کننده‌های وارد شده در مدل MLR منتخب

| شماره | علامت   | ضریب رگرسیون | اثر متوسط | گروه               | نام کامل  |
|-------|---------|--------------|-----------|--------------------|---|
| ۱     | BEHe2   | ۲۳/۷۴۵       | ۱۳/۰۸۷    | BCUT               | highest eigenvalue n.2 of Burden matrix/weighted by atomic sanderson electronegativities                        |
| ۲     | ATS3e   | ۲۳/۸۷۰       | ۳/۶۱۸     | 2D autocorrelation | Broto-Moreau autocorrelation of a topological structure-lag 3/ weighted by atomic sanderson electronegativities |
| ۳     | RDF095v | ۰/۱۱۷        | ۰/۰۴۸     | RDF                | Radial Distribution Function -9.5/ weighted by atomic van der waals volumes RDF descriptors                     |
| ۴     | G3u     | ۲۰/۰۹۱       | ۰/۴۷۸     | WHIM               | 3st component symmetry directional WHIM index/unweighted  |
| ۵     | Mor17m  | -۰/۸۲۵       | -۰/۰۲۹۳   | 3D-MorSE           | 3D-MorSe-signal 17/weighted by atomic masses  |
| ۶     | Mor30u  | ۰/۷۱۹        | ۰/۰۳۶۳    | 3D-MorSE           | 3D-MorSe-signal 30/unweighted   |

جدول ۵-۴ ماتریس همبستگی توصیف‌کننده‌های وارد شده در مدل منتخب را نشان می‌دهد. همانطور که ملاحظه می‌شود مقدار همبستگی بین پارامترها قابل قبول (کمتر از ۰/۹) می‌باشد. بنابراین این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول ۵-۴- ماتریس همبستگی توصیف‌کننده‌های وارد شده در مدل منتخب

| توصیف‌کننده | BEHe2  | ATS3e  | RDF095v | G3u    | Mor17m | Mor30u |
|-------------|--------|--------|---------|--------|--------|--------|
| BEHe2       | ۱      |        |         |        |        |        |
| ATS3e       | -۰/۴۹۷ | ۱      |         |        |        |        |
| RDF095v     | ۰/۴۵۷  | -۰/۵۷۷ | ۱       |        |        |        |
| G3u         | -۰/۶۲۵ | ۰/۳۷۸  | -۰/۳۰۴  | ۱      |        |        |
| Mor17m      | -۰/۲۷۹ | ۰/۴۹۲  | -۰/۱۷۲  | ۰/۴۴۸  | ۱      |        |
| Mor30u      | ۰/۳۷۸  | -۰/۳۲۶ | ۰/۵۵۴   | -۰/۵۶۳ | -۰/۴۱۰ | ۱      |

معادله خطی حاصل از این مدل نیز به صورت زیر می‌باشد:

$$pIC_{50} = -112/209 + 23/745 \text{ BEHe2} + 23/870 \text{ ATS3e} + 0/117 \text{ RDF095v} + 20/091 \text{ G3u} \\ -0/825 \text{ Mor17m} + 0/719 \text{ Mor30u} \quad (\text{معادله ۵-۱})$$

## ۵-۴- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN)

در این مرحله از تحقیق برای پیش‌بینی  $pIC_{50}$  ترکیبات ۵، ۶- دی هیدروکسی پیریمیدین ۴- کربوکسامیدها از مدل غیر خطی نیز استفاده شده است. برای ایجاد مدل غیرخطی که ارتباط مناسبی میان ساختار و  $pIC_{50}$  ترکیبات مورد نظر فراهم کند، از شبکه عصبی مصنوعی استفاده شد. سری داده‌ها به صورت تصادفی به دو سری آموزش (۲۶ ترکیب) و تست (۸ ترکیب) تقسیم شدند. سپس پارامترهای مؤثر بر آموزش موفق شبکه عصبی پس انتشار از قبیل تعداد متغیرهای ورودی، تعداد گره‌های لایه پنهان، نوع تابع انتقال، نوع تابع آموزش، تعداد دوره‌های آموزش و پارامتر  $\mu$  با استفاده از سری آموزش و با روش ارزیابی تقاطعی بهینه شدند. در این روش تابع کارایی شبکه، میانگین مربع خطای سری آموزش می‌باشد.

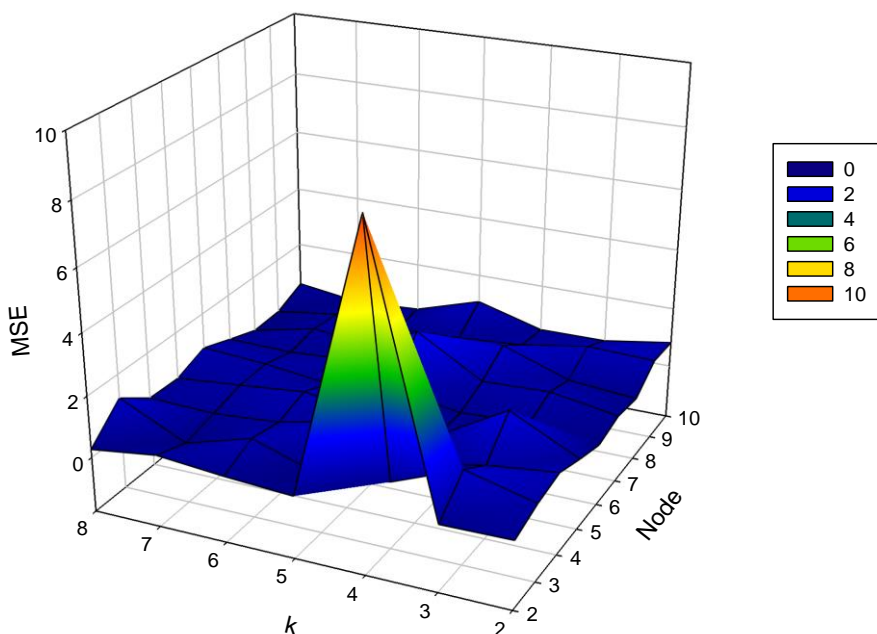
## ۵-۴-۱- بهینه‌سازی تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه و تعداد

### نرون‌های لایه مخفی

بر این اساس شبکه توسط دو الگوریتم آموزشی لونیگ-مارکوارت و تنظیم بایزین با ۲ تا ۸ متغیر ورودی، تعداد گره‌های از ۲ تا ۱۰ در لایه پنهان و همچنین توابع لگاریتم سیگموئیدی ( $\log\text{sig}$ )، تانژانت سیگموئیدی ( $\text{tansig}$ ) و تابع انتقال خطی ( $\text{purelin}$ )، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمامی این شبکه‌ها تعداد دوره‌های آموزش یکسان و برابر ۲۰ در نظر گرفته شد. همچنین از تابع انتقال خطی ( $\text{purelin}$ ) در لایه خروجی استفاده شد. همانطور که در بخش قبل گفته شد بهینه‌سازی از طریق ارزیابی تقاطعی بر روی سری آموزش انجام گرفت. معیار نیز به حداقل رساندن میانگین مربع خطای سری آموزش در نظر گرفته شد که نتایج این محاسبات در جداول (۵-۵)، (۵-۶)، (۶-۷)، (۷-۸)، (۸-۹)، (۹-۱۰) آورده شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۵-۳)، (۵-۴)، (۵-۵)، (۵-۶)، (۵-۷)، (۵-۸) ارائه شده است.

جدول ۵-۵ - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |
|--|----|--------|--------|--------|--------|--------|--------|--------|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |
| تعداد گره‌ها                                     | ۲  | ۰/۴۰۸۴ | ۰/۳۷۲۴ | ۹/۰۳۸۳ | ۰/۲۳۲۷ | ۰/۳۶۶۶ | ۰/۵۵۹۴ | ۰/۲۶۸۰ |
|  | ۳  | ۰/۷۰۹۴ | ۱/۲۶۵۷ | ۰/۳۷۹۶ | ۰/۰۹۶۲ | ۰/۸۱۳۳ | ۰/۱۸۵۵ | ۱/۲۱۷۵ |
|  | ۴  | ۰/۸۷۳۱ | ۰/۶۲۰۱ | ۰/۱۴۲۸ | ۱/۰۸۶۸ | ۰/۵۹۴۰ | ۰/۴۲۲۱ | ۰/۴۹۵۰ |
|  | ۵  | ۰/۵۱۸۶ | ۱/۶۲۱۴ | ۰/۳۵۰۶ | ۰/۵۳۱۶ | ۰/۸۸۲۶ | ۰/۶۵۷۳ | ۰/۴۵۶۹ |
|  | ۶  | ۱/۲۰۷۳ | ۰/۴۴۲۶ | ۰/۳۵۴۲ | ۰/۶۷۲۷ | ۰/۴۲۹۲ | ۰/۶۸۴۶ | ۰/۸۲۰۹ |
|  | ۷  | ۰/۴۱۶۴ | ۰/۷۷۹۳ | ۰/۶۹۱۰ | ۱/۹۷۰۸ | ۰/۵۵۹۹ | ۰/۴۱۳۸ | ۰/۴۵۳۲ |
|  | ۸  | ۰/۲۷۴۲ | ۰/۴۶۸۶ | ۱/۲۳۳۹ | ۰/۹۸۱۲ | ۰/۲۸۸۸ | ۰/۵۸۰۱ | ۰/۱۰۶۰ |
|  | ۹  | ۰/۸۷۲۵ | ۰/۹۴۰۵ | ۰/۵۷۹۳ | ۰/۴۰۸۶ | ۰/۳۲۱۶ | ۰/۴۴۴۸ | ۰/۱۴۵۷ |
|  | ۱۰ | ۰/۸۲۲۱ | ۰/۴۷۱۳ | ۰/۴۵۴۱ | ۱/۰۵۹۴ | ۰/۳۸۹۴ | ۰/۲۵۲۸ | ۰/۵۴۳۲ |

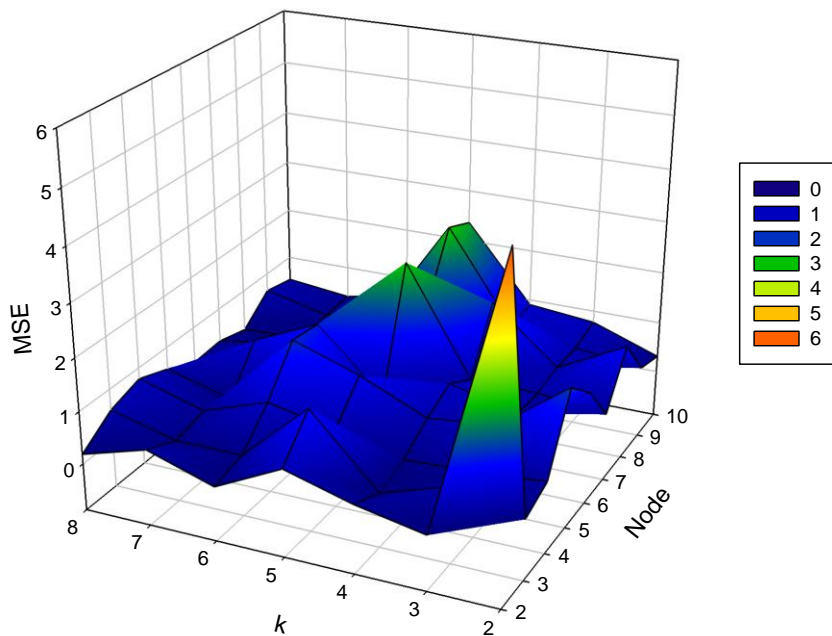


شکل ۵-۳- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی



جدول ۵-۶ - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

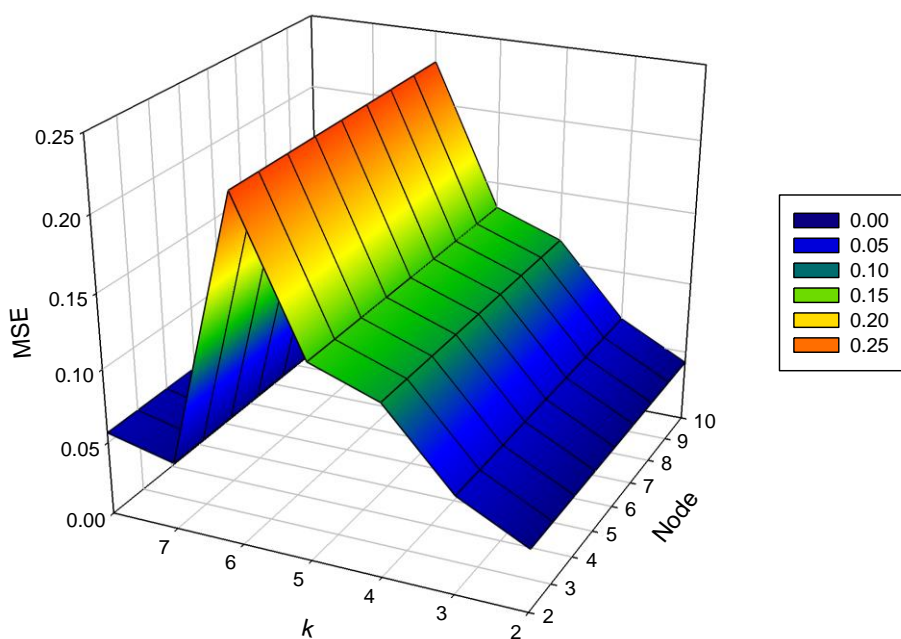
| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |
|--|----|--------|--------|--------|--------|--------|--------|--------|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |
| تعداد گره‌ها                                     | ۲  | ۵/۳۱۱۷ | ۰/۱۴۶۹ | ۰/۴۰۳۰ | ۰/۷۶۵۲ | ۰/۱۳۹۳ | ۰/۵۵۱۲ | ۰/۱۸۴۰ |
|  | ۳  | ۰/۲۵۲۶ | ۰/۵۶۰۶ | ۰/۱۳۶۸ | ۱/۴۲۹۰ | ۰/۲۷۳۸ | ۰/۲۷۷۴ | ۰/۵۹۹۷ |
|  | ۴  | ۰/۴۶۶۹ | ۰/۳۷۰۴ | ۰/۲۹۵۴ | ۰/۵۸۷۹ | ۰/۸۵۲۶ | ۰/۴۲۲۲ | ۰/۷۸۹۷ |
|  | ۵  | ۱/۷۵۳۵ | ۰/۷۹۴۳ | ۰/۶۵۱۲ | ۱/۲۱۱۷ | ۱/۶۳۶۱ | ۰/۲۱۸۵ | ۰/۵۷۱۴ |
|  | ۶  | ۱/۲۴۶۳ | ۰/۵۷۹۷ | ۰/۸۸۹۳ | ۰/۳۷۶۵ | ۱/۵۵۵۳ | ۱/۰۰۱۸ | ۰/۳۰۳۵ |
|  | ۷  | ۰/۳۹۷۲ | ۰/۵۴۵۹ | ۰/۵۶۶۰ | ۲/۶۰۶۳ | ۰/۴۰۶۶ | ۰/۵۶۵۹ | ۰/۳۴۴۶ |
|  | ۸  | ۱/۳۰۸۱ | ۰/۶۴۰۸ | ۱/۶۹۲۸ | ۱/۴۲۴۲ | ۰/۴۷۶۷ | ۰/۲۵۱۷ | ۰/۱۹۱۰ |
|  | ۹  | ۰/۴۸۳۸ | ۰/۶۳۴۸ | ۰/۸۴۳۵ | ۲/۶۰۱۱ | ۰/۷۴۵۳ | ۰/۳۵۶۸ | ۰/۶۴۳۶ |
|  | ۱۰ | ۰/۳۱۳۵ | ۰/۷۷۱۹ | ۰/۹۰۲۶ | ۲/۳۶۴۹ | ۰/۶۲۶۸ | ۰/۳۷۸۸ | ۰/۵۲۳۰ |



شکل ۵-۴ - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی

جدول ۵-۷ - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال خطی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

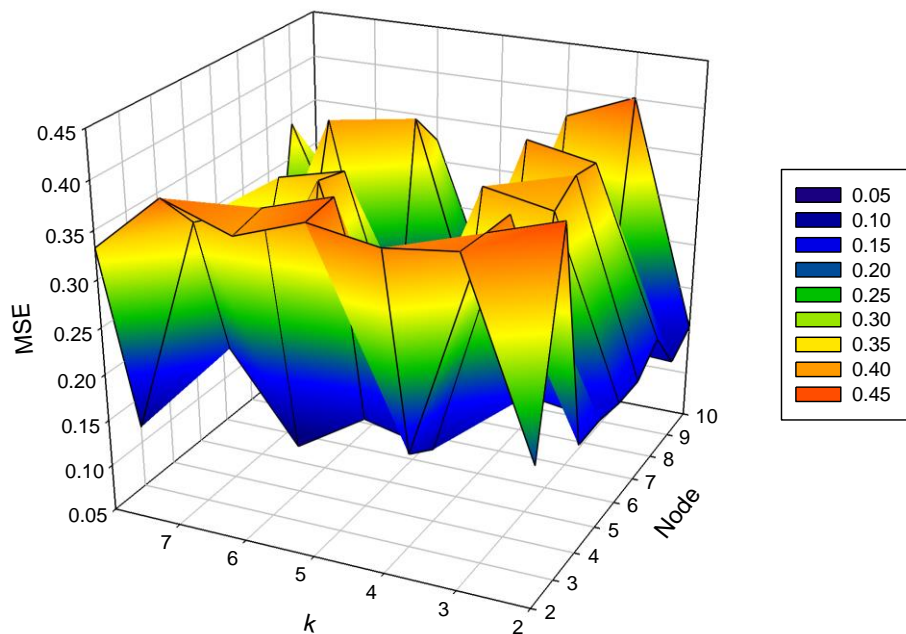
| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |  |
|--|----|--------|--------|--------|--------|--------|--------|--------|--|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |  |
| تعداد گره‌ها                                     | ۲  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۳  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۴  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۵  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۶  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۷  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۸  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۹  | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |
|  | ۱۰ | ۰/۲۲۹۰ | ۰/۱۳۲۳ | ۰/۱۱۵۹ | ۰/۰۶۶۱ | ۰/۰۴۲۳ | ۰/۰۵۶۸ | ۰/۰۴۵۸ |  |



شکل ۵-۵- نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال خطی

جدول ۵-۸- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

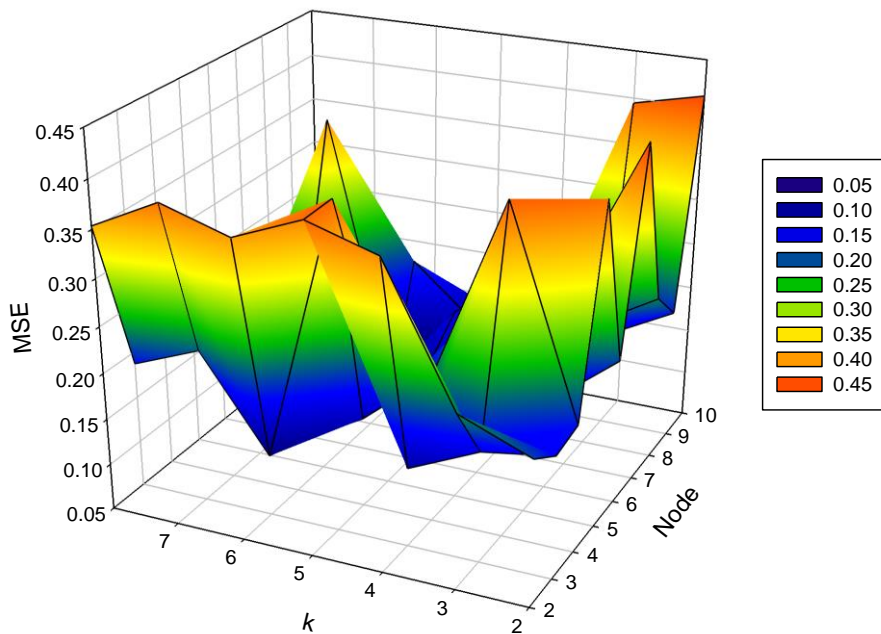
| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |
|--|----|--------|--------|--------|--------|--------|--------|--------|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |
| تعداد گره‌ها                                     | ۲  | ۰/۱۹۹۵ | ۰/۳۹۱۳ | ۰/۳۸۲۳ | ۰/۳۹۴۱ | ۰/۳۶۸۷ | ۰/۳۹۳۸ | ۰/۳۳۲۴ |
|  | ۳  | ۰/۴۱۲۲ | ۰/۳۸۸۸ | ۰/۱۵۱۷ | ۰/۳۹۹۴ | ۰/۳۷۶۹ | ۰/۳۵۱۰ | ۰/۱۱۷۲ |
|  | ۴  | ۰/۱۶۶۸ | ۰/۳۸۷۲ | ۰/۱۳۰۷ | ۰/۱۴۸۲ | ۰/۱۰۱۵ | ۰/۱۹۶۳ | ۰/۱۹۶۷ |
|  | ۵  | ۰/۱۶۸۶ | ۰/۱۷۹۹ | ۰/۱۷۳۵ | ۰/۱۴۱۵ | ۰/۳۶۷۶ | ۰/۲۲۳۴ | ۰/۲۹۸۴ |
|  | ۶  | ۰/۱۶۲۳ | ۰/۳۵۲۳ | ۰/۳۶۶۱ | ۰/۱۳۷۷ | ۰/۳۵۹۴ | ۰/۳۴۱۸ | ۰/۱۵۸۴ |
|  | ۷  | ۰/۱۶۲۰ | ۰/۳۷۰۶ | ۰/۱۴۳۳ | ۰/۱۳۹۳ | ۰/۰۷۸۱ | ۰/۲۲۰۰ | ۰/۱۵۲۱ |
|  | ۸  | ۰/۱۸۶۳ | ۰/۳۶۵۵ | ۰/۳۷۹۵ | ۰/۱۳۹۴ | ۰/۰۷۲۲ | ۰/۳۶۹۵ | ۰/۱۲۱۹ |
|  | ۹  | ۰/۱۳۷۲ | ۰/۱۴۱۶ | ۰/۰۹۳۹ | ۰/۱۳۵۸ | ۰/۳۶۴۷ | ۰/۱۸۰۲ | ۰/۳۳۸۶ |
|  | ۱۰ | ۰/۱۵۴۷ | ۰/۴۰۱۶ | ۰/۳۷۱۸ | ۰/۱۳۶۳ | ۰/۳۲۴۰ | ۰/۲۰۰۱ | ۰/۱۴۶۹ |



شکل ۵-۶- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول ۵-۹- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

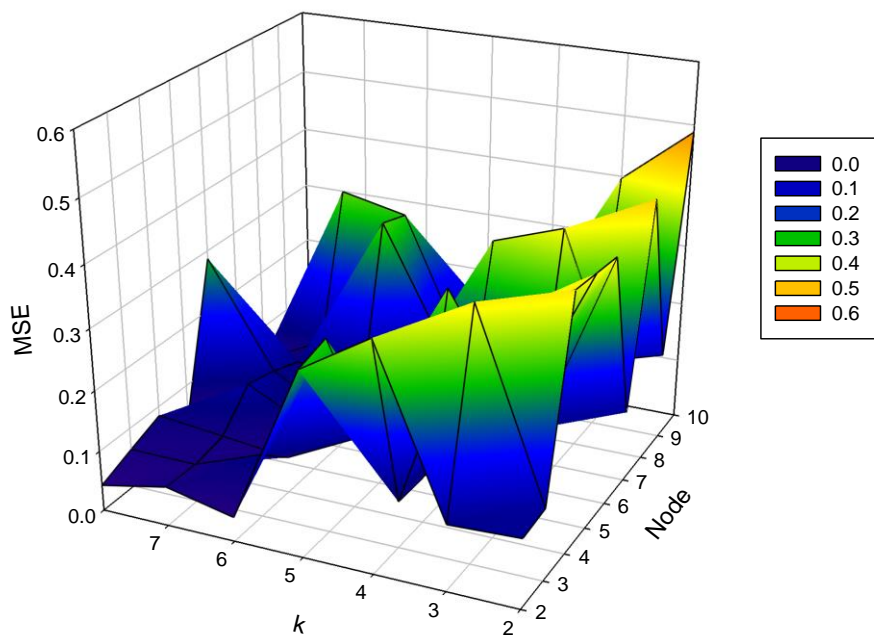
| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |
|--|----|--------|--------|--------|--------|--------|--------|--------|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |
| تعداد گره‌ها                                     | ۲  | ۰/۲۰۵۱ | ۰/۲۳۴۷ | ۰/۳۷۳۸ | ۰/۳۹۵۷ | ۰/۳۶۶۵ | ۰/۳۸۸۶ | ۰/۳۵۳۵ |
|  | ۳  | ۰/۱۸۱۸ | ۰/۱۶۹۶ | ۰/۱۳۵۳ | ۰/۳۹۷۰ | ۰/۱۱۶۰ | ۰/۲۱۶۰ | ۰/۱۸۷۲ |
|  | ۴  | ۰/۱۸۸۰ | ۰/۴۰۱۱ | ۰/۲۰۹۶ | ۰/۱۴۷۴ | ۰/۱۰۶۳ | ۰/۲۶۵۰ | ۰/۱۸۰۱ |
|  | ۵  | ۰/۳۹۴۲ | ۰/۲۳۷۸ | ۰/۲۱۴۱ | ۰/۱۴۴۰ | ۰/۰۹۲۴ | ۰/۲۰۸۸ | ۰/۱۶۲۱ |
|  | ۶  | ۰/۲۰۷۳ | ۰/۱۴۶۴ | ۰/۲۳۳۶ | ۰/۱۳۸۹ | ۰/۰۸۳۶ | ۰/۱۹۱۳ | ۰/۱۲۹۲ |
|  | ۷  | ۰/۴۱۴۷ | ۰/۲۵۶۷ | ۰/۰۹۴۴ | ۰/۱۴۶۷ | ۰/۰۸۴۰ | ۰/۱۹۵۵ | ۰/۱۴۲۱ |
|  | ۸  | ۰/۲۲۰۲ | ۰/۱۸۶۴ | ۰/۲۰۲۱ | ۰/۱۷۵۶ | ۰/۰۷۷۶ | ۰/۳۶۸۶ | ۰/۱۳۴۵ |
|  | ۹  | ۰/۱۹۲۴ | ۰/۱۵۲۷ | ۰/۱۷۴۸ | ۰/۱۳۸۵ | ۰/۲۰۰۲ | ۰/۱۸۱۶ | ۰/۱۴۸۰ |
|  | ۱۰ | ۰/۴۱۲۰ | ۰/۳۹۴۸ | ۰/۰۹۹۵ | ۰/۱۳۳۶ | ۰/۰۷۱۴ | ۰/۱۷۵۶ | ۰/۱۱۶۷ |



شکل ۵-۷- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی

جدول ۵-۱۰- مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال خطی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

| تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K |    |        |        |        |        |        |        |        |
|--|----|--------|--------|--------|--------|--------|--------|--------|
|  | ۲  | ۳      | ۴      | ۵      | ۶      | ۷      | ۸      |        |
| تعداد گره‌ها                                     | ۲  | ۰/۱۱۳۵ | ۰/۱۰۸۵ | ۰/۳۷۰۱ | ۰/۳۰۱۳ | ۰/۰۴۰۲ | ۰/۰۶۵۱ | ۰/۰۴۲۶ |
|  | ۳  | ۰/۱۱۸۲ | ۰/۴۱۰۸ | ۰/۰۷۶۶ | ۰/۳۱۴۶ | ۰/۱۰۴۰ | ۰/۰۶۲۸ | ۰/۰۶۴۱ |
|  | ۴  | ۰/۴۱۶۳ | ۰/۳۵۷۵ | ۰/۰۷۵۱ | ۰/۱۱۵۶ | ۰/۰۵۹۸ | ۰/۰۷۲۱ | ۰/۰۸۴۳ |
|  | ۵  | ۰/۴۰۹۷ | ۰/۱۶۹۸ | ۰/۳۴۷۴ | ۰/۱۲۷۲ | ۰/۰۶۸۳ | ۰/۱۲۵۲ | ۰/۰۴۳۴ |
|  | ۶  | ۰/۴۰۳۶ | ۰/۱۰۴۴ | ۰/۱۱۵۳ | ۰/۱۴۰۱ | ۰/۰۴۸۸ | ۰/۱۱۲۱ | ۰/۲۸۶۹ |
|  | ۷  | ۰/۱۱۷۴ | ۰/۳۹۹۴ | ۰/۳۶۱۳ | ۰/۱۰۲۵ | ۰/۰۴۵۸ | ۰/۰۹۳۹ | ۰/۰۴۸۱ |
|  | ۸  | ۰/۴۳۵۰ | ۰/۱۵۰۵ | ۰/۰۷۴۴ | ۰/۱۲۱۱ | ۰/۳۲۴۹ | ۰/۰۹۴۷ | ۰/۰۴۲۰ |
|  | ۹  | ۰/۱۴۴۲ | ۰/۱۱۰۰ | ۰/۱۷۶۹ | ۰/۰۷۷۱ | ۰/۳۰۹۸ | ۰/۳۳۳۳ | ۰/۰۵۵۶ |
|  | ۱۰ | ۰/۴۸۸۵ | ۰/۳۹۶۶ | ۰/۰۹۱۱ | ۰/۱۲۷۷ | ۰/۰۴۷۷ | ۰/۰۸۱۶ | ۰/۰۴۹۸ |



شکل ۵-۸- نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال خطی

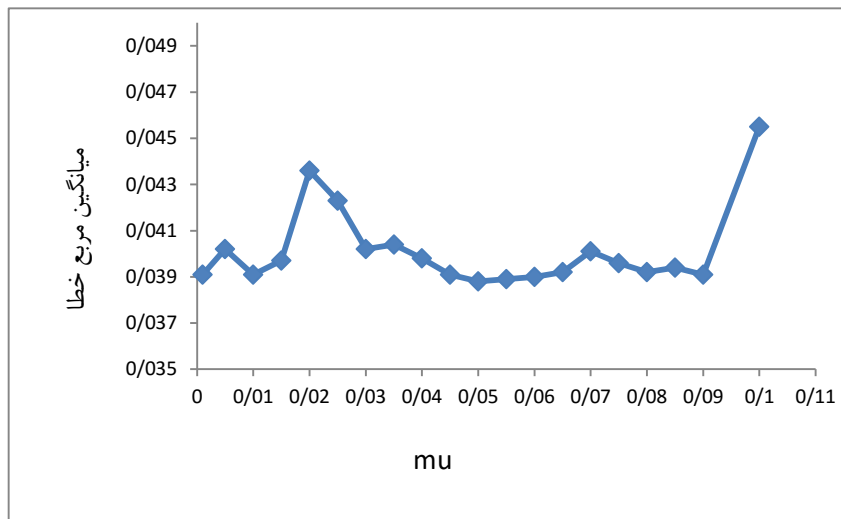
با توجه به نتایج به دست آمده، الگوریتم آموزشی تنظیم بایزین نسبت به الگوریتم آموزشی لونبرگ مارکوارت دارای میانگین مربعات خطای (MSE) کمتری است. علاوه بر آن دارای قدرت تعمیم‌پذیری بالاتری است بدین معنی که قادر است میان ترکیباتی که در سری آموزش استفاده شده و آن‌هایی که در سری آموزش استفاده نشده‌اند تطابق خوبی برقرار کند. همچنین تابع انتقال خطی با ۶ متغیر ورودی و تعداد دو گره در لایه پنهان دارای خطای کمتری نسبت به لگاریتم سیگموئیدی و تانژانت سیگموئیدی است. بنابراین تابع انتقال خطی به عنوان تابع انتقال لایه پنهان استفاده شد. مقدار  $\mu$  نیز در این شرایط برابر با مقدار پیش فرض آن در تابع آموزشی تنظیم بایزین یعنی ۰/۰۰۵ است

#### ۵-۴-۲- بهینه سازی پارامتر $\mu$

در این مرحله برای بهینه سازی پارامتر  $\mu$  ساختار شبکه با ۶ متغیر ورودی، ۲ گره در لایه پنهان، الگوریتم آموزشی تنظیم بایزین، تابع انتقال خطی (purelin) و تعداد دوره‌های آموزشی نیز ثابت و برابر با ۲۰ در نظر گرفته شد. سپس مقدار پارامتر  $\mu$  بین ۰/۰۰۱ تا ۰/۱ تغییر داده شد و مقدار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش محاسبه شد. سرانجام مقدار میانگین مربع خطا بر حسب  $\mu$  ترسیم گردید. نقطه‌ای که کمترین خطا را برای سری آموزش داشت، به عنوان مقدار بهینه انتخاب شد که طبق جدول (۵-۱۱) و شکل (۵-۹) مقدار  $\mu$  بهینه ۰/۰۵ به دست آمد.

جدول ۵-۱۱ - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت  $\mu$

| میانگین مربع خطا | $\mu$ مقادیر | میانگین مربع خطا | $\mu$ مقادیر |
|------------------|--------------|------------------|--------------|
| ۰/۰۳۸۸           | ۰/۰۵۰        | ۰/۰۳۹۱           | ۰/۰۰۱        |
| ۰/۰۳۸۹           | ۰/۰۵۵        | ۰/۰۴۰۲           | ۰/۰۰۵        |
| ۰/۰۳۹۰           | ۰/۰۶۰        | ۰/۰۳۹۱           | ۰/۰۱         |
| ۰/۰۳۹۲           | ۰/۰۶۵        | ۰/۰۳۹۷           | ۰/۰۱۵        |
| ۰/۰۴۰۱           | ۰/۰۷۰        | ۰/۰۴۳۶           | ۰/۰۲         |
| ۰/۰۳۹۶           | ۰/۰۷۵        | ۰/۰۴۲۳           | ۰/۰۲۵        |
| ۰/۰۳۹۲           | ۰/۰۸۰        | ۰/۰۴۰۲           | ۰/۰۳         |
| ۰/۰۳۹۴           | ۰/۰۸۵        | ۰/۰۴۰۴           | ۰/۰۳۵        |
| ۰/۰۳۹۱           | ۰/۰۹۰        | ۰/۰۳۹۸           | ۰/۰۴۰        |
| ۰/۰۴۵۵           | ۰/۱          | ۰/۰۳۹۱           | ۰/۰۴۵        |



شکل ۵-۹- نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر  $\mu$

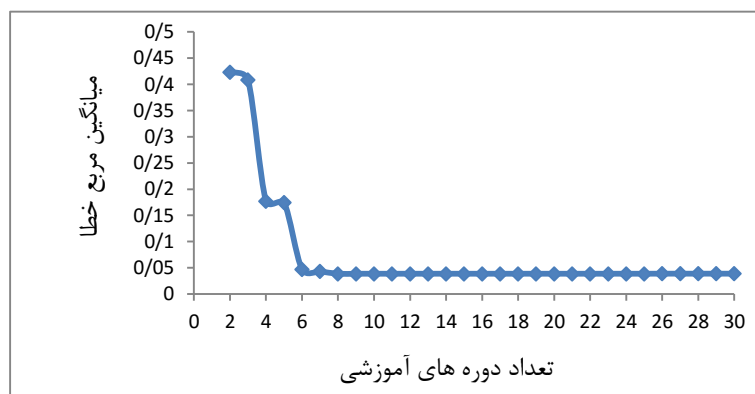
#### ۵-۴-۳- بهینه سازی تعداد دوره‌های آموزشی

تعداد دوره‌های آموزشی پارامتری است که برای جلوگیری از آموزش بیش از حد شبکه بهینه‌سازی می‌شود. برای این منظور، ابتدا مقدار بهینه تعداد نرون‌ها (۲) و مقدار بهینه  $\mu$  (۰/۰۵) در شبکه مورد نظر با تابع آموزش تنظیم بایزین و تابع انتقال خطی (purelin) قرار داده شد. تعداد دوره آموزش تغییر داده شد و برای هر دوره میانگین مربع خطاهای سری آموزش به روش ارزیابی تقاطعی محاسبه شد (جدول ۵-۱۲). سپس منحنی تعداد چرخه‌های آموزش بر حسب میانگین مربع خطاها بر طبق

شکل (۵-۱۰) رسم گردید. مقدار بهینه تعداد تکرار جایی است که میزان این خطا حداقل باشد. جدول (۵-۱۲) و شکل (۵-۱۰) نشان می‌دهد که با افزایش تعداد دوره‌های آموزش میانگین مربع خطا (MSE) کاهش می‌یابد تا جایی که در ۸ دور تکرار به حداقل خطا می‌رسد و پس از آن افزایش و یا کاهش چشمگیری در خطا مشاهده نمی‌شود و خطا تقریباً ثابت است، بنابراین ۸ بار تکرار به عنوان تعداد دوره‌های بهینه برای یادگیری شبکه انتخاب شد.

جدول ۵-۱۲- مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف

| تعداد دوره‌های آموزشی | میانگین مربع خطا | تعداد دوره‌های آموزشی | میانگین مربع خطا |
|-----------------------|------------------|-----------------------|------------------|
| ۲                     | ۰/۴۲۳۱           | ۱۷                    | ۰/۰۳۸۸           |
| ۳                     | ۰/۴۰۸۵           | ۱۸                    | ۰/۰۳۸۸           |
| ۴                     | ۰/۱۷۷۲           | ۱۹                    | ۰/۰۳۸۸           |
| ۵                     | ۰/۱۷۴۷           | ۲۰                    | ۰/۰۳۸۸           |
| ۶                     | ۰/۰۴۷۳           | ۲۱                    | ۰/۰۳۸۸           |
| ۷                     | ۰/۰۴۳۲           | ۲۲                    | ۰/۰۳۸۸           |
| ۸                     | ۰/۰۳۸۸           | ۲۳                    | ۰/۰۳۸۸           |
| ۹                     | ۰/۰۳۸۸           | ۲۴                    | ۰/۰۳۸۸           |
| ۱۰                    | ۰/۰۳۸۸           | ۲۵                    | ۰/۰۳۸۸           |
| ۱۱                    | ۰/۰۳۸۸           | ۲۶                    | ۰/۰۳۸۹           |
| ۱۲                    | ۰/۰۳۸۸           | ۲۷                    | ۰/۰۳۸۹           |
| ۱۳                    | ۰/۰۳۸۸           | ۲۸                    | ۰/۰۳۸۹           |
| ۱۴                    | ۰/۰۳۸۸           | ۲۹                    | ۰/۰۳۸۹           |
| ۱۵                    | ۰/۰۳۸۸           | ۳۰                    | ۰/۰۳۸۹           |
| ۱۶                    | ۰/۰۳۸۸           |                       |                  |

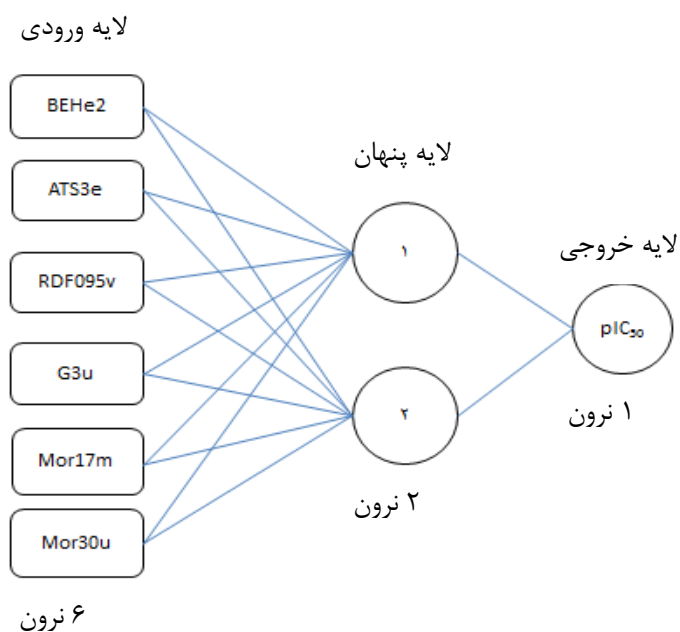


شکل ۵-۱۰- منحنی مقدار میانگین خطاها در برابر تغییر تعداد دوره‌های آموزشی



#### ۵-۴-۴- ساختار شبکه عصبی مصنوعی بهینه شده

با توجه به روند بهینه سازی که در قسمت‌های قبلی ذکر شد، شبکه عصبی مصنوعی بهینه شده دارای ۶ توصیف‌کننده، ۲ گره در لایه پنهان، تابع آموزشی تنظیم بایزین و تابع انتقال خطی است. شکل (۵-۱۱) ساختار شبکه عصبی بهینه شده را نشان می‌دهد.



شکل ۵-۱۱- تصویر شماتیک ساختار شبکه عصبی مصنوعی به دست آمده پس از بهینه سازی

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۵-۱۳) ارائه شده است.

جدول ۵-۱۳ - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

| Trainbr  | تابع آموزش             |
|----------|------------------------|
| pureline | تابع انتقال لایه پنهان |
| pureline | تابع انتقال لایه خروجی |
| ۲        | تعداد نرون لایه پنهان  |
| ۶        | تعداد متغیرهای ورودی   |
| ۸        | تعداد دوره‌های آموزش   |
| ۰/۰۵     | پارامتر mu             |

## ۵-۵- ارزیابی مدل‌های خطی و غیرخطی

### ۵-۵-۱- ارزیابی مدل‌های ANN و MLR با استفاده از سری تست

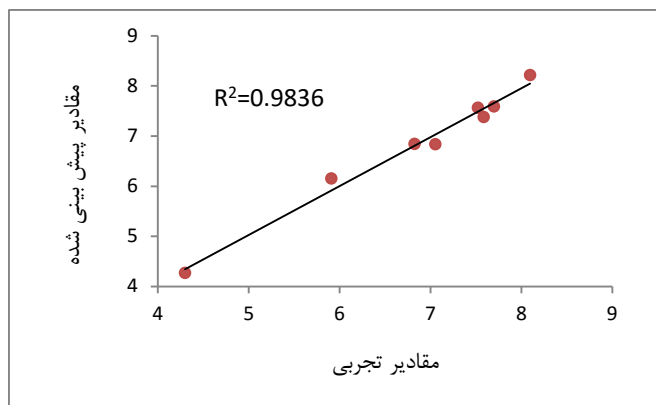
به منظور ارزیابی مدل‌های منتخب خطی و غیرخطی به دست آمده، سری تست مورد استفاده قرار گرفت. بدین ترتیب که در مدل خطی با استفاده از ضرایب مدل برتر سری آموزش و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر  $pIC_{50}$  مربوط به ۸ ترکیب که در مدل‌سازی استفاده نشده بودند، پیش‌بینی شد. نتایج حاصله در جداول ۵-۱۴، ۵-۱۵ برای سری تست آورده شده‌است.  $R^2$  حاصله در شکل ۵-۱۲ توانایی هر دو مدل خطی و غیرخطی را در پیش‌بینی فعالیت دارویی این ترکیبات نشان می‌دهد.

جدول ۵-۱۴- مقادیر تجربی و پیش‌بینی شده‌ی  $pIC_{50}$  به وسیله‌ی روش MLR برای سری تست

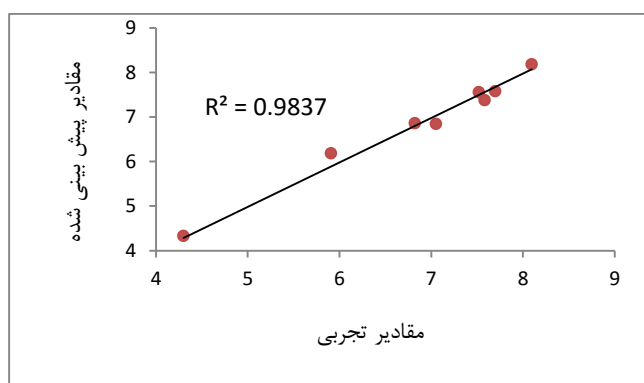
| شماره ترکیب | $pIC_{50}(M)$ |                    | درصد خطای نسبی |
|-------------|---------------|--------------------|----------------|
|             | مقدار تجربی   | مقدار پیش‌بینی شده |                |
| ۴           | ۳۰/۴          | ۴/۲۶               | -۰/۹۳          |
| ۹           | ۷۰/۷          | ۷/۵۹               | -۱/۴۳          |
| ۱۳          | ۸۲/۶          | ۶/۸۴               | ۰/۲۹           |
| ۱۷          | ۹۱/۵          | ۶/۱۵               | ۴/۰۶           |
| ۲۲          | ۵۲/۷          | ۷/۵۶               | ۰/۵۳           |
| ۲۹          | ۱۰/۸          | ۸/۲۱               | ۱/۳۶           |
| ۳۲          | ۰۶/۷          | ۶/۸۳               | -۳/۲۶          |
| ۳۴          | ۵۹/۷          | ۷/۳۸               | -۲/۷۷          |

جدول ۵-۱۵- مقادیر تجربی و پیش‌بینی شده‌ی  $pIC_{50}$  به وسیله‌ی روش ANN برای سری تست

| شماره ترکیب | $pIC_{50}(M)$ |                    | درصد خطای نسبی |
|-------------|---------------|--------------------|----------------|
|             | مقدار تجربی   | مقدار پیش‌بینی شده |                |
| ۴           | ۳۰/۴          | ۳۳/۴               | ۷۰/۰           |
| ۹           | ۷۰/۷          | ۵۷/۷               | -۱/۶۹          |
| ۱۳          | ۸۲/۶          | ۸۶/۶               | ۵۹/۰           |
| ۱۷          | ۹۱/۵          | ۱۸/۶               | ۵۷/۴           |
| ۲۲          | ۵۲/۷          | ۵۵/۷               | ۴۰/۰           |
| ۲۹          | ۱۰/۸          | ۱۸/۸               | ۹۹/۰           |
| ۳۲          | ۰۶/۷          | ۸۴/۶               | -۳/۱۲          |
| ۳۴          | ۵۹/۷          | ۳۷/۷               | -۲/۹۰          |



(الف)



(ب)

شکل (۵-۱۲) - نمودار مقادیر پیش بینی شده ی  $pIC_{50}$  بر حسب مقادیر تجربی برای سری تست توسط الف) روش MLR ، ب) روش ANN

#### ۵-۵-۲- ارزیابی مدل های MLR و ANN به روش حذف مرحله ای تک تک داده ها

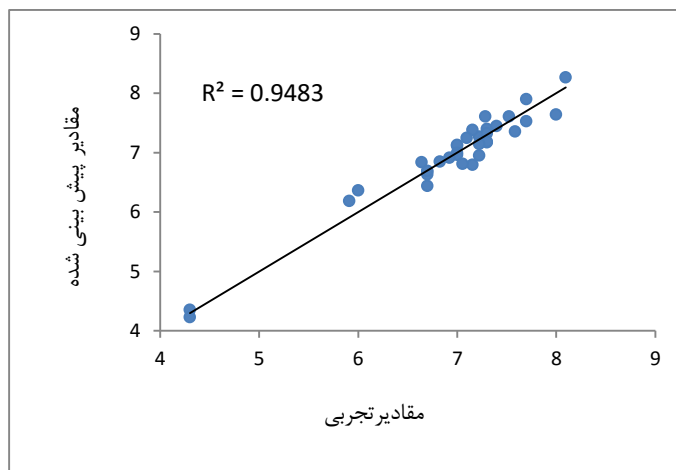
برای ارزیابی مدل به دست آمده علاوه بر استفاده از سری تست، روش حذف مرحله ای تک تک داده ها نیز به کار گرفته شد. در این روش هر بار یکی از ترکیبات از سری داده ها حذف و سپس مدل - سازی با ۳۳ ترکیب باقیمانده صورت گرفت. این مدل برای پیش بینی فعالیت دارویی ترکیب حذف شده به کار گرفته شد. سپس این فرایند برای تمام ترکیبات سری داده ها تکرار شد نتایج حاصله برای دو روش MLR و ANN در جداول ۵-۱۶ و ۵-۱۷ آورده شده است. شکل ۵-۱۳ نیز نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی برای هر دو روش نشان می دهد.

جدول ۵-۱۶- ارزیابی مدل MLR به روش حذف مرحله‌ای تک تک داده‌ها

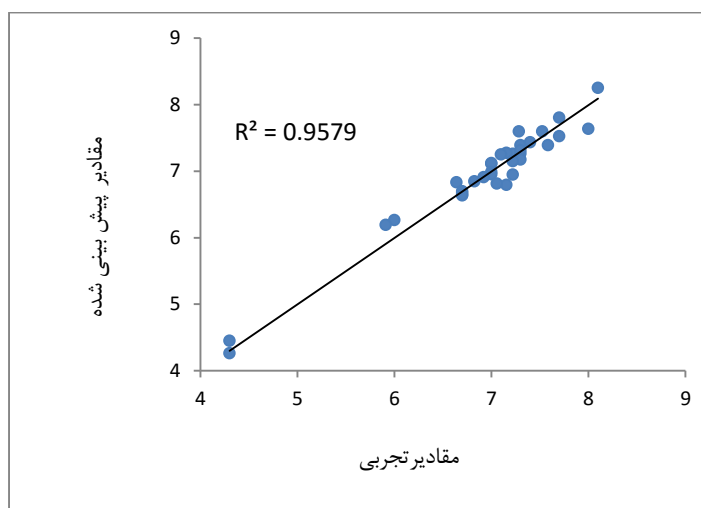
| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۱           | ۱۶/۷        | ۸۰/۶               | -۵/۰۳          |
| ۲           | ۰۰/۶        | ۳۶/۶               | ۶/۰۰           |
| ۳           | ۳۰/۴        | ۳۵/۴               | ۱/۱۶           |
| ۴           | ۳۰/۴        | ۲۳/۴               | -۱/۶۳          |
| ۵           | ۷۰/۷        | ۹۰/۷               | ۲/۶۰           |
| ۶           | ۲۲/۷        | ۹۵/۶               | -۳/۷۴          |
| ۷           | ۲۲/۷        | ۲۷/۷               | ۰/۶۹           |
| ۸           | ۰۰/۷        | ۱۲/۷               | ۱/۷۱           |
| ۹           | ۷۰/۷        | ۵۳/۷               | -۲/۲۱          |
| ۱۰          | ۳۰/۷        | ۴۰/۷               | ۱/۳۷           |
| ۱۱          | ۱۶/۷        | ۳۸/۷               | ۳/۰۷           |
| ۱۲          | ۷۰/۶        | ۶۷/۶               | -۰/۴۵          |
| ۱۳          | ۸۲/۶        | ۸۵/۶               | ۰/۴۴           |
| ۱۴          | ۳۰/۷        | ۱۷/۷               | -۱/۷۸          |
| ۱۵          | ۰۰/۷        | ۱۳/۷               | ۱/۸۶           |
| ۱۶          | ۰۰/۷        | ۹۹/۶               | -۰/۱۴          |
| ۱۷          | ۹۱/۵        | ۱۹/۶               | ۴/۷۴           |
| ۱۸          | ۷۰/۶        | ۶۳/۶               | -۱/۰۴          |
| ۱۹          | ۰۰/۷        | ۹۶/۶               | -۰/۵۷          |
| ۲۰          | ۷۰/۶        | ۶۹/۶               | -۰/۱۵          |
| ۲۱          | ۶۴/۶        | ۸۴/۶               | ۳/۰۱           |
| ۲۲          | ۵۲/۷        | ۶۱/۷               | ۱/۲۰           |
| ۲۳          | ۹۲/۶        | ۹۱/۶               | -۰/۱۴          |
| ۲۴          | ۳۰/۷        | ۳۲/۷               | ۰/۲۷           |
| ۲۵          | ۷۰/۶        | ۴۴/۶               | -۳/۸۸          |
| ۲۶          | ۳۰/۷        | ۱۸/۷               | -۱/۶۴          |
| ۲۷          | ۲۲/۷        | ۱۵/۷               | -۰/۹۷          |
| ۲۸          | ۴۰/۷        | ۴۴/۷               | ۰/۵۴           |
| ۲۹          | ۱۰/۸        | ۲۶/۸               | ۱/۹۸           |
| ۳۰          | ۰۰/۸        | ۶۴/۷               | -۴/۵۰          |
| ۳۱          | ۲۸/۷        | ۶۱/۷               | ۴/۵۳           |
| ۳۲          | ۰۶/۷        | ۸۱/۶               | -۳/۵۴          |
| ۳۳          | ۱۰/۷        | ۲۵/۷               | ۲/۱۱           |
| ۳۴          | ۵۹/۷        | ۳۶/۷               | -۳/۰۳          |

جدول ۵-۱۷- ارزیابی مدل ANN به روش حذف مرحله‌ای تک تک داده‌ها

| شماره ترکیب | مقدار واقعی | مقدار پیش بینی شده | درصد خطای نسبی |
|-------------|-------------|--------------------|----------------|
| ۱           | ۱۶/۷        | ۷۹/۶               | -۵/۱۷          |
| ۲           | ۰۰/۶        | ۲۶/۶               | ۴/۳۳           |
| ۳           | ۳۰/۴        | ۴۵/۴               | ۳/۴۹           |
| ۴           | ۳۰/۴        | ۲۶/۴               | -۰/۹۳          |
| ۵           | ۷۰/۷        | ۸۰/۷               | ۱/۳۰           |
| ۶           | ۲۲/۷        | ۹۵/۶               | -۳/۷۴          |
| ۷           | ۲۲/۷        | ۲۶/۷               | ۰/۵۵           |
| ۸           | ۰۰/۷        | ۱۱/۷               | ۱/۵۷           |
| ۹           | ۷۰/۷        | ۵۲/۷               | -۲/۳۴          |
| ۱۰          | ۳۰/۷        | ۳۹/۷               | ۱/۲۳           |
| ۱۱          | ۱۶/۷        | ۲۷/۷               | ۱/۵۴           |
| ۱۲          | ۷۰/۶        | ۶۷/۶               | -۰/۴۵          |
| ۱۳          | ۸۲/۶        | ۸۵/۶               | ۰/۴۴           |
| ۱۴          | ۳۰/۷        | ۲۶/۷               | -۰/۵۵          |
| ۱۵          | ۰۰/۷        | ۱۲/۷               | ۱/۷۱           |
| ۱۶          | ۰۰/۷        | ۹۹/۶               | -۰/۱۴          |
| ۱۷          | ۹۱/۵        | ۱۹/۶               | ۴/۷۴           |
| ۱۸          | ۷۰/۶        | ۶۳/۶               | -۱/۰۴          |
| ۱۹          | ۰۰/۷        | ۹۵/۶               | -۰/۷۱          |
| ۲۰          | ۷۰/۶        | ۶۹/۶               | -۰/۱۵          |
| ۲۱          | ۶۴/۶        | ۸۳/۶               | ۲/۸۶           |
| ۲۲          | ۵۲/۷        | ۶۰/۷               | ۱/۰۶           |
| ۲۳          | ۹۲/۶        | ۹۱/۶               | -۰/۱۴          |
| ۲۴          | ۳۰/۷        | ۳۱/۷               | ۰/۱۴           |
| ۲۵          | ۷۰/۶        | ۶۴/۶               | -۰/۹۰          |
| ۲۶          | ۳۰/۷        | ۱۷/۷               | -۱/۷۸          |
| ۲۷          | ۲۲/۷        | ۱۵/۷               | -۰/۹۷          |
| ۲۸          | ۴۰/۷        | ۴۳/۷               | ۰/۴۱           |
| ۲۹          | ۱۰/۸        | ۲۵/۸               | ۱/۸۵           |
| ۳۰          | ۰۰/۸        | ۶۳/۷               | -۴/۶۳          |
| ۳۱          | ۲۸/۷        | ۵۹/۷               | ۴/۲۶           |
| ۳۲          | ۰۶/۷        | ۸۱/۶               | -۳/۵۴          |
| ۳۳          | ۱۰/۷        | ۲۵/۷               | ۲/۱۱           |
| ۳۴          | ۵۹/۷        | ۳۹/۷               | -۲/۶۴          |



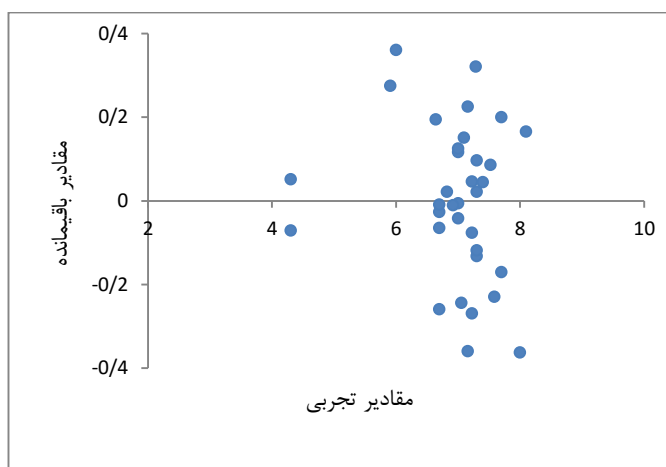
(الف)



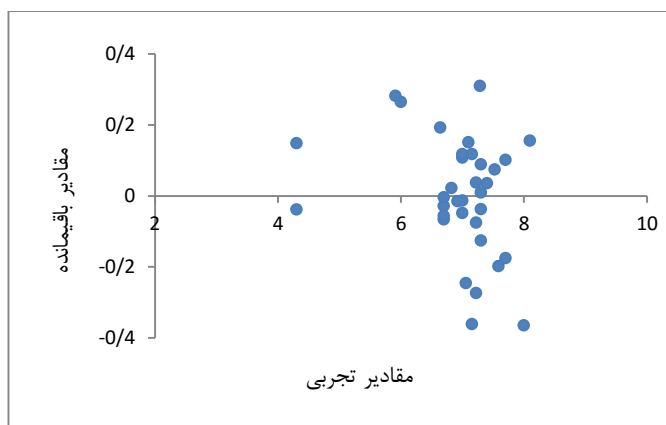
(ب)

شکل ۵-۱۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی  $IC_{50}$  به روش حذف مرحله‌ای تک تک داده‌ها (الف) روش MLR (ب) روش ANN

نمودار باقیمانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد. اگر باقیمانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند نشان می‌دهد که خطای معین در مدل وجود ندارد. شکل (۵-۱۴) نمودار باقیمانده‌ها بر حسب مقدار تجربی توسط مدل‌های خطی و غیرخطی نشان می‌دهد. پراکندگی نسبتاً متقارن نقاط در دو طرف محور افقی نشان دهنده عدم وجود خطای معین است.



(الف)



(ب)

شکل ۵-۱۴- نمودار باقیمانده‌ها بر حسب مقدار تجربی  $pic_{50}$  برای کل داده‌ها (الف) روش MLR (ب) روش ANN

۵-۵-۳- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری

مطابق جدول (۵-۱۸) هفت پارامتر آماری، جهت ارزیابی توانایی پیش‌گویی مدل‌های ساخته شده به روش خطی (MLR) و غیرخطی (ANN) به کار گرفته شد. مقادیر این پارامترها نشان می‌دهد که هر دو مدل ارائه شده از قدرت پیش‌گویی مناسبی برخوردار هستند.

جدول ۵-۱۸- پارامترهای آماری دو روش ANN و MLR

| پارامتر        | سری تست (n=۸) |        | کل داده ها (n=۳۴) |        |
|----------------|---------------|--------|-------------------|--------|
|                | MLR           | ANN    | MLR               | ANN    |
| MAE            | ۰/۱۲۵۰        | ۰/۱۲۷۵ | ۰/۱۴۶۸            | ۰/۱۲۸۵ |
| MSE            | ۰/۰۲۲۸        | ۰/۰۲۴۵ | ۰/۰۳۳۶            | ۰/۰۲۷۴ |
| PRESS          | ۰/۱۸۲۴        | ۰/۱۹۶۰ | ۱/۱۴۲۴            | ۰/۹۳۱۶ |
| SEP            | ۰/۱۵۱۰        | ۰/۱۵۶۵ | ۰/۱۸۳۳            | ۰/۱۶۵۵ |
| R <sup>2</sup> | ۰/۹۸۳۶        | ۰/۹۸۳۷ | ۰/۹۴۸۳            | ۰/۹۵۷۹ |
| REP(%)         | ۲/۱۹۶         | ۲/۲۷۶۳ | ۲/۶۳۷             | ۲/۳۸۱۰ |
| MRE            | ۱/۸۲۸         | ۱/۸۶۸  | ۲/۱۱۰۰            | ۱/۸۶۵۰ |

#### ۵-۴- ارزیابی شبکه به روش Y- تصادفی

برای بررسی هر گونه ارتباط تصادفی بین داده‌ها آزمون Y- تصادفی به کار گرفته شد. در این تکنیک، pIC<sub>50</sub> تجربی ترکیبات دارویی به صورت تصادفی جایجا و مدل‌های جدیدی ایجاد شدند. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۵-۱۹) نشان داده شده است. مقادیر پایین R<sup>2</sup> در این آزمون بیان‌گر تصادفی نبودن مدل اصلی می‌باشد.

جدول ۵-۱۹- مقادیر R<sup>2</sup> برای سری تست با استفاده از آزمون Y- تصادفی

| شماره تکرار آزمون | R <sup>2</sup> برای سری تست |
|-------------------|-----------------------------|
| ۱                 | ۰/۰۳۳                       |
| ۲                 | ۰/۵۴۴                       |
| ۳                 | ۰/۰۱۴                       |
| ۴                 | ۰/۰۴۸                       |
| ۵                 | ۰/۰۶۶                       |
| ۶                 | ۰/۷۶۶                       |
| ۷                 | ۰/۱۲۳                       |
| ۸                 | ۰/۰۳۲                       |
| ۹                 | ۰/۰۰۲۴                      |
| ۱۰                | ۰/۰۸۸                       |



## ۵-۶- بررسی ارتباط توصیف‌کننده‌های منتخب با فعالیت دارویی ترکیبات

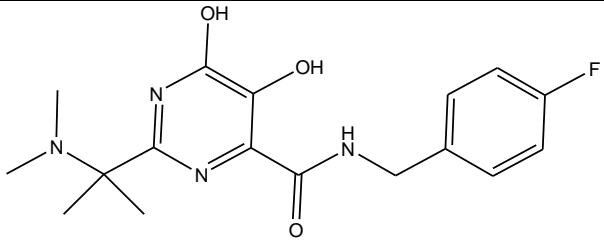
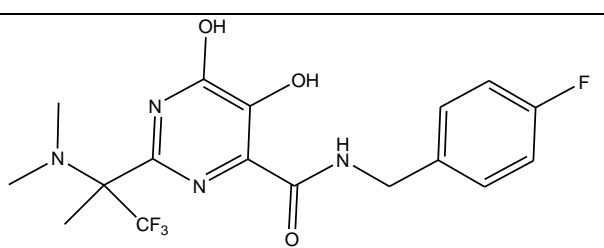
در این بخش با توجه به توصیف‌کننده‌های وارد شده در مدل‌های ANN و MLR یک بررسی اجمالی روی اثرات مختلف موجود در ثابت بازداری ترکیبات دارویی صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۶ توصیف‌کننده است که هر کدام بیانگر خصوصیات متفاوتی از ترکیبات مورد بررسی است. این توصیف‌کننده‌ها شامل BEHe2, ATS3e, RDF095v, G3u, Mor17m, و Mor30u می‌باشند.

## ۵-۶-۱- توصیف‌کننده‌های BCUT

همانطور که در بخش (۴-۷-۴) اشاره شد، این توصیف‌کننده‌ها از دسته توصیف‌کننده‌های دو بعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده (ماتریس بوردن) می‌باشند. ماتریس بوردن یک گراف تهی از هیدروژن را ارائه می‌دهد که در آن  $B_{ii}$ ، یعنی عناصر قطری در ارتباط با خاصیت اتمی مانند الکترونگاتیویته، حجم واندروالس، جرم اتمی و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند [۵۳ و ۵۴].

BEHe2 که در فصل پیش نیز به آن اشاره شد، بالاترین مقدار ویژه از ماتریس بوردن است که با الکترونگاتیویته ساندرسون وزن‌دار شده است. این توصیف‌کننده بزرگترین ضریب را در مدل به خود اختصاص داده است. هر چه الکترونگاتیویته اتم‌های تشکیل دهنده مولکول بیشتر باشد مقدار عددی این توصیف‌کننده افزایش پیدا می‌کند. با توجه به علامت مثبت اثر متوسط این توصیف‌کننده در معادله می‌توان گفت که با افزایش مقدار این توصیف‌کننده مقدار pIC<sub>50</sub> نیز افزایش می‌یابد. مثالی از اثر این توصیف‌کننده در جدول (۵-۲۰) ارائه شده است.

جدول ۵-۲۰- مثالی از اثر توصیف کننده BEHe2 بر فعالیت دارویی

| ترکیب   | BEHe2 | pIC <sub>50</sub> |
|---|-------|-------------------|
|  | ۳/۸۵۱ | ۷/۲۲              |
|  | ۳/۸۶۵ | ۸/۰۰              |

#### ۵-۶-۲- توصیف کننده‌های 2D-autocorrelation

این توصیف کننده زیر مجموعه توصیف کننده‌های خود همبستگی است. این دسته از توصیف کننده‌ها توصیف کننده‌های مولکولی هستند که بر اساس تابع خودهمبستگی ( $AC_1$ ) محاسبه می‌شوند.

تابع خودهمبستگی به صورت زیر تعریف می‌شود:

$$AC_1 = \int_a^b f(x) \cdot f(x + l) \cdot dx \quad (۲-۵)$$

که  $f(x)$  هر تابعی از متغیر  $x$  و  $l$  lag است که نشان دهنده‌ی یک بازه از  $x$  می‌باشد.  $a$  و  $b$  بازه‌ی کلی مورد مطالعه‌ی تابع را تعریف می‌کنند. تابع  $f(x)$  معمولاً یا یک تابع وابسته به زمان (مانند یک سیگنال الکتریکی وابسته به زمان) و یا یک تابع وابسته به فضا (مانند چگالی جمعیت در فضا) است [۶۴].

توصیف کننده ATS3e که در مدل نهایی دیده می‌شود با الکترونگاتیویته ساندerson وزن دار شده است. هرچه الکترونگاتیویته اتم‌های تشکیل‌دهنده ترکیبات بیشتر باشد، مقدار عددی این توصیف-کننده نیز افزایش پیدا می‌کند. علامت مثبت اثر متوسط این توصیف کننده در معادله نشان می‌دهد که با افزایش مقدار این توصیف کننده pIC<sub>50</sub> نیز افزایش می‌یابد.

### ۵-۶-۳- توصیف کننده های گروه RDF<sup>۱</sup>

می‌توان تابع توزیع شعاعی یک ترکیب شامل N اتم را به احتمال یافتن یک اتم در یک حجم کروی با شعاع r تعریف کرد و فرم کلی این تابع به صورت زیر است:

$$g(r) = f \sum_{i=1}^{N-1} \sum_{j>i}^N A_i A_j e^{\beta(r-r_{ij})^2} \quad (۳-۵)$$

که در این رابطه N تعداد اتم‌ها،  $A_i$  و  $A_j$  خاصیت اتمی ( الکترونگاتیویته، جرم اتمی، قطبش پذیری اتمی و...)،  $r_{ij}$  شعاع بین اتم i و j،  $f$  فاکتور مقیاس،  $\beta$  فاکتور دما که باعث حرکت و جابجایی اتم‌ها است، می‌باشند [۶۵].

از میان این توصیف کننده‌ها RDF<sup>۲</sup>095v که با حجم واندروالس اتمی وزن دار شده در مدل برتر دیده می‌شود. علامت مثبت اثر متوسط این توصیف کننده در معادله نشان می‌دهد که با افزایش مقدار این توصیف کننده مقدار pIC<sub>50</sub> نیز افزایش می‌یابد.

۱. Radial Distribution Function

۲. Radial Distribution Function -9.5/ weighted by atomic van der waals volumes RDF descriptors

#### ۵-۶-۴- توصیف کننده های WHIM<sup>۱</sup>

این شاخص از مختصات کارترین ساختار سه بعدی مولکول، با استفاده از کنفورمیری با حداقل انرژی محاسبه می شود و شامل اطلاعاتی درباره ی اندازه، شکل، تقارن و توزیع اتمی ساختار سه بعدی مولکول می باشد. این توصیف کننده از رابطه (۴-۵) به دست می آید:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j) (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (۴-۵)$$

که  $S_{jk}$  کوواریانس وزن دار شده بین کئوردینه  $j$  ام و  $k$  ام،  $A$  تعداد اتم ها،  $w_i$  وزن  $i$  امین اتم،  $q_{ij}$  و  $q_{ik}$  به ترتیب  $j$  امین و  $k$  امین کئوردینه های اتم  $i$  ام، و  $\bar{q}$  مقدار میانگین متناظر است. شش طرح وزن دار شدن پیشنهاد شده است که عبارتند از:

۱ - حالت بدون وزن ( $u$ ) - ۲ وزن دار شده با جرم اتمی - ۳ وزن دار شده با حجم و اندروالس - ۴ وزن دار شده با الکترون گاتیویته ساندرسون - ۵ وزن دار شده با حالت الکتروتوپولوژیکی کیپر و هال<sup>۲</sup> - ۶ وزن دار شده با قطبش پذیری [۶۶ و ۶۷].

از میان این توصیف کننده ها، توصیف کننده  ${}^3G3u$  در مدل ظاهر شده است که با توجه به علامت مثبت اثر متوسط این توصیف کننده می توان گفت که با افزایش مقدار این توصیف کننده مقدار  $pIC_{50}$  نیز افزایش می یابد.

۱ . Weighted Holistic Invariant Molecular descriptors

۲ . Kier-hall

۳ . 3st component symmetry directional WHIM index/unweighted

### ۵-۶-۵- توصیف‌کننده‌های 3D-MORSE

همانطور که در بخش (۴-۶-۲-۲) اشاره شد، توصیف‌کننده‌های 3D-MORSE از معادله‌ای که در مطالعات پراش الکترونی بکار می‌رود و امکان نمایش سه بعدی مولکول به صورت مقادیری ثابت فراهم می‌کند، قابل محاسبه‌اند. این توصیف‌کننده‌ها قادرند ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی، شیمیایی و بیولوژیکی را فراهم سازند [۵۰]. در تحقیق حاضر، برای پیش-بینی فعالیت دارویی، توصیف‌کننده  ${}^1\text{Mor}17\text{m}$  در مدل دیده می‌شود که با جرم اتمی وزن دار شده است. با افزایش جرم اتمی مولکول مقدار عددی این توصیف‌کننده افزایش پیدا می‌کند، بنابراین با توجه به علامت منفی اثر متوسط این توصیف‌کننده می‌توان گفت که افزایش مقدار این توصیف‌کننده سبب کاهش مقدار  $\text{pIC}_{50}$  می‌شود.

توصیف‌کننده  ${}^2\text{Mor}30\text{u}$  که با ویژگی خاصی از مولکول وزن دار نشده محاسبه می‌شود، دارای علامت مثبت در مدل است که اثر مستقیم بر روی مقدار  $\text{pIC}_{50}$  دارد.

### ۵-۷- نتیجه‌گیری

در این تحقیق از مدل‌های خطی و غیرخطی برای پیش‌بینی فعالیت ضد HIV یکسری مشتقات ۵، ۶-دی هیدروکسی پیریمیدین -۴- کربوکسامید استفاده شده است. بنابراین بر اساس این مدل‌ها می‌توان فعالیت ضد HIV را برای همین طبقه از ترکیبات که هنوز سنتز نشده‌اند پیش‌بینی کرد. در نتیجه می‌توان با صرف هزینه و وقت کمتر، پیش از سنتز ترکیب دارویی، با استفاده از مدل معتبر فعالیت دارویی ترکیب را پیش‌بینی کرد تا داروهایی با فعالیت دارویی مناسب با اثر بخشی بهتر تهیه شود.

۱ . 3D-MorSe-signal 17/weighted by atomic masses

۲ . 3D-MorSe-signal 30/unweighted

## آینده نگری

- ❖ محققان می‌توانند از نتایج حاصل از این تحقیق در طراحی ترکیبات (داروها) جدید استفاده کنند تا داروهایی با اثر بخشی بهتر تولید شود.
- ❖ از آن جایی که در این تحقیق توصیف‌کننده‌های به دست آمده از روش رگرسیون خطی چندگانه به عنوان ورودی شبکه عصبی مصنوعی قرار گرفته است و چون در روشهای رگرسیون خطی چندگانه تنها برهم‌کنشهای خطی بین پارامترها لحاظ می‌شود، بنابراین می‌توان از روشهایی همچون ژنتیک الگوریتم، طرح‌ریزی متوالی<sup>۱</sup> برای انتخاب توصیف‌کننده‌های مهم و روشهایی مانند SVM<sup>۲</sup> و LS-SVM<sup>۳</sup> به جای شبکه عصبی استفاده کرد.
- ❖ می‌توان از روشهای خطی دیگر مانند حداقل مربعات جزئی و آنالیز اجزای اصلی به جای روش رگرسیون خطی چندگانه استفاده کرد و نتایج حاصل را با هم مقایسه نمود.

---

۱. Successive Projection Algorithm

۲. Supported Vector Machine

۳. Least Square Supported Vector Machine

- [١]. <http://www.wikipedia.org>
- [٢]. Parlow J.J., Burney M.W., Case B.L., Girard T.J., Hall K.A., Harris P.K., Hiebsch R.R., Huff R.M., Lachance R.M., Mischke D.A., Rapp S.R., Woerndle R.S., Ennis M.D., (2010), "*Piperazinyl Glutamate Pyridinesas Potent Orally Bioavailable P2Y<sub>12</sub> Antagonists for Inhibition of Platelet Aggregation*", J. Med chem., 53, pp 2010-2037.
- [٣]. Sosnik A., Chieappetta D.A., Carcabose A.M., (2009), "*Drug delivery systems in HIV pharmacotherapy: what has been done and the challenges standing ahead*", J. controlled released., 138, pp 2-15.
- [٤]. Jalali-Heravi M., Asadollahi-Baboli M., (2008), "*QSAR analysis of platelet-derived growth inhibitors using GA-ANN and shuffling cross validation*", J. QSAR. Comb. Sci., 6, pp 750-757.
- [٥]. Arab Chamjangali M., (2009), "*Modeling of cytotoxicity data (CC<sub>50</sub>) of anti-HIV 1-[5-chlorophenyl] sulfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg- Marquardt artificial neural network*", J. Chem. Bio. Drug. Des., 73, pp 456-465.
- [٦]. De Candia M., Summo L., Carrieri A., Altomare C., Nardecchia A., Cellamare S., Carotti A., (2003), "*Investigation of Platelet Aggregation Inhibitory Activity by Phenyl Amides and Esters of Piperidinecarboxylic Acids*", J. Bioorg. Med. Chem., 11, pp 1439-1450.
- [٧]. Doble M., Karthikeyan S., Padmawar P.A., Akamanchi K.G., (2005), "*QSAR studies of paeonol analogues for inhibition of platelet aggregation*", J. Bioorg. Med. Chem., 338, pp 5996-6001.
- [٨]. Katritzky A.R., Pacureanu L.M., Slavov S., Dobchev D.A., Karelson M., (2006), "*QSAR study of antiplatelet agents*", J. Bioorg. Med. Chem., 14, pp 7490-7500.
- [٩]. Caballero J., Tundidor-Cambab A., Fernandez M., (2007), "*Modeling of the Inhibition Constant (K<sub>i</sub>) of Some Cruzain Ketone-Based Inhibitors Using 2D Spatial Autocorrelation Vectors and Data-Diverse Ensembles of Bayesian-Regularized Genetic Neural Networks*", J. QSAR. Comb. Sci., 26, pp 27-40.
- [١٠]. Jalali-Heravi M., Parastar F., (2000), "*Use of artificial neural network in a QSAR study of anti-HIV activity for a large group of HEPT derivatives*", J. Chem. Inf. Comput. Sci., 40, pp 147-154.
- [١١]. Goudarzi N., Goodarzi M., Chen T., (2009), "*QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm coupled with multiple linear regressions*", J. Med. Chem. Res., 10, pp 9542-9548.

- [۱۲]. Roy K., Leonard J.T., (2005), “*QSAR by LFER model of cytotoxicity data of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives using principal component factor analysis and genetic function approximation*”, *J. Bioorg. Med. Chem.* 13, pp 2967–2973.
- [۱۳]. Arab Chamjangali M., Beglari M., Bagherian G., (2007), “*Prediction of cytotoxicity data (CC<sub>50</sub>) of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg-Marquardt algorithm*”, *J. Mol. Graph. Model.* 26, pp 360–367.
- [۱۴]. Jalali-Heravi M., Mani-Varnosfaderani A., (2009), “*QSAR modeling of 1-(3,3-diphenylpropyl)-piperidinyl amides as CCR<sub>5</sub> modulators using multivariate adaptive regression spline and Bayesian regularization genetic neural networks*”, *J. QSAR Comb. Sci.*, 9, pp 946-958.
- [۱۵]. Ravichandran V., Mourya V.K., Agrawal., R.K., (2008), “*QSAR modeling of HIV-1 reverse transcriptase inhibitory activity with PETT derivatives*”, *Dig. J. Nanomater. Bios.*, 3, pp 9-17.
- [۱۶]. Patel A., Gupta L., Karthikeyan C., Trivedi P., (2010), “*QSAR studies on dihydro-alkoxy-benzyl-oxopyrimidines (DABOs) derivatives, as new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors*”, *J. Current Pharmaceutical Research* , 1, pp 19-25
- [۱۷]. Wold S., (1995), “*Chemometrics; what do we mean with it and what do we want from it*”, *J. Chemolab.*, 30, pp 109-115.
- [۱۸]. Manssnat D.L., Vandeginste B.G., Deming S.N., Kaufman L., (1998), “*Chemometrics, A Text Book*”, Elsevier., Amesterdom.
- [۱۹]. Howery D.G., Hirsch R.F., (1983), “*Chemometrics in the chemistry curriculum*”, *J. Chem. Ed.*, 60, pp 656-659.
- [۲۰]. Jalali-Heravi M., Garakani-Nejad Z., (1993), “*Prediction of gas chromatographic retention indices of some benzene derivatives*”, *J. Chromatogr.*, 648, pp 389-393.
- [۲۱]. Hemmatinejad B., (2005) “*Chemometrics in Iran*”, *Chemometr. Intell. Lab.*, 81, pp 202-208
- [۲۲]. Gundertofte K., Jorgensen F.S., (2000), “*Molecular modeling and prediction of bioactivity*”, Plenum Pub, New York.
- [۲۳]. Jurs P.C., (2005), “*Assessing the reliability of a QSAR model's predictions*”, *J. Mol. Graph. Model.*, 23(6), pp 503.



- [۲۴]. *Introduction to molecular descriptors*, <http://www.qsar.pe.kr/descriptors>.
- [۲۵]. 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.
- [۲۶]. Todeschini R., Consonni V., (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 98.
- [۲۷]. HyperChem7.0 Toronto, Canada: HyperCubeInc, <http://www.hyper.com>.
- [۲۸]. لواین ای. ان.، (۱۳۸۷)، "شیمی کوانتومی"، اسلامپور غ.، پارسافر غ.، مقاری ع.، نجفی ب.، جلد سوم چاپ اول.، [۲۸]. موسسه انتشارات فاطمی.
- [۲۹]. Atkins P.W., Friedman R.S., (1996), "*Molecular Quantum Mechanics*", 3<sup>rd</sup> Ed, oxford University press, New York, pp 307.
- [۳۰]. عرب چم جنگلی م.، (۱۳۸۶)، "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیتی) مشتقات ۵-فنیل-۱-فنیل - /ایمیدازول به وسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی. *H*/مینوا -
- [۳۱]. Platt J.R., (1947), "*Influence of Neighbor Bonds on Additive Bond Properties in Paraffins*". J.Chem.Phys., 15, pp 419-420.
- [۳۲]. Wiener H., (1947), "*Structural Determination of Paraffin Boiling Points*". J.Am.Chem.Soc., 69, pp 17-20.
- [۳۳]. Todeschini R., Consonni V., (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 307.
- [۳۴]. 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.
- [۳۵]. Galton F., (1894), "*Natural Inheritance*", New York, Mcmillan and Company.
- [۳۶]. فرشادفر ع.، (۱۳۸۰)، "اصول و روشهای پیشرفته آماری (تجزیه رگرسیون)"، چاپ دوم، انتشارات طاق بستان. [۳۶].
- [۳۷]. Todeschini R., Consonni V., (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 461.
- [۳۸]. Todeschini R., (2001), Milano Chemometrics and QSAR Group, Italy, *Dragon software vesion 2.1.*, <http://www.disat.unimib.it/vhml>.
- [۳۹]. SPSS 14.0 for Windows Evaluation version, Release 14.0.0, SPSS Inc., 2005.

[۴۰]. MATLAB, (The Language of Technical Computing). Version 7.5.0.342(R2007b), The Math Works Inc., 2007.

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

[۴۲]. Rosenblatt F., (1958), "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Cornell Aeronautical Lab, Psychological Review., 65, pp 386-408.

[۴۳]. 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, pp 2554-2558.

[۴۴]. 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

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

[۴۶]. سلیمی ه.، (۱۳۸۹) پایان نامه کارشناسی ارشد، "پیش‌بینی پارامتر حلالیت تعدادی حلال آلی با استفاده از روش‌های QSPR خطی و غیر خطی"، دانشگاه صنعتی شاهرود.

[۴۷]. منهای م.، (۱۳۸۷)، "مبانی شبکه‌های عصبی (هوش محاسباتی)"، جلد اول، چاپ پنجم، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

[۴۸]. <http://foram.takdownload.ir/threads/25029-؟> شبکه‌های مصنوعی چیست-؟

[۴۹]. کیا.م.، (۱۳۸۷) "شبکه‌های عصبی در MATLAB"، چاپ دوم، انتشارات کیان رایانه سبز.

[۵۰]. اشرفی م.، (۱۳۸۹)، پایان نامه کارشناسی ارشد، "مطالعه ارتباط کمی ساختار-فعالیت مشتقات تیوکربامات‌ها به عنوان دسته جدیدی از بازدارنده‌های غیر نوکلئوزیدی HIV"، دانشگاه صنعتی شاهرود.

[۵۱]. Kermani B.G., Schiffman S.S., Troy Nagle H., (2005), "Performance of the Levenberg-Marquardt neural network training method in electronic nose applications", *sensors and Actuators B*, 110, pp 13-22.

[۵۲]. Wang X.G, Tang Z., Tamura H., Ishii M., (2004), "A modified error function for the back propagation algorithm", *Neurocomputing*, 57, pp 477- 484.

[۵۳]. Wang X.G, Tang Z., Tamura H., Ishii M., Sun W., (2004), "An improved back propagation algorithm to avoid the local minima problem", *Neurocomputing*, 56, pp 455- 460.

[Δϕ]. Hajan M. T., Menhaj M. B., 1994, "Training feed forward networks with the Marquardt algorithm", IEE Trans. Neural network., pp 989-993.

[ΔΔ]. Caballero J., Fernandez L., Abreu J.I., Fernandez M., (2006), "Amino acid sequence autocorrelation vectors and ensembles of Bayesian-regularized genetic neural networks for prediction of conformational stability of human Lysozyme mutant", J. Chem. Inf. Model., 46, pp 1255-1268.

[Δϕ]. Todeschini R., Consonni V., (2000), "Handbook of Molecular Descriptors", John Wiley, New York, pp235.

[Δ∇]. Schuur J.H., Selzer P., Gasteiger J., (1996), "The coding of the three dimensional structure of molecules by molecular transform and its application to structure- spectra correlation and studies of biological activity", J. Chem. Inf. Comput. Sci., 36, pp 334-344.

[ΔΛ]. 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.

[Δ∧]. Marjanovic M., Kralj M., Supek F., Frkanec L., Piantunida I., Smuc T., Tusek-Bozic L., (2007), "Antitumor potential of crown ether: structure-activity relationship, cell cycle disturbances, and cell death studies of a series of Ionophores", J. Med. Chem., 50, pp 1007-1018.

[ϕ⋅]. Todeschini R., Consonni V., (2000), "Handbook of Molecular Descriptors", John Wiley, New York, pp 133.

[ϕ∧]. Todeschini R., Consonni V., (2000), "Handbook of Molecular Descriptors", John Wiley, New York, pp 481.

[ϕ∨]. Todeschini R., Consonni V., (2000), "Handbook of Molecular Descriptors", John Wiley, New York, pp 382.

[ϕ∗]. Pace P., Di Francesco E.M., Gardelli C., Harper S., Muraglia E., Nizi E., Orvieto F.,Petrocchi A., Poma M., Rowley M., Scarpelli R., Laufer R., Gonzalez Paz O.,MonteagudoE., Bonelli F., Hazuda D., Stillmock K.A., SummaV., (2007), "Dihydroxypyrimidine-4-carboxamides as Novel Potent and Selective HIV Integrase Inhibitors", J. Med. Chem.,50, pp 2225-2239.

[ϕϕ]. Todeschini R., and Consonni V., (2000), "Handbook of Molecular Descriptors", John Wiley, New York, pp 17-20.

[ϕΔ]. Hemmer M.C., Steinhauer V., Gasteiger J., (1999), "The prediction of the 3D structure of organic molecules from their infrared spectra", J. Vib. Spectrosc.,19, pp 151-164.

[११]. Todeschini R., Gromatica P., (1997), “3D- modeling and prediction by WHIM descriptors. Part5. Theory development and chemical meaning of WHIM descriptors”, J. Quant. Struct., Acta.,Relat., 16, pp 113-119.

[११]. Todeschini R., and Consonni V., (2000), “*Handbook of Molecular Descriptors*”, John Wiley, New York, pp 493.

**Abstract:**

In this project, quantitative structure–activity relationship studies was conducted on the inhibition constant ( $k_i$ ) of 77 drug compounds of Piperazinyl Glutamate Pyridine derivatives as Inhibition of platelet aggregation. The stepwise multiple linear regression method was used to select the most important descriptors. Then selected descriptors were used as input for QSAR model generation using multiple linear regression (MLR) and artificial neural network(ANN). The validation of the MLR and ANN models was performed using test set, leave-one-out and Y-Randomization techniques. The obtained results are shown, determination coefficients for prediction of inhibition constant of the test set by MLR and ANN models were 0.9458 and 0.9661 respectively.

In the second part of this study, MLR and ANN methods were used for modeling and accurate prediction of anti–HIV activities for a some of 5,6-Dihydroxypyrimidine-4-carboxamide derivatives. Training set was used in the selection the best MLR and ANN models by cross validation technique. The determination coefficient obtained for the test set by MLR and ANN models were 0.9836 and 0.9837 respectively. The results obtained are shown proper prediction power of proposed models.

**Keyword:** QSAR, Constant inhibition, Anti-HIV, Artificial neural network, Multiple linear regression.



**Shahrood University of Technology**

**Faculty of Chemistry**

**Prediction of Inhibition effects of some drug  
compounds using QSAR methods**

**Fatemeh Ameri**

Supervisor:

**Dr. N. Goudarzi**

Advisor:

**Dr. M. Arab Chamjangali**

Date: April – 2011