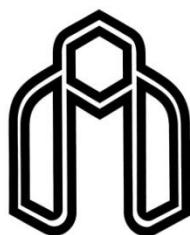


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ





دانشگاه صنعتی شاهرود

دانشکده : شیمی

گروه : شیمی فیزیک

عنوان:

مدلسازی QSPR خواص ترمودینامیکی کتون‌ها

دانشجو:

محبوبه صبوری

اساتید راهنما:

دکتر زهرا کلانتر

دکتر منصور عرب چم جنگلی

استاد مشاور:

دکتر حسین نیکوفرد

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

پائیز ۱۳۸۹

تقدیم بہ

مادر مہربان و پدر بزرگوارم

بہ پاس تمام محبت ہا، گذشت ہا و خوبی ہایشان

و ہمسر عزیزم

بہ پاس ہمراہی بی دریغش

هیچ کاری برای ذهن بهتر از شکرگزاری نیست « ادیسون »

خداوند بزرگ را سپاس می‌گویم که دروازه‌ی علم و حکمتش را بر بنده‌ی کوچکش اندکی گشود و از بیکران علمش قطره‌ای نصیبم ساخت تا این لطف و مرحمت، در قالب این دیباچه مهیا گردد.

در این راه دور، استادی اساتید و معلمی معلمانی بس بزرگوار، همچون چراغی فروزان، روشنگر راه من بود که با شفقت و تدبیر، مرا چون کودکی نوپا، قدم به قدم پیش برده و راه نمایانده- اند. هیچ حرف و کلمه‌ای اندازه و مقدار آن را ندارد تا به رسم سپاس من، به ایشان تقدیم شود. به همین الفاظ اندک و کوچک اکتفا می‌کنم تا نامی از ایشان در ابتدا برده باشم و سپاسی تقدیمشان؛ که "من لم یشکر المخلوق، لم یشکر الخالق".

با تمام احترام و ادب نام می‌برم از استاد ارجمند سرکار خانم دکتر کلانتر، که در تمام این مدت و در قدم به قدم این راه، حق استادی را بر من تمام کردند.

تشکر می‌کنم از اساتید محترم آقای دکتر عرب چم جنگلی و دکتر نیکوفرد که در تمام مراحل تکمیل مطالعات، مرا از تجارب ارزشمند خود بی نصیب نگذاشتند.

یاد می‌کنم از دو اسطوره‌ی صبر و شکیبایی، مهر و گذشت "پدر و مادر عزیزم" که اگر نبود یاری بی‌دریغشان، قامت ناتوان من در برابر سختی‌ها و مرارت‌های این راه استوار نمی‌ماند. به همان اندازه از همراهی همیشگی همسر مهربان و برادر عزیزم کمال تشکر را دارم.

و سپاسگزارم از زحمات خالصانه کارمندان دانشکده شیمی دانشگاه صنعتی شاهرود و همراهی همیشگی دوستان عزیزم خانم‌ها طبسی، محمدی و جمشیدی.

"حاصل تحقیق پیش رو، شاید جبران گوشه‌ای از زحماتشان"

چکیده

مدل شبکه‌ی عصبی مصنوعی، برای پیش‌بینی دانسیته‌ی کتون‌ها در محدوده‌ی فشار MPa ۲-۳۰۰ و محدوده‌ی دمایی K ۲۷۳-۳۳۸ به‌کار گرفته شد. توصیف‌کننده‌ها از میان ۱۸ دسته‌ی مختلف از توصیف‌کننده‌ها در نرم‌افزار DRAGON با استفاده از روش رگرسیون مرحله‌ای چندگانه و تکنیک stepwise انتخاب شدند. توصیف‌کننده‌های انتخاب شده شامل ATS4e و RDF040e بوده و همچنین دو توصیف‌کننده‌ی محاسباتی T و p برای ساخت مدل به‌کار گرفته شدند. داده‌های حاصل، به‌صورت گزینشی به سه دسته‌ی آموزش، ارزیابی و تست به ترتیب با تعداد ۳۲۴، ۱۰۷ و ۱۰۷ تقسیم شدند به‌گونه‌ای که هر سری بهترین نماینده‌ی کل داده‌ها باشد. پس از مدلسازی و بهینه‌کردن پارامترهای شبکه، مدل ساخته شده، با استفاده از سری تست، مورد بررسی قرار گرفت. خطای مربع میانگین (MSE) برای سری تست، توسط مدل‌های ANN و MLR به ترتیب ۰/۹۳۶۰ و ۰/۱۷۳۵ بودند. نتایج به‌دست آمده نشان می‌دهد که هر دوی روش‌های خطی و غیر خطی توانایی دقیقی در پیش‌بینی دانسیته دارند، اگرچه ANN نتایج دقیق‌تری را نشان می‌دهد.

کلمات کلیدی: شبکه‌ی عصبی مصنوعی - رگرسیون مرحله‌ای چندگانه - توصیف‌کننده - کتون‌ها

لیست مقالات مستخرج از پایان نامه

ارائه مدل رگرسیون خطی چندگانه برای محاسبه دانسیته کتونها

(هفتمین همایش ملی شیمی دانشگاه پیام نور شیراز)

2- Prediction of Thermodynamic properties of Ketones Using Artificial Neural Network

(سیزدهمین سمینار شیمی فیزیک ایران – دانشگاه صنعتی شیراز و دانشگاه شیراز)

فهرست مطالب

صفحه	عنوان
ز.....	چکیده.....
ط.....	فهرست مطالب.....
م.....	فهرست اشکال.....
س.....	فهرست جداول.....

فصل اول: مقدمه

۱.....	۱-۱- مقدمه.....
۲.....	۲-۱- هدف تحقیق.....
۳.....	۳-۱- روش کار.....
۳.....	۴-۱- پیشینه‌ی تحقیق.....

فصل دوم: کمومتریکس

۵.....	۱-۲- کمومتریکس.....
۶.....	۲-۲- ارتباط کمی ساختار - خاصیت (QSPR).....
۸.....	۳-۲- ارتباط کمی ساختار - فعالیت (QSAR).....
۸.....	۴-۲- روش‌های اعتبارسنجی.....
۹.....	۲-۴-۲- اعتبارسنجی تقاطعی.....
۹.....	۲-۴-۲- تقسیم کردن مجموعه به دو دسته‌ی آموزش و ارزیابی.....
۹.....	۲-۴-۳- روش اعتبارسنجی بیرونی.....
۱۰.....	۲-۵-۲- رگرسیون خطی چندگانه (MLR).....
۱۲.....	۲-۵-۱- انتخاب سری مولکولی.....

- ۱۲.....۲-۵-۲- محاسبه‌ی توصیف‌کننده‌ها
- ۱۴.....۲-۵-۲-۱- توصیف‌کننده‌های کوانتوم شیمیایی
- ۱۴.....۲-۵-۲-۲- توصیف‌کننده‌های توپولوژی
- ۱۵.....۲-۵-۳- توصیف‌کننده‌های هندسی
- ۱۵.....۲-۵-۴- توصیف‌کننده‌های فیزیکو- شیمیایی
- ۱۵.....۲-۵-۵- توصیف‌کننده‌های الکتروتوپولوژی
- ۱۵.....۲-۵-۳- انتخاب توصیف‌کننده‌های مناسب
- ۱۶.....۲-۵-۳-۱- روش ورود اجباری
- ۱۷.....۲-۵-۳-۲- روش حذف مرحله‌ای
- ۱۷.....۲-۵-۳-۳- روش افزایش مرحله‌ای
- ۱۷.....۲-۵-۴- ساخت مدل
- ۱۸.....۲-۵-۵- استفاده از مدل جهت پیش‌بینی خاصیت موردنظر
- ۱۸.....۲-۶-۶- نرم‌افزارهای مورد استفاده در این پژوهش
- ۱۸.....۲-۶-۱- نرم‌افزار Hyperdem 7
- ۱۹.....۲-۶-۲- نرم‌افزار Dragon 3.0
- ۲۰.....۲-۶-۳- نرم‌افزار spss 13
- ۲۱.....۲-۶-۴- نرم‌افزار MATLAB

فصل سوم: شبکه‌ی عصبی مصنوعی

- ۲۱.....۳-۱- شبکه عصبی
- ۲۲.....۳-۲- تاریخچه شبکه‌های عصبی مصنوعی
- ۲۴.....۳-۳- ساختار شبکه‌های عصبی زیستی
- ۲۵.....۳-۴- مقایسه مغز و رایانه

- ۲۷.....۳-۵- ساختار شبکه عصبی مصنوعی.....
- ۲۸.....۳-۶- توابع محرک.....
- ۲۹.....۳-۷- تقسیم بندی لایه‌های شبکه‌های عصبی مصنوعی.....
- ۲۹.....۳-۷-۱- شبکه‌ی عصبی تک‌لایه.....
- ۳۰.....۳-۷-۲- شبکه عصبی چند لایه.....
- ۳۱.....۳-۸- آموزش شبکه‌های عصبی.....
- ۳۱.....۳-۹- روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی.....
- ۳۲.....۳-۹-۱- الگوریتم نزول گرادیانی.....
- ۳۶.....۳-۹-۲- الگوریتم کمترین میانگین مربع خطا (LMS).....
- ۳۷.....۳-۹-۳- الگوریتم پس انتشار خطا.....
- ۳۷.....۳-۹-۴- الگوریتم لونبرگ - مارکوارت.....
- ۴۱.....۳-۹-۵- الگوریتم تنظیم بایزین.....
- ۴۱.....۳-۹-۵-۱- تابع عملکرد اصلاح شده.....
- ۴۲.....۳-۹-۵-۲- تنظیم اتوماتیک بایزین.....
- ۴۲.....۳-۱۰- محاسن شبکه‌های عصبی مصنوعی.....
- ۴۳.....۳-۱۱- معایب شبکه‌های عصبی مصنوعی.....
- ۴۴.....۳-۱۲- کاربردهای شبکه عصبی مصنوعی.....

فصل چهارم: مدلسازی QSPR خواص ترمودینامیکی کتون‌ها

- ۴۶.....۴-۱- مقدمه.....
- ۴۶.....۴-۲- انتخاب سری داده‌ها.....
- ۷۲.....۴-۳- محاسبه‌ی توصیف‌کننده‌ها.....
- ۷۲.....۴-۴- انتخاب توصیف‌کننده‌ها با استفاده از روش ورود مرحله‌ای.....

- ۷۴-۴-۱- مدلسازی به روش MLR..... ۷۴
- ۷۶-۴-۵- مدلسازی به روش شبکه عصبی مصنوعی..... ۷۶
- ۴-۵-۱- انتخاب تابع انتقال، تابع آموزش تعداد دوره و تعداد گره در لایه پنهان آموزش..... ۷۶
- ۴-۶- هندسه‌ی شبکه عصبی مصنوعی بهینه‌شده..... ۸۳
- ۴-۷-۱- ارزیابی شبکه و مقایسه‌ی روش‌های رگرسیون خطی چندگانه و شبکه‌ی عصبی مصنوعی..... ۸۴
- ۴-۷-۱- نمودار برگشت..... ۸۴
- ۴-۷-۲- نمودار خطای باقیمانده..... ۸۷
- ۴-۷-۳- ارزیابی شبکه با استفاده از پارامترهای آماری..... ۹۴
- ۴-۷-۴- ارزیابی شبکه با استفاده از یک ترکیب جدید..... ۹۵
- ۴-۷-۵- ارزیابی شبکه با استفاده از تست Y-randomation..... ۱۰۰
- ۴-۸-۱- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با دانسیته..... ۱۰۱
- ۴-۸-۱- توصیف‌کننده ATS4e..... ۱۰۱
- ۴-۸-۲- توصیف‌کننده RDF040e..... ۱۰۲
- ۴-۸-۳- توصیف‌کننده فشار p..... ۱۰۳
- ۴-۸-۴- توصیف‌کننده دما T..... ۱۰۴
- ۱۰۵- آینده‌نگری..... ۱۰۵
- ۱۰۶- منابع..... ۱۰۶

فهرست اشکال

عنوان.....	صفحه.....
شکل ۳-۱- نمایی از دو نرون زیستی.....	۲۵.....
شکل ۳-۲- ساختمان یک نرون محاسباتی با بردار ورودی دارای R عضو.....	۲۷.....
شکل ۳-۳- ساختار یک شبکه‌ی تک لایه با R ورودی و S نرون.....	۲۹.....
شکل ۳-۴- ساختار یک شبکه سه لایه.....	۳۱.....
شکل ۴-۱- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی با تابع تبدیل logsig و تابع آموزش br.....	۷۸.....
شکل ۴-۲- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی با تابع تبدیل logsig و تابع آموزش lm.....	۷۹.....
شکل ۴-۳- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی با تابع تبدیل tansig و تابع آموزش br.....	۸۰.....
شکل ۴-۴- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی با تابع تبدیل tansig و تابع آموزش lm.....	۸۱.....
شکل ۴-۵- تصویر شماتیک ساختار شبکه‌ی عصبی مصنوعی به‌دست آمده پس از بهینه سازی.....	۸۳.....
شکل ۴-۶- نمودار برگشت دانسیته پیش‌بینی شده توسط MLR بر حسب دانسیته تجربی برای الف) سری تست و ب) سری ارزیابی.....	۸۵.....
شکل ۴-۷- نمودار برگشت دانسیته پیش‌بینی شده توسط ANN بر حسب دانسیته تجربی برای الف) سری تست و ب) سری آموزش.....	۸۶.....
شکل ۴-۸- نمودار خطای باقیمانده نسبت به مقدار تجربی با استفاده از روش MLR برای الف) سری تست و ب) سری ارزیابی.....	۹۱.....

شکل ۴-۹- نمودار خطای باقیمانده نسبت به مقدار تجربی با استفاده از روش ANN برای

الف) سری تست و ب) سری ارزیابی.....۹۲

شکل ۴-۱۰- نمودار دانسیته پیش‌بینی بر حسب دانسیته تجربی برای مولکول ۲- هگزانون.....۹۵

فهرست جداول

عنوان.....	صفحه.....
جدول ۱-۱- فهرست توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon.....	۱۹.....
جدول ۱-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی استون.....	۴۷.....
جدول ۲-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی ۲ - بوتانول.....	۴۹.....
جدول ۳-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی ۲- پنتانول.....	۵۴.....
جدول ۴-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی ۴- متیل - ۲- پنتانول.....	۵۸.....
جدول ۵-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی ۳- پنتانول.....	۶۳.....
جدول ۶-۴- مقادیر مولکولی k, T, p و دانسیته‌ی مولکولی ۲- هگزانون.....	۶۷.....
جدول ۷-۴- ماتریس هم‌بستگی متغیر های مستقل وارد شده.....	۷۳.....
جدول ۸-۴- مشخصات مدل منتخب توسط روش رگرسیون خطی چندگانه.....	۷۳.....
جدول ۹-۴- مقادیر F, R و خطای استاندارد برای مدل‌های ۲ تا ۴.....	۷۴.....
جدول ۱۰-۴- مقادیر توصیف‌کننده‌های وارد شده در مدل منتخب.....	۷۵.....
جدول ۱۱-۴- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل logsig	
و تابع آموزش br.....	۷۸.....
جدول ۱۲-۴- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل logsig	
و تابع آموزش Im.....	۷۹.....
جدول ۱۳-۴- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل tansig	
و تابع آموزش br.....	۸۰.....
جدول ۱۴-۴- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل tansig	
و تابع آموزش Im.....	۸۱.....
جدول ۱۵-۴- توابع و مقادیر پارامترهای مهم شبکه‌ی عصبی.....	۸۲.....

جدول ۴-۱۶- مقادیر تجربی دانسیته و مقادیر پیش‌بینی شده توسط تکنیک‌های ANN و MLR

(۷: سری ارزیابی، t : سری تست)..... ۸۸

جدول ۴-۱۷- مقایسه آماره‌های به‌دست‌آمده از دو روش ANN و MLR..... ۹۴

جدول ۴-۱۸- مقادیر R ، R^2 ، F ، SE و MSE برای مولکول ۲- هگزانون..... ۹۵

جدول ۴-۱۹- مقادیر دانسیته‌ی تجربی، دانسیته‌ی پیش‌بینی شده و درصد خطا برای

مولکول ۲- هگزانون..... ۹۶

جدول ۴-۲۰- مقادیر R^2 برای سری تست و سری ارزیابی بعد از چندین تست \bar{Y} -تصادفی. ۱۰۰

۱-۱- مقدمه

برای درک مکانیسم فرآیندهای مختلف شیمیایی، کشف و توسعه‌ی مواد جدید، حفظ محیط زیست و زمینه‌های دیگر شیمی، هنوز توانایی حل مسائل به طور کامل وجود ندارد و برای عملی کردن بعضی از مسائل، نیاز به سیستم‌های بسیار پیچیده‌ای می‌باشد که انجام آن‌ها در گرو صرف هزینه‌های بسیار و مطالعات گسترده است.

در جهت حل این مشکل، روش‌های محاسباتی کمومتریکس می‌توانند مفید باشند. تجزیه و تحلیل آماری و ریاضی داده‌های شیمیایی معمولاً تحت عنوان کمومتریکس یاد می‌شود. به عبارتی دیگر کمومتریکس یک روش کارآمد برای خلاصه‌کردن اطلاعات مفید از یک سری داده مشخص و پیش‌بینی سری دیگر داده‌هاست. درحقیقت هدف کمومتریکس، بهبود بخشیدن فرآیندهای اندازه‌گیری و استخراج اطلاعات شیمیایی مفیدتر از داده‌های اندازه‌گیری شده فیزیکی و شیمیایی می‌باشد. چندین تعریف برای کمومتریکس بیان شده است که غالباً در متن‌های تجزیه‌ای به‌کار می‌روند. یکی از جامع‌ترین تعاریف به‌صورت زیر است:

کمومتریکس شاخه‌ای از شیمی است که از ریاضی و آمار و منطق استفاده می‌کند برای اینکه:

الف) فرآیندهای تجربی بهینه را طراحی و انتخاب کند.

ب) حداکثر اطلاعات شیمیایی قابل حصول را از تحلیل اطلاعات شیمیایی فراهم کند.

ج) بتوان اطلاعات بیشتری در مورد سیستم‌های شیمیایی به‌دست آورد.

کمومتریکس به عنوان یک شاخه علمی جوان در دو دهه‌ی اخیر به سرعت توسعه پیدا کرده است. این رشد سریع مدیون پیشرفت سریع دستگاه‌های هوشمند و خودکار آزمایشگاهی و همچنین امکان استفاده از کامپیوترهای قدرتمند و نرم‌افزارهای ساده است. بنابراین کمومتریکس به عنوان یک وسیله در همه‌ی قسمت‌های شیمی و بیشتر در زمینه شیمی تجزیه مورد استفاده قرار گرفته است. امروزه در کمومتریکس از روش‌های مختلف ریاضی و آماری استفاده شده است. همچنین در

پیشرفت‌های دو دهه‌ی اخیر از شبکه‌های عصبی مصنوعی^۱، منطق فازی و الگوریتم ژنتیکی نیز استفاده شده است.

این روش‌ها جزء دسته‌ای از علوم کامپیوتر به نام هوش مصنوعی^۲ (AI) قرار می‌گیرند که تفاوت اساسی با سایر روش‌های محاسباتی و همچنین با یکدیگر دارند. در سیستم‌های ریاضی متداول، دانسته‌ها به صورت قواعد روشن و از پیش مشخص شده‌ای در سیستم وجود دارد، در حالی که این سیستم‌ها قواعد را توسط آموزش از مثال‌هایی که به آن‌ها داده می‌شوند فرا می‌گیرند. از میان روش‌های فوق استفاده از شبکه‌های عصبی مصنوعی در شیمی نقش برجسته‌ای پیدا کرده و کاربرد آن در شیمی به سرعت در حال افزایش است.

یکی از مهم‌ترین ویژگی‌های شبکه عصبی مصنوعی در این است که شبکه‌های عصبی، سیستمی برنامه‌ریزی شده با قواعد از پیش تعیین شده نمی‌باشند و می‌توانند طی فرآیندی که آموزش نامیده می‌شود ساختار درونی خود را به صورت تجربی به نحوی تنظیم کنند که بهترین پاسخ ممکن را برای داده‌های ورودی جدید ایجاد کند.

۲-۱- هدف تحقیق

در مطالعه‌ی سیستم‌های شیمیایی، بیوشیمیایی و طبیعی دانستن خواص مختلف فیزیکی و شیمیایی مواد خالص و یا مخلوط‌ها تحت شرایط مختلف ترمودینامیکی لازمست. یکی از مهم‌ترین این خواص، دانسیته می‌باشد. اما همیشه یافتن مقادیر تجربی معتبر در شرایط مختلف دما و فشار، خصوصاً از منابع علمی امکانپذیر نیست. به‌علاوه اندازه‌گیری تجربی این خواص مشکل بوده و به صرف هزینه و وقت زیادی نیاز دارد. در این شرایط استفاده از روش‌های مختلف علمی برای پیش‌بینی

^۱Artificial Neural Network(ANN)

^۲Artificial Intelligence (AI)

دانسیتیه مطلوب نظر می‌باشد. در این کار، هدف پیش‌بینی دانسیته‌ی کتون‌ها بر حسب دما و فشار با استفاده از روش نوین پردازش اطلاعات، شبکه‌ی عصبی مصنوعی می‌باشد.

۳-۱- روش کار

در این پایان نامه به منظور مطالعه‌ی ارتباط کمی ساختار - دانسیته از دو روش رگرسیون خطی چندگانه^۳ و شبکه‌ی عصبی مصنوعی استفاده شده است. در بخش اول، از رگرسیون خطی چندگانه به منظور ایجاد ارتباط خطی بین دانسیته و ساختار ۶ مولکول آلی از دسته‌ی کتون‌ها شامل استون، ۲- پنتانول، ۲- بوتانول، ۴- متیل-۲- پنتانول، ۳- پنتانول و ۲- هگزانول با شرایط مختلف از نظر دما و فشار استفاده شد. در بخش دوم این پایان نامه، شبکه‌ی عصبی مصنوعی به منظور بررسی روابط غیرخطی بین دانسیته و ساختار این ترکیبات به کار رفته است. پس از رسم ساختار مولکول‌ها، توصیف‌کننده‌های مولکولی^۴ از ساختار ترکیبات مورد نظر توسط نرم‌افزارهای مناسب استخراج شد. سپس با استفاده از توصیف‌کننده‌های محاسبه شده توسط نرم‌افزارهای آماری، مدل مناسب به دست آمد.

۴-۱- پیشینه‌ی تحقیق

محاسبه‌ی دانسیته‌ی کتون‌ها بر حسب دما و فشار، اولین بار با استفاده از قاعده‌ی همدمای خطی اصلاح شده^۵ (MLIR) انجام گرفته است [۱]. این روش با تلفیقی از معادله‌ی حالت با روش سهم گروه‌ها توانسته است دانسیته‌ی کتون‌ها را با متوسط خطای نسبی کمتر از ۱/۱ درصد و ماکزیموم خطای نسبی ۱/۵ درصد پیش‌بینی کند. پس از آن در سال ۲۰۰۹، محاسبه‌ی دانسیته بر حسب دما و

³Multiple Linear Regression(MLR)

⁴Molecular Descriptors

⁵ Modifical Linear Isotherm Regularity

فشار، با تلفیقی از شبکه عصبی موجک و روش سهم گروه‌ها انجام گرفته است که در آن دانسیته‌ی کتون‌ها با متوسط خطای نسبی کمتر از ۱/۰۵ درصد و ماکزیموم خطای نسبی کمتر از ۱/۹۵ درصد پیش‌بینی گردیده است [۲].

۲-۱- کموتریکس

اگر چه هنوز توانایی حل مسائل مختلف در زمینه‌ی درک مکانیسم فرایندهای مختلف شیمیایی، کشف و توسعه‌ی مواد جدید و زمینه‌های دیگر شیمی هنوز به طور کامل وجود ندارد، اما روش‌های محاسباتی کموتریکس در جهت حل این مشکلات به میزان زیادی مفید واقع شده‌اند. در حدود ۱۵۰ سال است که دانشمندان در حال مطالعه و بررسی فعالیت بیولوژیکی ترکیبات مختلف می‌باشند. امروزه، در هر سال میلیاردها دلار توسط شرکت‌های شیمیایی و دارویی صرف انجام آزمایش‌های گوناگون بر روی انواع آفت‌کش‌ها و بررسی میزان سمیت محصولات طبیعی و مصنوعی می‌شود. در این راستا، آنچه در تمامی این زمینه‌ها ضروری به نظر می‌رسد، استفاده از ابزاری جهت پیش‌بینی فعالیت شیمیایی ترکیبات مختلف بدون انجام آزمایش و تنها با استفاده از اطلاعات ساختاری ترکیب می‌باشد.

از قدرتمندترین تکنیک‌ها که خواص مولکول‌ها را به ویژگی‌ها ساختاری آنها نسبت می‌دهد کموتریکس است. از نظر شیمی‌دانان، فعالیت‌ها و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. این نوع از مطالعات به بررسی کمی ساختار با فعالیت^۶ QSAR و همچنین بررسی کمی ارتباط ساختار با خاصیت^۷ QSPR معروف می‌باشند. منظور از فعالیت، اثر بیولوژیکی ترکیبات مختلف بر ساختار بدن موجود زنده است. تاثیر ترکیبات دارویی گوناگون بر روی سلول‌های سرطانی، مثالی از این نوع است. منظور از خاصیت نیز خواص فیزیکی ترکیبات مختلف مانند حلالیت می‌باشد. علی‌رغم نامگذاری متفاوت، نحوه‌ی عملکرد در هر روش یکسان است با این تفاوت که در QSAR فعالیت و در QSPR خاصیت مورد مطالعه و بررسی قرار می‌گیرد [۳]. نتایج این مطالعات علاوه بر شفاف‌سازی نحوه‌ی ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمان آنها، به پژوهشگران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند.

^۶ Quantitative Structure-Activity Relationship (QSAR)

^۷ Quantitative Structure-property Relationship (QSPR)

عبارت کموتریکس توسط اسوانت ولد^۸ و بروس کوواسکی^۹ در اوایل سال ۱۹۷۰ تعریف شد [۴].
قبلاً عباراتی مانند بیومتریکس^{۱۰} و اقتصادسنجی^{۱۱} نیز در زمینه‌های علوم زیستی و اقتصاد معرفی شده اند و پس از آن بود که انجمن بین‌المللی کموتریکس تاسیس شد. هاوری و هیرش^{۱۲} [۵] در اوایل سال ۱۹۸۰ پیشرفت‌های کموتریکس را در مراحل مختلف رده‌بندی کردند. هاوری و هیرش تحقیق خود را به طور ویژه به همبستگی تعداد زیادی از داده‌های شیمیایی به ویژگی‌های مولکولی مربوطه اختصاص دادند. این اقدامات اولیه، اساس یکی از شاخه‌های مهم کموتریکس یعنی ارتباط کمی ساختار - فعالیت‌ها را فراهم کرد. اولین مرحله قبل از ۱۹۷۰ است که در آن تعدادی از روش‌های ریاضی توسعه یافتند و در زمینه‌های مختلف ریاضی و مهندسی استفاده شدند. دومین مرحله‌ی تکامل کموتریکس در سال ۱۹۷۰ به وقوع پیوست. در این سال، شیمیدانان نه تنها روش‌های در دسترس برای آنالیز داده‌ها بلکه روش‌های نوینی را برای برآوردن نیازهایشان توسعه دادند. مرحله‌ی بعدی تکامل کموتریکس، از اوایل سال ۱۹۸۰ توسط هاوری و هیرش [۵] و سپس توسط براون [۴] آغاز شد. پس از آن کموتریکس به عنوان یک ابزار برای شیمیدانان و یک شاخه‌ی وابسته‌ی شیمی به شمار آمد. با پیشرفت در علوم کامپیوتر، اطلاعات، آمار و ریاضی کاربردی عناصر جدیدی در کموتریکس معرفی گردید. به طور مثال شبکه‌ی عصبی مصنوعی [۶، ۷] و الگوریتم ژنتیکی [۸] نمونه‌هایی از این روش‌ها هستند.

۲-۲- ارتباط کمی ساختار - خاصیت (QSPR) اساس این روش ارتباط بین

خاصیت یک ترکیب خاص با ساختار شیمیایی آن می‌باشد. در این روش خاصیت به صورت تابعی از خواص فیزیکی - شیمیایی در قالب یک مدل ارائه می‌شود و

⁸ SvantWold

⁹ Bruce R.Kowalski

¹⁰ Biometrics

¹¹ Econometrics

¹² Howery and Hirsch

در نهایت این مدل برای پیش‌بینی خاصیت سایر ترکیبات مشابه مورد استفاده قرار می‌گیرد. استفاده از چنین مدلی برای پیش‌بینی خاصیت ترکیبات مختلف مانع از اتلاف وقت و هزینه خواهد شد. ایجاد ارتباط بین خاصیت یک ترکیب و ساختار آن با استفاده از چند توصیف‌کننده‌ی مولکولی مناسب امکان‌پذیر است. توصیف‌کننده‌های مولکولی پارامترهای نظری هستند که به ساختار مولکول ارتباط دارند و برای هر مولکول مقادیر خاصی را به خود اختصاص می‌دهند [۳]. این توصیف‌کننده‌ها از روی ساختارهای مولکولی و با استفاده از الگوریتم‌های شناخته شده‌ی ریاضی، نظریه گراف شیمیایی [۹]، روش مکانیک کوانتومی و غیره محاسبه می‌شوند.

امروزه QSPR به صورت موفقیت‌آمیزی برای پیش‌بینی بسیاری از خواص فیزیکی ترکیبات شیمیایی به کار گرفته می‌شود. از جمله کاربردهای QSPR می‌توان به مطالعه حلالیت، گرانش و زمان بازداری ترکیبات در ستونهای کروماتوگرافی اشاره نمود. در واقع از این روش می‌توان برای پیش‌بینی بسیاری از خواص فیزیکی که با استفاده از ساختار مولکولی ترکیبات تعیین می‌شوند، استفاده نمود.

مزایای استفاده از QSPR عبارتند از:

۱- تعداد توصیف‌کننده‌های مورد استفاده در این روش کم است و این خاصیت منجر به ایجاد مدل‌های قوی می‌شود.

۲- توصیف‌کننده‌های استفاده شده در این مدل، معانی فیزیکی تعریف شده‌ای دارند. این معانی فیزیکی برای جستجوی اطلاعات فیزیکی - شیمیایی که وابستگی بالایی با خواص موردنظر دارند، مفید هستند.

۳- مدل‌های حاصل شده، از لحاظ نظری برای ترکیبات مختلف کاربرد دارند.

به همین جهت روش‌های QSPR به طور گسترده در پیش‌بینی خواص فیزیکی - شیمیایی مختلف مثل نقطه‌ی جوش، نقطه‌ی ذوب، فشار بخار، خواص بحرانی، حلالیت و... استفاده می‌شوند [۱۰-۱۳].

در مطالعات QSPR برای مدلسازی از تکنیک‌های مختلفی مانند رگرسیون خطی چندگانه و شبکه‌های عصبی مصنوعی در مدل‌سازی استفاده می‌شود.

۲-۳- ارتباط کمی ساختار - فعالیت (QSAR)

از کاربردهای دیگر روش‌های کموتریکس می‌توان به پیدا کردن ارتباط کمی ساختار - فعالیت و ارزیابی داده‌های تجزیه‌ای اشاره کرد. کوچکترین تغییر و تحول در ساختار مولکول‌های مشابه، باعث ایجاد فعالیت‌های شیمیایی متفاوت می‌شود. به طور کلی QSAR رابطه‌ای متعادل و هماهنگ بین ساختارهای مولکولی و فعالیت شیمیایی ترکیبات برقرار می‌کند [۱۷-۱۴]. QSAR در زمینه‌هایی مانند طراحی دارو و برآورد آسیب‌های محیطی ترکیبات شیمیایی مورد استفاده قرار می‌گیرد. معمول‌ترین روشهایی که در QSPR مورد استفاده قرار می‌گیرد عبارتند از:

- رگرسیون خطی چندگانه (MLR)

- آنالیز اجزای اصلی^{۱۳} (PCA)

- حداقل مربعات جزئی^{۱۴} (PLS) [۱۹، ۱۸].

۲-۴- روش‌های اعتبارسنجی^{۱۵} [۲۰]

روش‌های اعتبارسنجی برای ارزیابی اعتبار و قدرت پیش‌بینی مدل‌های حاصل از سری داده توسط رگرسیون چند متغیری^{۱۶} و روش‌های دسته‌بندی^{۱۷} به کار می‌روند. روش‌های اعتبارسنجی عبارتند از: اعتبارسنجی تقاطعی، اعتبارسنجی بیرونی و تقسیم کردن مجموعه به دو دسته آموزش و ارزیابی.

¹³ Principal Component Analysis (PCA)

¹⁴ Partial Least Squares (PLS)

¹⁵ Validation Technique

¹⁶ Multivariate Regression

¹⁷ Classification Methods

۲-۴-۱- اعتبارسنجی تقاطعی

اعتبارسنجی تقاطعی یکی از رایج‌ترین روش‌های اعتبارسنجی است که در آن، در هر مرحله یک داده یا گروه کوچکی از داده‌ها کنار گذاشته می‌شود [۲۱] و برای هر مجموعه داده‌ی باقی‌مانده در هر مرحله، مدلی ایجاد می‌گردد. سپس از روی مدل حاصل، داده‌های کنار گذاشته شده مورد بررسی و پیش‌بینی قرار می‌گیرند. مجدور تفاوت میان پاسخ واقعی و پاسخ پیش‌بینی شده^{۱۸} با عبارت $press$ بیان می‌شود. با استفاده از $press$ نهایی مقادیر مربع ضریب همبستگی (R^2) و انحراف استاندارد خطای پیش‌بینی^{۱۹} (SDEP) محاسبه می‌شود [۲۲].

۲-۴-۲- تقسیم کردن مجموعه به دو دسته‌ی آموزش و ارزیابی

در این روش، مجموعه‌ی داده‌ها به دو سری آموزش و ارزیابی تقسیم می‌شوند. سپس با استفاده از داده‌های سری آموزش، مدل بهینه شده و توسط سری ارزیابی قدرت پیش‌بینی آن، بررسی می‌شود. تقسیم‌بندی داده‌ها به دو سری آموزش و ارزیابی به صورت کاملاً تصادفی انجام می‌شود.

۲-۴-۳- روش اعتبارسنجی بیرونی^{۲۰} [۲۳]

در این روش، مجموعه‌ی داده‌ها علاوه بر سری آموزش و سری ارزیابی، به سری تست نیز تقسیم می‌شود. در این روش مدل با استفاده از سری آموزش بهینه می‌شود سپس قدرت پیش‌بینی مدل توسط سری ارزیابی، بررسی و در نهایت یک بررسی اضافی روی سری تست جهت قدرت پیش‌بینی مدل انجام می‌شود.

¹⁸ Predictive Residual Sum of Squares

¹⁹ Standard Deviation Error of Prediction

²⁰ External Validation

۲-۵- رگرسیون خطی چندگانه (MLR)

روش رگرسیون خطی چندگانه برای مدل‌سازی رابطه‌ی خطی بین متغیر وابسته و یک یا چند متغیر مستقل به کار می‌رود. در این روش متغیرهای مستقل، متغیرهایی هستند که فرض می‌شود در تابعی که برای مدل‌سازی مناسب است سهم دارند. متغیرهای وابسته، متغیرهایی هستند که اغلب از اندازه‌گیری‌های تجربی بدست می‌آیند و وابستگی آماری میان آنها و تعدادی از متغیرهای مستقل موردنظر است.

برای داده تجربی و متغیر مستقل، مدل رگرسیون خطی چندگانه توسط معادله‌ی (۱-۱) نمایش داده می‌شود:

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

در معادله‌ی فوق $\beta_0, \beta_1, \beta_p, \dots$ و ضرایب رگرسیون X_1, X_2, \dots, X_p متغیرهای مستقل و Y_i متغیر وابسته موردنظر می‌باشد.

در طی فرایند محاسبه‌ی مدل مناسب، فاکتورهای آماری که صحت مدل رگرسیون را نشان می‌دهند، نیز محاسبه می‌شوند. این فاکتورها عبارتند از: t, F, R, R^2 و ضریب همبستگی میان توصیف‌کننده‌ها.

✳️ آزمون F (آزمون فیشر): آزمون F ، آزمون معنادار بودن آماری در تحلیل رگرسیون ساده و چند متغیره و مشابه آزمون‌های معنی‌داری آماری در تحلیل واریانس است [۲۴].

نسبت F از رابطه (۲-۲) بدست می‌آید:

$$F = \frac{SS_{reg} / DF_{reg}}{SS_{res} / DF_{res}} \quad (2-2)$$

SS_{reg} : مجموع مجذورات رگرسیون

DF_{reg} : درجه آزادی رگرسیون

SS_{res} : مجموع مجذورات باقی مانده‌ها

DF_{res} : درجه آزادی باقی مانده‌ها

☀️ آماره‌ی R : این فاکتور، ضریب همبستگی بین مقدار پیش‌بینی شده و مقدار واقعی را نشان می‌دهد. مقدار این ضریب، بین صفر تا یک متغیر است. مقادیر کوچک R ، نشان‌دهنده‌ی ارتباط خطی کم (و گاهی صفر) بین متغیر وابسته و متغیرهای مستقل را نشان می‌دهد.

☀️ آماره‌ی R^2 : مجذور ضریب همبستگی چند متغیره‌ی R^2 ، واریانس مشترک بین متغیرهای وابسته و متغیرهای مستقل را نشان می‌دهد. فصل مشترک دو زیرمجموعه را به صورت مقداری بیان می‌کند. از آن جا که R^2 بخشی از واریانس متغیر وابسته را که توسط متغیرهای مستقل به حساب آمده معین می‌کند به R^2 ضریب تعیین نیز می‌گویند [۲۵].

☀️ آماره‌ی t : آزمون معنی‌داری حضور هر متغیر مستقل در معادله رگرسیون، آزمون t می‌باشد. مقدار t از تقسیم مقدار ضریب هر متغیر مستقل بر خطای استاندارد آن به دست می‌آید.

☀️ ضریب همبستگی میان توصیف‌کننده‌ها: این ضریب که ضریب همبستگی پیرسون^{۲۱} نیز نامیده می‌شود معیاری از ارتباط خطی بین دو متغیر مستقل را نشان می‌دهد. مقادیر ضریب همبستگی بین -۱ تا +۱ متغیر هستند که علامت (+) آن نشان‌دهنده جهت ارتباط خطی مستقیم و علامت (-) آن نشان‌دهنده‌ی ارتباط خطی غیرمستقیم و علامت قدرمطلق آن، بزرگی این همبستگی را نشان می‌دهد [۲۴].

به طور کلی مطالعات QSPR شامل پنج فعالیت عمده است:

- ۱- انتخاب سری مولکولی
 - ۲- محاسبه‌ی توصیف‌کننده‌ها
 - ۳- انتخاب توصیف‌کننده‌های مناسب
 - ۴- ایجاد مدل
 - ۵- استفاده از مدل جهت پیش‌بینی خاصیت ترکیبات جدید
- در ادامه، این پنج مورد توضیح داده خواهد شد.

²¹ Pearson Correlation

۲-۵-۱- انتخاب سری مولکولی

اولین مرحله در یک فعالیت QSPR، جمع‌آوری و انتخاب یک سری مولکولی است که خاصیت موردنظر آنها به طور تجربی در دسترس باشد. لازم به ذکر است که خاصیت موردنظر ترکیبات انتخاب شده باید در شرایط یکسان اندازه‌گیری شده باشد. بدیهی است که هر چه اطلاعات تجربی قابل دسترسی برای طراحی مدل بیشتر باشد، بدون تردید مدل کارایی بهتری خواهد داشت. به منظور ارزیابی قابلیت‌های پیش‌بینی یک مدل از طریق روش‌های اعتبارسنجی، مجموعه داده‌ها به ۳ دسته تقسیم می‌شود: سری آموزش (مجموعه‌ای که برای بهینه سازی استفاده می‌شود)، سری ارزیابی (مجموعه‌ای که برای سنجش پیش‌بینی مدل بهینه شده استفاده می‌شود) و سری تست (که یک مجموعه داده جدید برای انجام اعتبارسنجی اضافی برای مدل به دست آمده از سری آموزش است). بدیهی است که هرچه سری داده‌ها بزرگتر و متنوع‌تر باشد مدل حاصل از آن معتبرتر و قدرت پیش‌بینی بالاتری خواهد داشت.

۲-۵-۲- محاسبه‌ی توصیف‌کننده‌ها

توصیف‌کننده‌های مولکولی مقادیری هستند که ساختار و شکل مولکول را توصیف می‌کنند و به پیش‌بینی فعالیت و خصوصیات مولکول‌ها در آزمایش‌های پیچیده کمک می‌کنند [۲۶]. توصیف‌کننده‌های مولکولی اطلاعات شیمیایی مربوط به ساختار یک مولکول را طی یک فرایند منطقی و ریاضی به اعداد تبدیل می‌کنند. از این اعداد در تفسیر خواص مولکولی استفاده می‌شود.

ویژگی‌های توصیف‌کننده‌های مولکولی عبارتند از [۲۷]:

۱- ساده بودن

۲- مستقل بودن

- ۳- تفسیر ساختار مولکولی
- ۴- عدم همبستگی با سایر توصیف کننده‌ها
- ۵- تغییر منظم با تغییر تدریجی در ساختارها
- ۶- وابستگی صحیح به اندازه مولکولی
- ۷- تمایز بین ایزومرهای مختلف مولکول

انواع کاربردهای توصیف کننده‌های مولکولی عبارتند از:

- ۱- آنالیز و مدل‌سازی فعالیت ترکیبات دارویی
- ۲- ارزیابی خصوصیات شیمیایی
- ۳- مدل‌سازی و ارزیابی فعالیت مولکولی در سم‌شناسی
- ۴- پیش‌بینی خصوصیات، فعالیت و رفتارهای مولکول‌هایی که هنوز سنتز نشده‌اند

محاسبه‌ی توصیف کننده‌ها معمولاً با استفاده از محاسبات تئوری یا آزمایش‌های عملی انجام می‌گیرد. روش محاسبه‌ی توصیف کننده‌ها به نوع آنها وابسته است. توصیف کننده‌ها معمولاً به چند دسته‌ی کلی زیر تقسیم می‌شوند:

- توصیف کننده‌های کوانتوم شیمیایی^{۲۲}
- توصیف کننده‌های توپولوژی^{۲۳}
- توصیف کننده‌های هندسی^{۲۴}
- توصیف کننده‌های فیزیکوشیمیایی
- توصیف کننده‌های الکتروتوپولوژی

²² Quantum Chemical Descriptors

²³ Topological Descriptors

²⁴ Geometric Descriptors

در ادامه این توصیف‌کننده‌ها و روش محاسبه آنها توضیح داده خواهد شد.

۲-۵-۱- توصیف‌کننده‌های کوانتوم شیمیایی

این دسته از توصیف‌کننده‌ها که با استفاده از محاسبات کوانتومی به دست می‌آیند نشان‌دهنده‌ی خواص الکترونی مولکول‌ها و برهم‌کنش‌های بین آن‌ها هستند. با استفاده از روش‌های کوانتومی می‌توان کمیت‌های متعددی را که تعیین‌کننده میزان واکنش‌پذیری، شکل و خواص پیوندی مولکولی می‌باشند، تعیین نمود. این مقادیر عددی ممکن است به کل مولکول یا قسمتی از آن مربوط باشند. از آن‌جا که این نوع توصیف‌کننده‌های تئوری در بردارنده‌ی اطلاعات فیزیکو-شیمیایی شناخته شده‌ای هستند، استفاده از آن‌ها در مدل‌ها متداول است. در محاسبه‌ی توصیف‌کننده‌های کوانتوم شیمیایی هیچ‌گونه خطای معین یا نامعین مربوط به اندازه‌گیری‌های تجربی، وجود ندارد. با این حال خطای ناشی از فرضیهایی که برای تسریع محاسبات استفاده می‌شوند، اجتناب‌ناپذیر است. از آن‌جا که در تعیین این نوع توصیف‌کننده‌ها برای یک سری مولکول مشابه خطای محاسباتی به طور تقریبی برای همه ترکیبات سری ثابت است، لذا این خطا در بدست آوردن نتایج نهایی تأثیر قابل توجهی نخواهد داشت [۲۸]. انرژی، ممان دوقطبی و بار اتمی، نمونه‌هایی از این نوع توصیف‌کننده‌ها به‌شمار می‌آیند.

۲-۵-۲- توصیف‌کننده‌های توپولوژی

توصیف‌کننده‌های توپولوژی، توصیف‌کننده‌هایی هستند که اطلاعاتی راجع به ساختار ترکیب ارائه می‌دهند. این توصیف‌کننده‌ها بر اساس توپولوژی (ریخت یا ساختار کلی مولکولی) به دست می‌آیند. تعدادی از این توصیف‌کننده‌ها که بر اساس توپولوژی مولکول به دست می‌آیند عبارتند از: تعداد اتم‌ها، تعداد هترو اتم‌ها، تعداد حلقه‌ها، تعداد پیوندهای ساده و دوگانه ترکیب و غیره.

۲-۵-۲- توصیف‌کننده‌های هندسی

این دسته از توصیف‌کننده‌ها اطلاعاتی راجع به ساختار هندسی مولکول ارائه می‌دهند. حجم و سطح مولکول نمونه‌هایی از این نوع هستند که با استفاده از محاسبات کوانتومی به دست می‌آیند. این نوع توصیف‌کننده‌ها با استفاده از مختصات سه بعدی اتم‌های تشکیل دهنده مولکول محاسبه می‌شوند، بنابراین قبل از محاسبه این توصیف‌کننده‌ها، می‌بایست ساختمان مولکول بهینه شود. به دلیل استفاده از روش‌های کوانتومی در محاسبه‌ی توصیف‌کننده‌های هندسی، می‌توان آن‌ها را از نوع توصیف‌کننده‌های کوانتوم شیمیایی نیز به شمار آورد.

۲-۵-۲-۴- توصیف‌کننده‌های فیزیکوشیمیایی

توصیف‌کننده‌های فیزیکوشیمیایی اطلاعات ساختاری مولکول‌ها را ارائه می‌دهند. از جمله این توصیف‌کننده‌ها می‌توان به نقطه جوش و پارامتر حلالیت اشاره کرد.

۲-۵-۲-۵- توصیف‌کننده‌های الکتروتوپولوژی

این نوع توصیف‌کننده‌ها تلفیقی از توصیف‌کننده‌های الکترونی و توپولوژی هستند که مشخصات الکترونی مولکول را با توجه به ساختار و توپولوژی آن بیان می‌کنند. از جمله این توصیف‌کننده‌ها می‌توان به نسبت ممان دوقطبی به قطبش پذیری اشاره کرد.

۲-۵-۳- انتخاب توصیف‌کننده‌های مناسب

مهمترین بخش در ایجاد یک مدل خوب و کارآمد، انتخاب توصیف‌کننده‌های مناسب است. پس از محاسبه‌ی توصیف‌کننده‌ها، تعدادی از آنها به عنوان توصیف‌کننده‌های مناسب جهت ساخت مدل

انتخاب می‌شوند. این مرحله شامل یافتن توصیف‌کننده‌های حاوی اطلاعات مفید است به طوری که قدرت پیش‌بینی مدل در سطح قابل قبولی باشد.

برای این منظور، توصیف‌کننده‌ها به عنوان متغیرهای مستقل و خاصیت موردنظر به عنوان پارامتر وابسته در نظر گرفته می‌شود. سپس از بین توصیف‌کننده‌های موجود، توصیف‌کننده‌هایی که با متغیر وابسته همبستگی بیشتری دارند انتخاب می‌شوند. برای انتخاب توصیف‌کننده‌های مناسب باید به نکات زیر توجه کرد:

۱- توصیف‌کننده‌هایی که بیش از ۹۰٪ مقادیر آنها یکسان باشند حذف می‌شوند.

۲- از بین توصیف‌کننده‌های که همبستگی بالای ۰/۹ دارند، توصیف‌کننده‌هایی که با متغیر وابسته همبستگی کمتری دارند، حذف می‌شوند.

۳- توصیف‌کننده‌هایی که محاسبه و تفسیر آنها مشکل است حذف می‌شوند.

روشهای مختلفی برای انتخاب متغیرهای وابسته به عنوان توصیف‌کننده‌های مناسب وجود دارد که در ادامه چند مورد آنها آمده است [۲۹]:

۲-۵-۳-۱- روش ورود اجباری^{۲۵}:

در این روش همه‌ی توصیف‌کننده‌های محاسبه شده (متغیرهای مستقل) در مرحله مدل‌سازی به کار گرفته می‌شوند. به این ترتیب، اثرات مثبت یا منفی متغیرهای مستقل، برای وارد شدن در مدل و میزان همبستگی آنها با خاصیت مورد مطالعه (متغیر وابسته) در نظر گرفته نمی‌شود. از این رو برای ساخت مدل استفاده از این روش انتخاب متغیر توصیه نمی‌شود.

²⁵ Stepwise Regression Methods

۲-۵-۳-۲- روش حذف مرحله‌ای:

در این روش ابتدا تمامی توصیف‌کننده‌ها در ساخت مدل به کار گرفته می‌شوند. در مرحله‌ی بعد اثر حذف تک تک توصیف‌کننده‌ها از مدل مورد بررسی قرار می‌گیرد. به این ترتیب که اگر با حذف یک توصیف‌کننده آماری F و R^2 (ضریب همبستگی) به مقدار قابل توجهی تغییر نکند، توصیف‌کننده موردنظر حذف می‌شود و در غیر اینصورت به عنوان یک پارامتر موثر در ساخت مدل نگهداشته خواهد شد. در نهایت پس از حذف توصیف‌کننده‌ای اضافی، متغیرهای باقی‌مانده به عنوان توصیف‌کننده‌های منتخب در مرحله ساخت مدل مورد استفاده قرار می‌گیرند.

۲-۵-۳-۳- روش افزایش مرحله‌ای:

در این روش ابتدا ضریب همبستگی R تک تک توصیف‌کننده‌ها با خاصیت موردنظر، محاسبه می‌شود، سپس توصیف‌کننده‌ها بر اساس مقدار عددی R به صورت نزولی مرتب می‌شوند و توصیف‌کننده‌ای که بیشترین ضریب همبستگی را تولید کرده است به عنوان مبنا انتخاب می‌شود. بدیهی است که این توصیف‌کننده بیشترین ارتباط را با خاصیت موردنظر دارد. پس از آن توصیف‌کننده‌های بعدی به ترتیب وارد مدل می‌شوند و در هر مرحله، اثر ورود هر کدام مورد بررسی قرار می‌گیرد. اگر با ورود یک توصیف‌کننده آماری F و R به اندازه‌ی قابل توجهی افزایش یافت، توصیف‌کننده‌ی مورد بررسی نگه داشته می‌شود و در غیر اینصورت از آن چشم‌پوشی خواهد شد.

۲-۵-۴- ساخت مدل

هدف از ساخت مدل‌های ریاضی، برقراری ارتباط بین توصیف‌کننده‌ها و خاصیت فیزیکی-شیمیایی مورد بررسی است. پس از بسط و توسعه مدل برای سری آموزش، این مدل می‌تواند برای پیش‌بینی خاصیت موردنظر به کار گرفته شود. از مدل‌های عمومی و پرکاربرد در روش‌های QSPR، مدل

²⁶ Correlation Coefficient

رگرسیون خطی چندگانه (MLR) است. از دیگر روش‌های مدل‌سازی در مطالعات مربوط به QSPR می‌توان به شبکه عصبی و الگوریتم‌های ژنتیکی اشاره کرد.

۲-۵-۵- استفاده از مدل جهت پیش‌بینی خاصیت موردنظر

پس از محاسبه و انتخاب توصیف‌کننده‌ها برای سری مولکولی مورد مطالعه و آموزش مدل، می‌توان از مدل جهت پیش‌بینی این خاصیت برای سایر مولکول‌های مشابه استفاده نمود. برای انجام این کار کافی است توصیف‌کننده‌هایی که برای ساخت مدل استفاده شده‌اند، برای ترکیب جدید نظر محاسبه شوند. با به دست آوردن این توصیف‌کننده‌ها به عنوان متغیرهای مستقل و قراردادن آنها در مدل، خاصیت موردنظر ترکیب به عنوان خروجی (متغیر وابسته) به سادگی به دست می‌آید. به این ترتیب بدون آزمایش‌های عملی و صرف وقت و هزینه، می‌توان به خاصیت ترکیب موردنظر دست یافت.

۲-۶- نرم‌افزارهای مورد استفاده در این پژوهش

در این تحقیق از بسته‌های نرم‌افزاری 7 Hyperdem [۳۰] برای رسم ساختار ترکیبات مورد بررسی، Dragon 3.0 [۳۱] برای محاسبه‌ی توصیف‌کننده‌های مولکولی، SPSS 13 [۳۲] برای انجام آنالیزهای آماری رگرسیون‌های خطی چندگانه و MATLAB 2008a [۳۳،۳۴] به منظور برنامه نویسی شبکه‌ی عصبی مصنوعی استفاده شد که در ادامه به معرفی آنها پرداخته خواهد شد.

۲-۶-۱- نرم‌افزار 7 Hyperdem

با استفاده از این نرم‌افزار ابتدا شکل تقریبی مولکول روی صفحه نمایش کامپیوتر رسم می‌گردد. سپس مولکول توسط روش‌های مکانیک - کوانتومی بهینه می‌شود. روش‌های مکانیک - کوانتومی

موجود در این نرم‌افزار عبارتند از روش دینامیک مولکولی، روش‌های از آغازین^{۲۷} و روش‌های نیمه تجربی مثل AM1 و PM3. با استفاده از این نرم‌افزار می‌توان اطلاعاتی در زمینه‌ی زوایای پیوندی، طول پیوندها، زوایای پیچشی، بار اتم‌ها و غیره به دست آورد.

۲-۶-۲- نرم‌افزار Dragon 3.0

این نرم‌افزار توسط گروه QSAR و کمومتریکس دانشگاه میلانو طراحی شده است [۳۵] و از آن برای محاسبه‌ی توصیف‌کننده‌های مولکولی استفاده می‌کنند. این توصیف‌کننده‌های مولکولی به منظور ارزیابی روابط ساختار - فعالیت یا ساختار - ویژگی مولکولی به کار می‌روند. نرم‌افزار Dragon توانایی محاسبه بیش از ۱۴۹۷ توصیف‌کننده مختلف را دارا است که این توصیف‌کننده‌ها در ۱۸ دسته اصلی قرار گرفته‌اند. در جدول ۱-۱ فهرست این توصیف‌کننده‌ها آورده شده است.

جدول ۱-۱- فهرست توصیف‌کننده‌های محاسبه شده توسط نرم‌افزار Dragon

ردیف	نام توصیف‌کننده	ردیف	نام توصیف‌کننده
۱	توصیف‌کننده‌های زیر ساختاری	۱۰	توصیف‌کننده‌های هندسی
۲	توصیف‌کننده‌های توپولوژیکی	۱۱	توصیف‌کننده‌های RDF
۳	شمارنده‌های مولکولی مورس	۱۲	توصیف‌کننده‌های سه بعدی
۴	توصیف‌کننده‌های BCUT	۱۳	توصیف‌کننده‌های WHIM
۵	شاخص‌های بار	۱۴	توصیف‌کننده‌های GETAWAY
۶	خودارتباطی‌های دو بعدی	۱۵	گروه‌های عاملی
۷	توصیف‌کننده‌های بار	۱۶	اجزای میان اتمی
۸	شاخص‌های آروماتیسیته	۱۷	توصیف‌کننده‌های تجربی
۹	پروپرونده‌های مولکولی راندیک	۱۸	خصوصیات مولکولی

²⁷ Ab Initio Methodes

این نرم‌افزار نیاز به پرونده‌های ساختار مولکولی دارد که قبلاً توسط سایر نرم‌افزارهای مولکولی مثل Hyperdem تامین شده‌اند. Dragon تنها توصیف‌کننده‌های مولکولی را نمایش می‌دهد و هیچ آنالیز QSAR را انجام نمی‌دهد.

۲-۶-۳- نرم‌افزار spss 13

این نرم‌افزار، توانایی تحلیل کلی اطلاعات را دارد و جزء عمومی‌ترین نرم‌افزارهای آماری است. در این پروژه، از این نرم‌افزار جهت حذف متغیرهایی با همبستگی بالا، به دست آوردن بهترین مدل و ارزیابی و اعتبار مدل رگرسیون خطی چندگانه استفاده شده است.

۲-۶-۴- نرم‌افزار MATLAB

این نرم‌افزار کارایی بسیار بالایی در محاسبات تخصصی دارد. اولین نگارش آن مربوط به دانشگاه نیومکزیکو و استانفورد در سال ۱۹۷۰ است. با گذشت زمان MATLAB به عنوان یک نرم‌افزار قوی و در عین حال ساده برای کاربران تبدیل شده است. مدل‌سازی به روش شبکه عصبی مصنوعی از جمله کاربردهای مهم این نرم‌افزار محسوب می‌شود که در تحقیق حاضر مورد استفاده قرار گرفته است. اطلاعات ترکیبات به صورت ماتریس به عنوان ورودی به نرم‌افزار داده می‌شود. سپس در محیط MATLAB با استفاده از فرامین آرایه‌ها و امکانات ویژه آن، مدل‌سازی غیرخطی داده‌های فعالیت یا خاصیت ترکیبات امکانپذیر است.

سایر روش‌های کموتریکس شامل آنالیز (رگرسیون) اجزای اصلی، روش کمترین مربعات جزئی، الگوریتم ژنتیکی [۳۶] و غیره نیز توسط این نرم‌افزار قابل اجرا است.

۳-۱- شبکه عصبی

در چنددهه‌ی اخیر، اندیشه‌ی شبیه‌سازی مغز انسان، دانشمندان را بر آن داشته است که در مسیرانتقال تواناییهای مغز به رایانه گام بردارند. زیرا بعد از هزاران سال تکامل، مغز می‌تواند وقایع پیرامون خود را تجزیه و تحلیل کند. دانشمندان با بهره‌گیری از یافته‌های علمی درباره اصول و ساز و کارهای محاسباتی مغز انسان که عملکرد بسیار سریع و در عین حال دقیقی دارد، شبکه‌های عصبی مصنوعی^{۲۸} (ANN) را به عنوان یک مدل بسیار ساده و ابتدایی از مغز، ارائه کردند. همان‌طور که مغز انسان می‌تواند با استفاده از تجربیات قبلی مسائل جدید را تجزیه و تحلیل کند، شبکه‌های عصبی مصنوعی نیز در صورت آموزش قادر به ارائه‌ی جواب مناسب برای مسائل جدیدی، که قبلاً با آن روبرو نشده است، می‌باشد [۳۷]. البته رسیدن به ماشینی که بتواند با مغز انسان رقابت کند تقریباً محال است. اما اگر یک سامانه بتواند تنها درصد کمی از کارایی مغز را داشته باشد، توانایی فوق العاده‌ای در پردازش سیگنال^{۲۹} (شکل، صحبت و اطلاعات)، کنترل (رباتها، سامانه‌های قدرت^{۳۰} و وسایل نقلیه هوشمند) و بسیاری از زمینه‌های دیگر خواهد داشت [۳۸]. شبکه‌های عصبی مصنوعی جزء دسته‌ای از سیستم‌های دینامیکی هستند که با پردازش روی مجموعه‌ای از ورودی‌ها و خروجی‌های صحیح، دانش نهفته در ورای آن‌ها را پیدا می‌کنند. به همین دلیل به این سیستم‌ها هوشمند^{۳۱} گفته می‌شود زیرا بر اساس محاسبات روی داده‌های عددی یا مثال‌ها قوانین کلی را فرا می‌گیرند [۳۹] و براساس آن متغیرهای خود را به گونه‌ای تنظیم می‌کنند که در صورت گرفتن ورودی‌های جدید، بتوانند پاسخ درستی ایجاد کنند. از شبکه‌های عصبی مصنوعی در زمینه‌های مختلف استفاده شده است. به عنوان مثال از کاربردهای شبکه‌ی عصبی مصنوعی در زمینه‌ی شیمی می‌توان به تشخیص گروه‌های عاملی و تفکیک طیف‌ها [۴۰]، پیش‌بینی خصوصیات آشکارساز یونیزاسیون شعله [۴۱]، پیش‌بینی جابه‌جایی

²⁸ Artificial Neural Networks

²⁹ Signal Processing

³⁰ Power Systems

³¹ Intelligent

شیمیایی کربن در طیف $^{13}\text{C}-\text{NMR}$ [۴۱] و اندازه‌گیری نمونه‌های مختلف به روش اسپکتروفتومتری [۴۲،۴۳] اشاره کرد. در مورد خواص انتقالی سیالات نیز شبکه‌های عصبی مصنوعی بکار گرفته شده اند که به دو مورد پیش‌بینی گرانشی مایعات [۴۴] و پیش‌بینی گرانشی گازها در محدوده‌ی وسیعی از دما و فشار [۴۰] می‌توان اشاره کرد. در این تحقیق هدف تعیین و پیش‌بینی دانسیته‌ی کتون‌ها با استفاده از توصیف‌کننده‌های مولکولی در محدوده‌ی وسیعی از دما و فشار است.

۲-۳- تاریخچه شبکه‌های عصبی مصنوعی

برخی از پیش‌زمینه‌های شبکه‌های عصبی مصنوعی را می‌توان در اواخر قرن نوزدهم و اوایل قرن بیستم جستجو کرد. در این دوره کارهای اساسی در فیزیک، روانشناسی و فیزیولوژی عصبی توسط دانشمندانی چون هرمان فون هلمهولتز^{۳۲}، ارنست ماخ^{۳۳} و ایوان پاولف^{۳۴} انجام گرفت. در دهه‌ی چهل قرن بیستم دیدگاهی جدید راجع به شبکه‌های عصبی پدیدار شد. در سال ۱۹۴۳ وان مک کلوت^{۳۵} و والتر پیتز^{۳۶} [۴۵] با استفاده از یک مدل منطقی که امروزه بلوک اصلی سازنده‌ی اکثر شبکه‌های عصبی است، نشان دادند که شبکه‌های عصبی می‌توانند هر تابع حسابی و یا منطقی را محاسبه کنند [۴۶]. این مدل فرضیه‌هایی در مورد عملکرد نرون‌ها ارائه می‌کرد اما محدودیت‌های تکنولوژی در آن زمان اجازه کار بیشتر به آن‌ها را نداد. در همان دوران، دونالد هب^{۳۷} عمل شرط‌گذاری کلاسیک را که توسط پاولف مطرح شده بود به عنوان خاصیت نرون‌ها معرفی کرد و سپس مکانیسمی را برای یادگیری نرون‌های زیستی پیشنهاد کرد [۴۷]. نخستین کاربرد عملی شبکه‌های مصنوعی در دهه‌ی

³² Hermann Von Helmholtz

² Ernest Mach

³ Ivan Pavlov

⁴ Warren Mcculloch

⁵ Walter Pitts

⁶ Donald Hebb

قرن بیستم مطرح شد. فرانک روزنبلات^{۳۸} در سال ۱۹۵۸ شبکه پرسپترون^{۳۹} را معرفی کرد و به کمک آن توانست الگوها را از هم شناسایی کند. پرسپترون دارای سه لایه می‌باشد که لایه میانی به عنوان لایه پیوند شناخته می‌شود. سپس برنارد ویدرو^{۴۰} و تد هوف^{۴۱} در سال ۱۹۶۰ شبکه عصبی تطبیقی خطی (آدالاین)^{۴۲} را با قانون یادگیری جدید مطرح کردند که از لحاظ ساختار و قابلیت شبیه شبکه پرسپترون بود [۴۷]. هر دوی این شبکه‌ها، پرسپترون و آدالاین، محدودیت داشتند و فقط قادر به طبقه‌بندی الگوهایی بودند که به طور خطی متمایز می‌شدند. علت این بود که آن‌ها قانون یادگیری را برای شبکه‌های عصبی تک لایه‌ای مطرح کرده بودند که توانایی محدودی جهت تخمین توابع داشتند. آن‌ها توانستند شبکه‌های چند لایه‌ای را مطرح کنند اما نتوانستند الگوریتم‌های یادگیری شبکه‌های چند لایه‌ای را بهبود بخشند. این محدودیت‌ها در سیستم‌های تک لایه و چند لایه در سال ۱۹۶۹ توسط مینسکی^{۴۳} و پاپرت^{۴۴} در کتابی مورد بررسی قرار گرفت. در سال ۱۹۷۲ تئو کوهونن^{۴۵} و جیمز اندرسون^{۴۶} به طور مستقل و بدون اطلاع از یکدیگر، شبکه‌های عصبی جدیدی را مطرح کردند که قادر به ذخیره‌سازی اطلاعات بودند. استفان گروسبرگ^{۴۷} در سال ۱۹۷۶ روی شبکه‌های خود سازمانده کار می‌کرد. پیشرفت‌های شبکه عصبی تا دهه‌ی هشتاد قرن بیستم ادامه یافت. در دهه‌ی هشتاد ایده به‌کارگیری شبکه‌ها در زمینه‌های جدیدی مطرح شد. دهه‌ی هشتاد قرن بیستم را می‌توان زمان تولد دوباره شبکه‌های عصبی دانست چرا که در این دوران دو کار بسیار بزرگ صورت گرفت:

³⁸ Frank Rosenblat

⁸ Perceptron

⁴⁰ BernurdWidrow

⁴¹ Ted Hoff

⁴² ADAptiveLINear Element (ADALINE)

⁴³ Minsky

⁴⁴ Papert

⁴⁵ TeoKohonen

⁴⁶ James Anderson

⁴⁷ Stefan Grossberg

۱- در سال ۱۹۸۲ فیزیکدان آمریکایی جان هاپفیلد^{۴۸} شبکه‌های چندلایه و الگوریتم یادگیری دارای بازخورد را معرفی کرد. جان هاپفیلد از مکانیک آماری برای توضیح عملکرد شبکه استفاده کرد و روشی را برای حل مسائل غیر خطی ارائه کرد. از این شبکه‌ها، برای ذخیره سازی اطلاعات استفاده می‌شود.

۲- در سال ۱۹۸۶ دیوید راملهارت^{۴۹} و جیمز مک‌کلند^{۵۰} الگوریتم پس انتشار را مطرح کردند و از آن زمان تا کنون این زمینه علمی روند رشد و پیشرفت را ادامه داده است که البته بیشتر این پیشرفت‌ها به ساختارهای نوین و روش‌های یادگیری جدید مربوط می‌شود [۴۸].

پس از آن‌زمان کاربردهای شبکه‌ی عصبی مصنوعی در شیمی و علوم وابسته به آن مانند بیوشیمی، مهندسی شیمی و داروسازی رواج پیدا کرد [۴۹]. مسلماً در آینده نیز که ساختار مغز بیشتر شناخته خواهد شد قابلیت‌های بیشتری به شبکه‌های عصبی مصنوعی افزوده خواهد شد و این شبکه‌ها خواهند توانست به عنوان یک ابزار علمی برای حل مسائل جدید آن زمان مطرح شوند [۳۹،۴۷].

۳-۳- ساختار شبکه‌های عصبی زیستی

شبکه‌های عصبی مصنوعی شباهت ناچیزی به هم‌تایان زیستی خود، یعنی شبکه‌های عصبی مصنوعی زیستی و در راس آنها، مغز دارند. در اینجا فقط ویژگی‌هایی از عملکرد مغز، که برای توسعه شبکه‌های عصبی مصنوعی الگوبرداری شده اند، مورد بررسی قرار می‌گیرد. مغز از عناصری به نام نرون^{۵۱} تشکیل شده است. هر نرون می‌تواند به سه جزء اصلی تقسیم‌بندی شود: دندریت^{۵۲}، بدنه‌ی

⁴⁸ John Hopfield

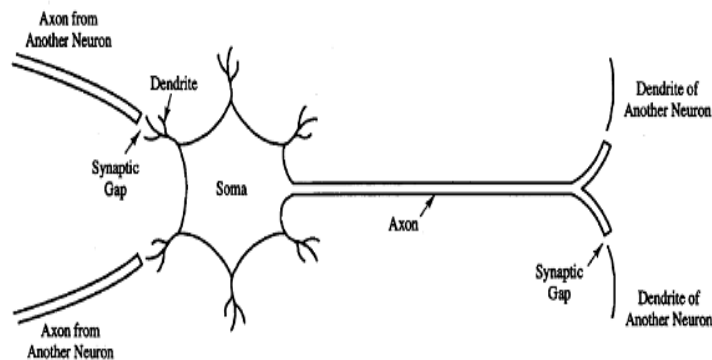
⁴⁹ David Rummelhart

⁵⁰ James Mcland

⁵¹ Neuron

⁵² Dendrite

سلول^{۵۳} و آکسون^{۵۴} (شکل (۳-۱)). دندریته‌ها گیرنده‌های نرون هستند که مثل شاخه‌های درخت دارای انشعاب‌های فراوان بوده و درون آنها سیگنال‌های الکتریکی جریان دارد تا به بدنه‌ی سلول برسد. بدنه‌ی سلول، انجام عملیات لازم روی اطلاعات ورودی را بر عهده دارد که به‌طور ساده می‌توان آن را شبیه جمع‌زدن اطلاعات ورودی دانست. آکسون یک رشته طولانی است که اطلاعات خروجی را از بدنه‌ی سلول گرفته و به خارج نرون می‌فرستد. محل اتصال آکسون یک سلول و دندریته سلول دیگر سیناپس^{۵۵} نام دارد. نحوه‌ی پیکربندی نرون‌ها و قدرت سیناپس‌های ویژه‌ی آن‌ها، که توسط یک فرایند الکتروشیمیایی پیچیده تعیین می‌شود، عملکرد شبکه‌ی عصبی را مشخص می‌کند [۴۷].



شکل ۳-۱- نمایشی از دو نرون زیستی

۳-۴- مقایسه مغز و رایانه

مغز در انجام فعالیت‌هایی مثل دیدن و شنیدن و یادگیری از طریق تجربه، کارآیی چشمگیری دارد. روند تکامل مغز متأثر از فعالیت‌هایی بوده که اهمیت بیشتری داشته است و از آن‌جا که دیدن و شنیدن، نسبت به جمع‌کردن دقیق اعداد اهمیت بیشتری داشته و دارد، مغز در این جنبه تکامل بیشتری پیدا کرده است. البته این فعالیت‌های به ظاهر ساده نیاز به پردازش اطلاعات عظیمی دارد که به علت ساختار موازی مغز به راحتی انجام می‌شود به این صورت که واحدهای محاسباتی ساده

⁵³ Soma

⁵⁴ Axon

⁵⁵ Synapse

(نرون‌ها) به همراه یکدیگر و به صورت مشترک این پردازش را انجام می‌دهند. این تقسیم‌کار علاوه بر اینکه باعث می‌شود یک فرایند پیچیده مثل دیدن یا شنیدن براحتی اتفاق بیفتد، مزایای دیگری نیز دارد، که از آن جمله می‌توان به موارد زیر اشاره نمود:

۱- به این دلیل که تعداد نرون‌های درگیر در یک فعالیت زیاد است، سهم هریک از آنها به تنهایی حائز اهمیت نیست. بنابراین اگر یکی، راه خطا در پیش‌گیرد نتیجه‌ی آن تاثیر چندانی بر نتیجه‌ی کلی ندارد. این ویژگی پردازش توزیع شده^{۵۶} نام دارد و مزیت آن قابل اغماض بودن لغزش‌های احتمالی اجزاء است.

۲- توان انجام فعالیت در حالی که فقط تعدادی از نرون‌ها به درستی کار می‌کنند تحمل خطا^{۵۷} نام دارد و به علت داشتن این ویژگی مغز بون توجه به از دست دادن تعدادی از نرون‌ها (به دلایل مرگ طبیعی) به فعالیت خود ادامه می‌دهد. اما در صورت ادامه مرگ نرون‌ها و جایگزین نشدن آن‌ها به اصطلاح، فرسایش تدریجی اتفاق می‌افتد و عملکرد مغز به آهستگی تنزل پیدا می‌کند. اما در مورد رایانه‌ها به جای چندین واحد پردازش نسبتاً کند و متصل به یکدیگر، یک یا چند واحد پردازش بسیار سریع به کار گرفته شده که قادر است میلیون‌ها محاسبه را در هر ثانیه انجام دهد. به همین دلیل انجام عملیات ریاضی روی اعداد، توسط رایانه‌ها بسیار سریع و راحت است. در مقابل رایانه‌ها از انجام کارهایی که نیاز به پردازش انواع گوناگون داده‌ها به صورت موازی دارند عاجزند (مثل دیدن). همچنین به علت عدم توانایی در پردازش توزیع شده، رایانه‌ها توانایی تحمل خطا و اغماض در مقابل خطاها را ندارند و اگر واحد پردازش یک رایانه از کار بیفتد، رایانه در انجام مسئولیتی که بر عهده‌ی آن گذاشته شده، ناتوان می‌ماند [۵۰].

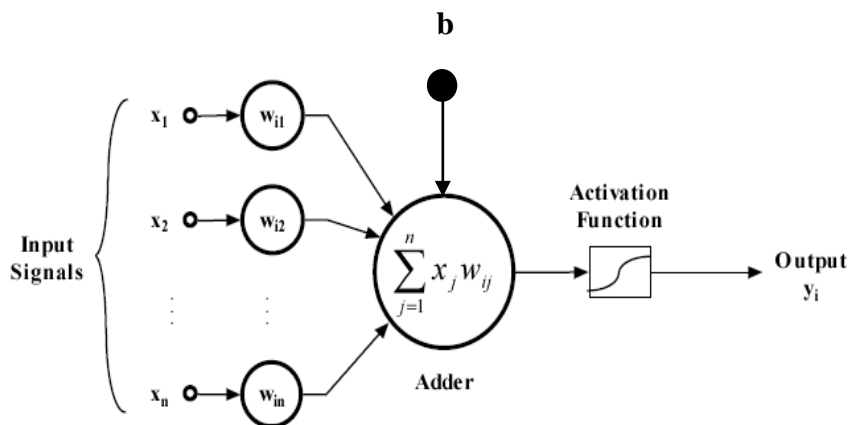
⁵⁶ Distributed Processing

⁵⁷ Fault Tolerance

این مسائل موجب بروز تمایل به ایجاد رایانه‌هایی شده است که از عناصر ساده‌تر با اتصالات پر تعداد بین عناصر، تشکیل شده باشند تا وظایف محوله را به صورت مشترک انجام دهند و بیشتر به مغز شبیه باشند تا به رایانه و در این راه استفاده از شبکه‌های عصبی مصنوعی، مناسب است.

۳-۵- ساختار شبکه عصبی مصنوعی

مهمترین جزء تشکیل دهنده هر شبکه‌ی عصبی مصنوعی، سلول عصبی مصنوعی یا نرون است. نرون عملکرد ساده‌ای دارد. این عملکرد عبارتست از: دریافت ورودی‌ها، وزن‌دهی به آن‌ها و محاسبه مجموع آن‌ها به عنوان سیگنال تحریک در مرحله اول و سپس اعمال تابع تبدیل بر روی آن‌ها و در نهایت ساختن یک خروجی در مرحله دوم. در متون فنی برای نمایش مدل ساده‌ای از نرون محاسباتی از شکلی مشابه شکل (۳-۲) استفاده می‌کنند.



شکل ۳-۲- ساختمان یک نرون محاسباتی با بردار ورودی دارای R عضو [۹]

هر ورودی در وزن متناظرش ضرب شده و به اصطلاح وزن‌دار می‌شود، سپس تمام ورودی‌های وزن‌دار با هم جمع می‌شوند. ورودی دیگری با عنوان b در نظر گرفته می‌شود که به معنای سیگنالی

به اندازه‌ی واحد است که در بایاس^{۵۸} ضرب شده و با حاصلضرب ورودی‌ها در وزن‌ها جمع می‌شود تا ورودی خالص را ایجاد کند. سپس تابع محرک روی آن اثر کرده و خروجی نرون را ایجاد می‌کند [۳۷]. بایاس شبیه یک وزن است با این تفاوت که بایاس روی ورودی ثابت یک، ۱، اعمال می‌شود و حسن آن افزودن یک متغیر اضافی به شبکه است. به همین دلیل شبکه‌ای که بایاس دارد نسبت به شبکه‌ی فاقد بایاس قوی‌تر است. اگر نرون دارای R ورودی باشد می‌توان ورودی‌ها را به صورت بردار $1 \times R$ نشان داد که هر یک از ورودی‌ها در یک عضو از ماتریس وزن (w) ضرب شده و حاصل با بایاس جمع می‌شود تا ورودی خالص (n) را تشکیل دهد.

$$n = WX + b = W_{i,1}X_1 + W_{i,2}X_2 + \dots + W_{i,n}X_n + b \quad (۱-۳)$$

تابع محرک (f) روی ورودی خالص (n) اثر می‌کند تا خروجی (a) را بدهد (در مورد یک نرون تنها، ماتریس وزن‌ها (w) فقط یک ردیف دارد).

$$y = f(n) = f(wx + b) \quad (۲-۳)$$

۳-۶- توابع محرک

توابع محرک می‌توانند نسبت به n خطی یا غیرخطی باشند. توابع محرک مورد استفاده در شبکه‌های عصبی بر حسب کاربردشان، توابع متفاوتی هستند. معروفترین توابع محرک عبارتند از: تابع محرک آستانه‌ای حدی^{۵۹}، تابع محرک خطی^{۶۰} و تابع محرک سیگموئید^{۶۱}.

⁵⁸ Bais

⁵⁹ Hard limit transfer function

⁶⁰ Linear Transfer Function

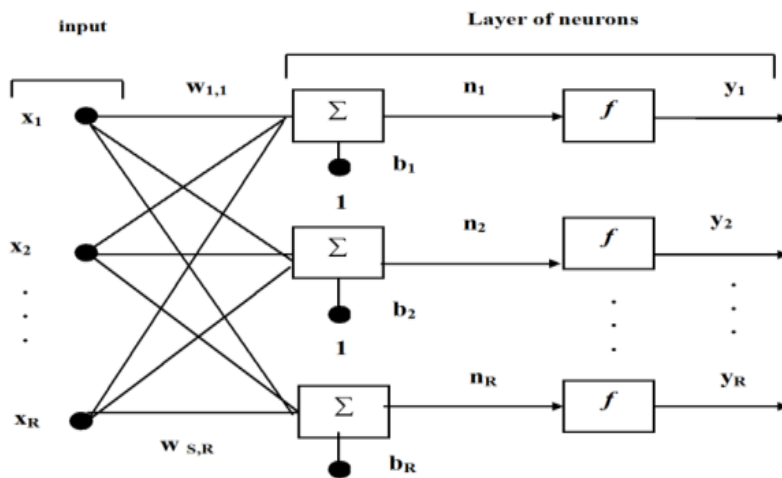
⁶¹ Logarithmic Sigmoid Transfer Function

۳-۷- تقسیم بندی لایه‌های شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی به صورت تک لایه یا چندین لایه می‌باشند.

۳-۷-۱- شبکه‌ی عصبی تک‌لایه

معمولاً یک نرون، حتی با تعداد زیاد ورودی، به تنهایی برای حل مسائل گوناگون کافی نیست، بنابراین باید چند نرون به طور موازی به کار گرفته شوند که آرایش چند نرون به صورت موازی یک لایه را تشکیل می‌دهند. شکل (۳-۳) یک شبکه تک‌لایه‌ای با R عنصر ورودی و S نرون را نشان می‌دهد.



شکل ۳-۳- ساختار یک شبکه‌ی تک‌لایه با R ورودی و S نرون [۹]

در این شبکه هر عنصر از بردار ورودی (P) به تمام نرون‌ها متصل است و به ازای وزن‌های متعلق

به این ورودی‌ها یک ستون از ماتریس وزن‌ها شکل می‌گیرند. بنابراین ماتریس وزن‌ها چنین است:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdot & \cdot & \cdot & w_{1R} \\ w_{21} & w_{22} & \cdot & \cdot & \cdot & w_{2R} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ w_{S1} & w_{S2} & \cdot & \cdot & \cdot & w_{SR} \end{bmatrix}_{S \times R} \quad (3-3)$$

همچنین هر نرون دارای یک بایاس (b_i) است که بردار b بصورت زیر است:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_s \end{bmatrix}_{s \times 1} \quad (4-3)$$

البته نرون‌ها هر کدام ورودی خالص، تابع محرک و خروجی مخصوص به خود را دارند. در مسائل مختلف، شبکه ماتریس‌های W و b را تغییر می‌دهد تا مقدار بهینه به دست آید [۴۷].

۳-۷-۲- شبکه عصبی چند لایه

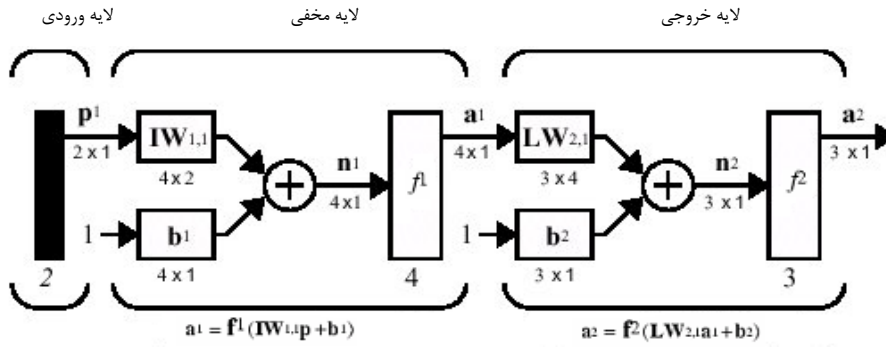
هنگامی که (به علت دینامیک و قوانین شبکه‌های عصبی) به پارامترهای قابل تنظیم بیشتری نیاز است از شبکه‌های چندلایه استفاده می‌شود. در یک شبکه چند لایه‌ای، هر لایه ماتریس وزن w ، بردار بایاس b ، بردار ورودی خالص n و بردار خروجی a مخصوص به خود را دارد که برای مشخص کردن لایه‌های مختلف از بالانویس^{۶۲} استفاده می‌شود [۳۹]. هر شبکه چندلایه‌ای شامل لایه ورودی^{۶۳}، لایه خروجی^{۶۴} و یک یا چند لایه پنهان^{۶۵} یا میانی می‌باشد. در این توپولوژی، نرون‌ها علاوه بر این که در لایه خود به شکل محدودی به یکدیگر اتصال داده شده‌اند، از طریق اتصال بین لایه‌ها نیز با یکدیگر ارتباط دارند. در این حالت گروهی از نرون‌ها از طریق ورودی‌های خود با جهان واقعی ارتباط دارند، گروهی دیگر از نرون‌ها نیز از طریق خروجی‌های خود، جهان خارج را می‌سازند که لایه-ی خروجی نام دارند و بقیه‌ی نرون‌ها که از دید پنهان هستند لایه‌های پنهان را تشکیل می‌دهند [۵۱]. در شکل (۴-۳) یک شبکه سه لایه نشان داده شده است

⁶² Superscript

⁶³ Input ILayer

⁶⁴ Output Layer

⁶⁵ Hidden Layer



شکل ۳-۴- ساختار یک شبکه سه لایه

با توجه به شکل (۳-۴)، یک شبکه سه لایه با R ورودی و K نرون در هر لایه را نشان می‌دهد. همان‌طور که در این شکل می‌بینید خروجی‌های هر لایه، ورودی لایه بعدی خواهد بود و لایه‌ای که خروجی آن همان خروجی شبکه است، لایه خروجی نام دارد.

همان‌طور که قبلاً توضیح داده شد، در شبکه‌های تک لایه باید مقادیر W و b بهینه شوند، اما در شبکه‌های چند لایه علاوه بر آنها باید تعداد لایه‌های مخفی و تعداد نرون‌ها در لایه مخفی نیز توسط کاربر بهینه شوند. که همین امر باعث افزایش حجم محاسبات و در نتیجه افزایش مدت زمان آموزش خواهد شد.

در بسیاری از شبکه‌های عصبی، اتصالات بین نرون‌ها به گونه‌ای است که نرون‌های لایه میانی، ورودی خود را از تمام نرون‌های لایه پایینی خود دریافت می‌کنند بدین ترتیب، سیگنال‌ها به تدریج از یک لایه نرونی به لایه‌های بعدی حرکت می‌کنند و در نهایت به لایه آخر یا لایه خروجی شبکه می‌رسد. چنین سیری در شبکه‌ی عصبی به پیش‌خور^{۶۶} معروف است. در نوع دیگری از ارتباط بین نرون‌ها، خروجی یک نرون به همان نرون یا نرون‌های دیگر در همان لایه یا لایه‌های قبلی متصل می‌شود که چنین ارتباطی در شبکه‌های عصبی پس‌خور^{۶۷} وجود دارد. ارتباطات بین نرونی در شبکه‌ی عصبی از اهمیت بسیار زیادی برخوردار است و به نوعی قدرت یک شبکه‌ی عصبی را تعیین می‌کنند.

^{۶۶} Feed forward

^{۶۷} Feed back

۳-۸- آموزش شبکه‌های عصبی

پس از طراحی هر شبکه عصبی، باید پارامترهای b و W به ازای مجموعه‌ای از سیگنال‌های ورودی به گونه‌ای تنظیم شوند که سیگنال‌های خروجی شبکه خروجی مطلوبی را تشکیل دهند. چنین فرایندی آموزش شبکه‌ی عصبی نامیده می‌شود. برای آموزش دادن شبکه‌های عصبی دو روش کلی وجود دارد:

۱- یادگیری با ناظر^{۶۸} ۲- یادگیری بدون ناظر^{۶۹}

در یادگیری با ناظر، قاعده یادگیری توسط مجموعه‌ای از مثال‌ها برای ایجاد شبکه‌ی مناسب فراهم می‌شود. در این روش ورودی‌ها و خروجی‌های واقعی منطبق با این ورودی‌ها در اختیار شبکه قرار می‌گیرند. زمانی که ورودی‌ها برای شبکه به کار می‌روند، خروجی‌های محاسبه شده توسط شبکه با مقادیر واقعی مقایسه می‌شوند. سپس قاعده یادگیری برای تنظیم وزن‌ها و بایاس‌های شبکه جهت نزدیکی هرچه بیشتر خروجی شبکه به مقادیر واقعی استفاده می‌شود. در روش یادگیری بدون ناظر، خود شبکه‌ی عصبی باید به تنهایی و بدون کمک داده‌های خروجی، کار آموزش را انجام دهد. در این روش، شبکه می‌آموزد که الگوهای ورودی را به تعداد متناهی از گروه‌ها تقسیم‌بندی کند [۵۲]. پس از آن که یک شبکه‌ی عصبی به اندازه‌ی کافی آموزش دید، طراح یا کاربر شبکه می‌تواند پارامترهای شبکه را قفل کند. در این مرحله شبکه‌ی عصبی برای استفاده و حل مسائل جدید آماده خواهد بود.

۳-۹- روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی

اصول بهینه‌سازی الگوریتم‌ها در طی قرن هفدهم توسط ریاضی دانانی نظیر کپلر^{۷۰}، فرمات^{۷۱}،

⁶⁸ Supervised learning

⁶⁹ Unsupervised learning

⁷⁰ Kepler

⁷¹ Fermat

نیوتن^{۷۲} و لایبنیز^{۷۳} کشف شد. سپس از سال ۱۹۵۰ به بعد این اصول مجدداً به کار گرفته شدند و محققان شبکه‌های عصبی نیز از ماحصل تلاش ریاضی دانان در این زمینه برای آموزش شبکه‌ها استفاده کردند.

در تمامی الگوریتم‌های بهینه‌سازی یک مقدار اولیه برای تابع حدس زده می‌شود و سپس این مقدار طی مراحل که از قاعده زیر پیروی می‌کند بهینه می‌شود:

$$X_{k+1} = X_k + \alpha P_k \quad (5-3)$$

در رابطه فوق، X_k تخمین نقطه مینیمم تابع در مرحله k ام است و X_{k+1} تخمین مرحله بعدی نقطه مینیمم تابع می‌باشد و P_k نیز بردار جستجو می‌باشد، بدین صورت که از روی آن، تخمین مرحله بعد به دست می‌آید. α یک مقدار اسکالر بین صفر و یک است و سرعت یادگیری نامیده می‌شود. انتخاب P_k و α روش‌های مختلف مینیمم‌سازی را از یکدیگر جدا می‌کند که در زیر به برخی از آن‌ها اشاره می‌شود.

۳-۹-۱- الگوریتم نزول گرادیانی^{۷۴}

روش نزول گرادیانی یکی از ابتدایی‌ترین و در عین حال موثرترین تکنیک‌ها در آموزش شبکه است. تابع n متغیره $F(x)$ را به صورت زیر در نظر بگیرید:

$$F(x) = F(x_1, x_2, \dots, x_n) \quad (6-3)$$

بردار گرادیان برای این تابع به صورت زیر تعریف می‌شود:

$$\nabla F(x) = \left[\frac{\partial}{\partial x_1} F(x) \quad \frac{\partial}{\partial x_2} F(x) \quad \dots \quad \frac{\partial}{\partial x_n} F(x) \right]_{n \times 1}^T \quad (7-3)$$

⁷² Newton

⁷³ Leibniz

⁷⁴ Gradient Descent Algorithm

در این رابطه T ، ترانهاده‌ی ماتریس نامیده می‌شود و باعث جابه‌جایی سطر و ستون‌های ماتریس می‌شود. گرادیان تابع خاصیت جالبی دارد که کاربرد منحصر به فردی برای آن به وجود آورده است. برای استفاده از این خاصیت یک تابع عملکرد یا تابع خطای E تعریف می‌شود که بیانگر اختلاف بین مقادیر واقعی و پاسخ شبکه است. تابع انتخاب شده بایستی نسبت به تک تک پارامترهای شبکه مشتق پذیر باشد. در این صورت گرادیان تابع خطا در هر نقطه از فضای پارامترها قابل محاسبه خواهد بود.

اگر E تابع خطا باشد، گرادیان آن بصورت زیر تعریف می‌شود:

$$\nabla E(W)^T \Big|_{W=W_k} = \left[\frac{\partial}{\partial w_1} E(W) \quad \frac{\partial}{\partial w_2} E(W) \quad \dots \quad \frac{\partial}{\partial w_n} E(W) \right]_{n \times 1}^T \Big|_{W=W_k} \quad (۸-۳)$$

W برداری است که مولفه‌هایش پارامترهای شبکه هستند.

از طرف دیگر زمانی که از رابطه (۳-۶) استفاده می‌شود باید تابع خطا در هر تکرار کاهش یابد بدین معنی که:

$$E(W_{k+1}) < E(W_k) \quad (۹-۳)$$

بسط تیلور مرتبه اول تابع E_k حول نقطه W_k بصورت رابطه (۳-۱۰) است:

$$E(W_{k+1}) = E(W_k) + \nabla E(W)^T \Big|_{W=W_k} \Delta W_k \quad (۱۰-۳)$$

برای اینکه رابطه (۳-۹) برقرار باشد باید:

$$\nabla E(W)^T \Big|_{W=W_k} \Delta W_k = g_k^T \Delta W_k < 0 \quad (۱۱-۳)$$

اگر رابطه (۳-۵) را بصورت زیر بازنویسی کنیم:

$$W_{k+1} = W_k + \alpha P_k \Rightarrow \Delta W_k = \alpha P_k \quad (۱۲-۳)$$

با قراردادن رابطه (۳-۱۲) در رابطه (۳-۱۱) به دست می‌آید:

$$\alpha \Delta E(W)_k = \alpha g_k^T p_k < 0 \quad (13-3)$$

و از آنجا که ضریب یادگیری α ، یک مقدار مثبت است، در نتیجه داریم:

$$g_k^T p_k < 0 \quad (14-3)$$

که این نامساوی جهت نزول تابع $E(W)$ را تعریف می‌کند.

هر بردار p_k که در این معادله صدق می‌کند، مسیر نزول^{۷۵} نام دارد و مقدار تابع با حرکت در این مسیر کاهش می‌یابد. مسیری که $g_k^T p_k$ مربوط به آن منفی‌تر از بقیه باشد باعث بیشترین نزول در مقدار تابع می‌شود.

زمانی که بردار مسیر و بردار گرادیان تابع در جهت عکس یکدیگر باشند، مقدار عبارت $g_k^T p_k$ حداقل می‌شود به این دلیل که معادله (۱۴-۳) حاصل ضرب درونی دو بردار می‌باشد.

بنابراین برداری که مسیر سریع‌ترین نزول را نشان می‌دهد، چنین است:

$$p_k = -g_k \quad (15-3)$$

طبق رابطه (۱۵-۳) اگر در جهت گرادیان حرکت کنیم، مقدار تابع در هر نقطه از فضای پارامترها، به سرعت افزایش می‌یابد و بر عکس اگر در خلاف جهت گرادیان حرکت کنیم، مقدار تابع با همان سرعت کاهش می‌یابد. در نتیجه رابطه (۱۵-۳) به‌گونه‌ای پارامترهای شبکه را تنظیم می‌کند که باعث مینیمم شدن تابع خطا می‌گردد [۴۷]

با جایگذاری رابطه (۱۵-۳) در (۱۲-۳) به‌دست می‌آید:

$$W_{k+1} = W_k - \alpha g_k = W_k - \alpha \nabla E(W) \Big|_{W=W_k} \quad (16-3)$$

در این الگوریتم سرعت یادگیری α نقش حیاتی دارد. انتخاب مقادیر بزرگ α با وجود آموزش سریع شبکه و اتلاف کمتر زمان، در بسیاری موارد باعث ناپایداری آموزش شبکه شده و نتایج مناسبی حاصل نمی‌گردد. از طرف دیگر، انتخاب مقادیر کوچک α باعث افزایش مدت زمان آموزش شبکه

⁷⁵ Descent Direction

می‌شود اما مسیر تغییرات تابع هموارتر خواهد بود. لذا برای داشتن یک ضریب یادگیری مناسب می‌توان از روش‌های زیر استفاده کرد:

۱- د

رابطه ابتدای آموزش شبکه، می‌توان از مقادیر بزرگ α استفاده کرد که میزان خطا زیاد است. پس از کاهش یافتن خطا، برای حفظ همگرایی شبکه، باید مقادیر کوچکی برای α در نظر گرفت [۴۷].

۲- ۱

استفاده از روش نزول گرادیانی با ترم ممنتوم: در این روش، برای اینکه α بزرگ باشد و ناپایداری رخ ندهد، طبق رابطه (۱۷-۳) ضریبی از وزن‌های مرحله قبل به مرحله فعلی اضافه می‌شود، که ضریب η ، ضریب ممنتوم است.

$$\Delta W_k^{new} = -\alpha \nabla E(W) \Big|_{W=W_{kk}} + \eta \Delta W_k^{old} \quad (17-3)$$

طبق رابطه (۱۷-۳) در دور $(k+1)$ ام، به جای $\Delta W_k = -\alpha \nabla E(W) \Big|_{W=W_k}$ در رابطه (۱۸-۳)، ΔW_k^{new} به عنوان تغییر وزن‌ها مورد استفاده قرار می‌گیرد [۳۹].

۳-۹-۲- الگوریتم کمترین میانگین مربع خطا (LMS)^{۷۶}

در این الگوریتم از روابط (۱۸-۳) و (۱۹-۳) برای بهینه‌سازی ماتریس وزن‌ها W و بردار بایاس b استفاده می‌شود:

$$W(k+1) = W(k) + 2a \times e(k) p^T(K) \quad (18-3)$$

$$b(k+1) = b(k) + 2a \times e(k) \quad (19-3)$$

که در آن P بردار ورودی، e بردار خطا و a سرعت یادگیری است. علامت T بالای P به معنی ترانپوز است. برای دور قبلی با $W(k)$ و $b(k)$ و برای دور فعلی با $W(k+1)$ و $b(k+1)$ نشان داده می‌شود.

⁷⁶ Last Mean Square Algorithm

هنگامی که هر یک از این ورودی‌ها برای شبکه به کار گرفته می‌شود، خروجی شبکه با هدف مقایسه می‌شود. خطای محاسبه شده به عنوان تفاوت میان خروجی واقعی و خروجی شبکه محاسبه می‌گردد. هدف، کمینه کردن میانگین مجموع این خطاها در k امین تکرار است [۵۴،۵۳].

۳-۹-۳- الگوریتم پس انتشار خطا^{۷۷}

این الگوریتم حالت عمومی‌تری از الگوریتم LMS است که برای شبکه‌های چندلایه به کار گرفته می‌شود. الگوریتم پس انتشار خطا در اصل یک الگوریتم سریع‌ترین نزول با شاخص عملکرد میانگین مربع خطا^{۷۸} (MSE) است که برای آموزش شبکه‌های چندلایه با تابع محرک غیرخطی، بکار می‌رود. در این روش ابتدا وزن‌های شبکه بصورت تصادفی انتخاب می‌شود، ورودی‌ها با این وزن‌های تصادفی به شبکه اعمال شده و خروجی شبکه محاسبه می‌گردد. سپس مقدار تابع خطا محاسبه می‌شود. برای مینیمم کردن تابع خطا، خطای خروجی شبکه از یک لایه به لایه پیشین آن انتشار می‌یابد و به همین دلیل اصطلاح پس انتشار برای این الگوریتم به کار می‌رود [۵۷-۵۵].

۳-۹-۴- الگوریتم لونبرگ - مارکواردت^{۷۹}

الگوریتم لونبرگ - مارکواردت الگوریتمی با کارایی بالا است که می‌تواند تا صد مرتبه سریع‌تر از الگوریتم‌های قبلی باشد. بعلاوه اینکه الگوریتم لونبرگ برای کمینه‌کردن توابع درجه دوم کارایی زیادی دارد، روش بسیار مناسبی برای آموزش شبکه‌هایی است که شاخص عملکرد آنها حداقل متوسط مربع خطا است. الگوریتم لونبرگ-مارکواردت به منظور رسیدن به سرعت آموزش بالا بدون نیاز به محاسبه‌ی ماتریس هسین^{۸۰} طراحی شد [۵۸].

⁷⁷ Back propagation Algorithm

⁷⁸ Mean square error

⁷⁹ Lvenberg Marquardt Algorithm

⁸⁰ Hessian

در این الگوریتم، که شکل تغییر یافته روش نیوتن است تابع عملکرد، شکل مجموع مربعات را دارد. معادله‌ی مربوط به روش نیوتن برای محاسبه‌ی پارامترهای مینیمم‌کننده‌ی تابع عملکرد به صورت زیر است:

$$X_{k+1} = X_k - A_k^{-1} g_k \quad , \quad A_k = \nabla^2 F(x) \Big|_{x=x_k} \quad , \quad g_k = \nabla F(x) \Big|_{x=x_k} \quad (20-3)$$

اگر $F(x)$ جمع مربع‌ها باشد:

$$F(x) = \sum_{i=1}^N v_i^2(x) = V^T(x) V(x) \quad (21-3)$$

که $V(x)$ بردار خطا است، عنصر j ام بردار شیب چنین است:

$$[\nabla F(x)]_j = \frac{\partial F(x)}{\partial x_j} = 2 \sum_{i=1}^N v_i(x) \frac{\partial v_i(x)}{\partial x_j} \quad (22-3)$$

که بردار شیب بصورت ماتریسی چنین نوشته می‌شود:

$$\nabla^2 F(x) = 2J^T(x) V(x) \quad (23-3)$$

در رابطه (۲۳-۳)، $J(k)$ ماتریس ژاکوبین^{۸۱} نام دارد و به صورت زیر است:

$$J(x) = \begin{bmatrix} \frac{\partial V_1(x)}{\partial x_1} & \frac{\partial V_1(x)}{\partial x_2} & \dots & \frac{\partial V_1(x)}{\partial x_n} \\ \frac{\partial V_2(x)}{\partial x_1} & \frac{\partial V_2(x)}{\partial x_2} & \dots & \frac{\partial V_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial V_N(x)}{\partial x_1} & \frac{\partial V_N(x)}{\partial x_2} & \dots & \frac{\partial V_N(x)}{\partial x_n} \end{bmatrix} \quad (24-3)$$

برای استفاده از رابطه (۲۰-۳) عناصر ماتریس هسین نیز مورد نیاز می‌باشد، عنصر (k, j)

از ماتریس هسین طبق رابطه (۲۵-۳) به دست می‌آید:

$$\left[\nabla^2 F(x) \right]_{k,j} = \frac{\partial^2 F(x)}{\partial x_k \partial x_j} = 2 \sum_{i=1}^N \left\{ \frac{\partial v_i(x)}{\partial x_k} \frac{\partial v_i(x)}{\partial x_j} + v_i(x) \frac{\partial^2 v_i(x)}{\partial x_k \partial x_j} \right\} \quad (25-3)$$

و بطور کلی ماتریس هسین را می‌توان چنین بیان کرد:

⁸¹ Jacobian matrix

$$\nabla^2 F(x) = 2J^T(x)J(x) + 2S(x) \quad (26-3)$$

که در آن $S(x)$ کوچک و قابل صرف نظر کردن می‌باشد:

$$S(x) = \sum_{i=1}^N V_i(x) \nabla^2 V_i(x) \cong 0 \quad (27-3)$$

بنابراین می‌توان بطور تقریبی ماتریس هسین را چنین نوشت:

$$\nabla^2 F(x) = 2J^T(x)J(x) \quad (28-3)$$

با جایگذاری روابط (23-3) و (28-3) در رابطه (20-3) روش گaus نیوتن به دست می‌آید:

$$X_{k+1} = X_k - [2J^T(X_k)J(X_k)]^{-1} 2J^T(X_k)V(X_k) \quad (29-3)$$

$$= X_k - [J^T(X_k)J(X_k)]^{-1} J^T(X_k)J(X_k)$$

مشکلی که این روش با آن روبروست، امکان معکوس پذیر نبودن عبارت $H = J^T J$ می‌باشد.

پس باید به نحوی آن را به یک تابع معکوس پذیر تبدیل کرد. اینکار با افزودن یک مقدار ثابت به تمام

اعضای ماتریس انجام می‌شود. یعنی به جای عبارت $H = J^T J$ عبارت G بکار برده می‌شود.

$$G = H + \mu I \quad (30-3)$$

که در این رابطه μ مقدار عددی و I ماتریس یکانی است.

با انجام این اصلاح در روش گaus-نیوتن، الگوریتم لوببرگ - مارکوارتت به دست می‌آید:

$$X_{k+1} = X_k - [J^T(X_k)J(X_k) + \mu_k I]^{-1} J^T(X_k)V(X_k) \quad (31-3)$$

و یا

$$\Delta X_k = -[J^T(X_k)J(X_k) + \mu_k I]^{-1} J^T(X_k)V(X_k) \quad (32-3)$$

الگوریتم لوببرگ - مارکوارتت با یک μ_k کوچک (مثلاً $\mu = 0.01$) شروع می‌شود. اگر یک دور

منجر به مقدار کوچک تری برای متوسط مربعات خطاها نشود، نتایج مربوط به آن دور نادیده گرفته

می‌شود و تغییرات وزن‌ها و بایاس‌ها اعمال نمی‌شود. در دور بعدی μ_k در فاکتوری بزرگتر از یک (

مثلاً ۱۰) ضرب شده و محاسبات با μ جدید آغاز می‌شود و اگر یک دور به مقدار کوچکتری برای

متوسط مربعات خطاها بیانجامد، نتایج مربوط به آن دور اعمال می‌شود (تغییرات لازم در وزن‌ها و بایاس‌ها اعمال می‌شود) و μ_k در فاکتوری کوچکتر از یک (مثلاً ۰/۱) ضرب می‌شود و در واقع μ_k کاهش داده می‌شود و دور بعدی با μ جدید شروع می‌شود. به این ترتیب الگوریتم لونبرگ-مارکواردت حالت بینابین الگوریتم سریعترین نزول و روش نیوتن است [۴۷].

پارامترهای آموزش برای الگوریتم لونبرگ - مارکواردت (train lm) عبارتند از:

دوره‌ها (epochs)

نمایش (show)

هدف (goal)

زمان (time)

گرادیان - مینیمم (min-grad)

رد شدن - ماکزیموم (max-grad)

ضریب ممنتوم (mu)

کاهش - mu (mu-dec)

افزایش - mu (mu-inc)

ماکزیموم - mu (mu-max)

کاهش - حافظه (mem-reduct)

وضعیت آموزش برای هر یک از تکرارهای الگوریتم نمایش داده می‌شود. پارامترهای دیگر زمانی

که آموزش متوقف می‌شود، تعیین می‌شوند.

یکی از مشکلاتی که در خلال آموزش شبکه رخ می‌دهد، انطباق بیش از حد^{۸۲} داده‌هاست. در این

این حالت خطای مجموعه‌ی آموزش مقدار بسیار کوچکی می‌شود در حالی که هنگام کار با داده‌های

⁸² Overfitting

جدید، شبکه، خطای بسیار بزرگی را نشان می‌دهد. اصطلاحاً گفته می‌شود که شبکه الگوهای آموزش داده شده را به خاطر می‌سپارد، اما قابلیت تعمیم به حالات جدید را ندارد.

یک روش برای افزایش قدرت تعمیم شبکه، استفاده از یک شبکه‌ی بزرگ است. هر چه شبکه بزرگتر باشد برای تطبیق اطلاعات مناسب‌تر است ولی مشکلی که وجود دارد این است که از قبل دانستن میزان اندازه‌ی شبکه مقدور نیست.

۳-۹-۵- الگوریتم تنظیم بایزین^{۸۳}

یکی دیگر از روش‌های افزایش قدرت شبکه، روش تنظیم^{۸۴} نام دارد. این روش شامل یک تابع عملکرد اصلاح‌شده می‌باشد.

۳-۹-۵-۱- تابع عملکرد اصلاح شده

تابع عملکردی که برای شبکه‌های عصبی پیش‌خور مورد استفاده قرار می‌گیرد برابر با مجموع مجذورهای شبکه خطاهای شبکه است.

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - e_i)^2 \quad (3-33)$$

بهبود تنظیم شبکه زمانی میسر است که تابع عملکرد را با عبارتی تصحیح کنیم که شامل مجموع مجذورهای وزن‌ها و بایاس‌های شبکه می‌شود.

$$mesreg = \gamma mse + (1 - \gamma) msw \quad (3-34)$$

γ نرخ عملکرد است و

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (3-35)$$

⁸³ Bayesian Regularization Algorithm

⁸⁴ Regularization

استفاده از این تابع عملکرد موجب می‌شود که شبکه وزن‌ها و بایاس‌های کوچک‌تری داشته باشد و باعث می‌شود که پاسخ شبکه کمترین انطباق بیش از حد را داشته باشد. مشکل روش تنظیم این است که انتخاب بهینه مقدار پارامتر نرخ عملکرد مشکل است. برای این منظور از روشی به نام بایزین استفاده می‌کنیم که در آن پارامترهای تنظیم به صورت اتوماتیک تنظیم می‌شوند.

۳-۹-۵-۲- تنظیم اتوماتیک بایزین

این تنظیم از روشی به نام بایزین که توسط دیوید مک‌کی^{۸۵} مطرح شد، نتیجه گرفته شده است. در این روش، وزن‌ها و بایاس‌های شبکه و اعداد تصادفی با توزیع‌های مشخص فرض می‌شود. پارامترهای تنظیم به وسیله‌ی واریانس‌هایی که مرتبط با این توزیع‌ها هستند، تنظیم می‌شوند. این پارامترها را می‌توانیم توسط تکنیک‌های آماری تخمین بزنیم. الگوریتم تنظیم بایزین، در قالب تابع trainbr در جعبه ابزار MATLAB به کار برده می‌شود [۵۹].

۳-۱۰- محاسن شبکه‌های عصبی مصنوعی

۱- قابلیت یادگیری

قابلیت یادگیری یعنی توانایی تنظیم پارامترهای شبکه، به نحوی که با تغییر ورودی‌ها، شبکه بتواند خود را با شرایط جدید وفق دهد. در نتیجه در هر زمینه‌ای که مجموعه داده‌های آموزش را دریافت کند (بدون توجه به نوع ورودی و خروجی) در آن زمینه کارآمد باشد.

۲- پردازش اطلاعات بصورت موازی

⁸⁵ David McKay

سلول‌های هر لایه، بطور همزمان روی ورودی‌های آن لایه اثر می‌کنند و علی‌رغم مستقل بودن، جمعاً ورودی را پردازش کرده و خروجی را تحویل می‌دهند. در نتیجه، بار یک کار، بر دوش چند پردازشگر تقسیم شده و سرعت پردازش بالا می‌رود.

۱- مقاوم بودن

چون هر سلول کار مخصوص به خود را انجام می‌دهد و رفتار کلی شبکه، برآیند رفتارهای محلی سلول‌های متعدد است، اگر چند سلول به‌طور محلی خطا کنند، خروجی نهایی به خاطر درست عمل کردن اکثر نرون‌ها از این خطاها مصون می‌ماند. بعبارت دیگر نرون‌ها خطای یکدیگر را تصحیح می‌کنند. در این صورت شبکه توانایی تحمل خطا و مقاومت در مقابل اشتباه را دارد.

۲- قابلیت تعمیم

بعد از آموزش شبکه با چند الگو، شبکه قادر است در مقابل یک ورودی تجربه نشده قرار بگیرد و خروجی مناسب آن را ارائه نماید. بدین معنی که شبکه از آموخته‌های خود برای پاسخ دادن به مجهولات استفاده کند.

۳- پراکندگی آموخته‌ها در کل شبکه

وقتی شبکه آموزش می‌بیند، تغییرات در وزن‌ها و بایاس تثبیت می‌شود و چون تعداد عناصر ماتریس‌های وزن و بردارهای بایاس عموماً زیاد است، تغییرات متمرکز نیستند و لذا اگر بخشی از نرون‌ها حذف شوند باز هم احتمال اینکه شبکه بتواند به ورودی تجربه نشده پاسخ صحیح بدهد، زیاد است [۳۹].

۳-۱۱- معایب شبکه‌های عصبی مصنوعی

۱- مدت زمان آموزش

اگر شبکه‌های عصبی مصنوعی برای حل مسائل بزرگ به کار گرفته شوند، مدت زمان زیادی صرف آموزش شبکه می‌شود و حتی سریع‌ترین الگوریتم‌ها ممکن است برای این مسائل، نیاز به ساعت‌ها آموزش داشته باشند.

۱- کمینه‌های محلی

تمام شبکه‌ها، احتمال گیرافتادن در کمینه‌های محلی را دارند. بنابراین هیچ‌گاه نمی‌توان گفت که یک حل معین، بهترین جواب ممکن است و اگر شبکه جواب‌های غیر قابل قبول ارائه داد، باید آموزش دادن شبکه مجدداً انجام شود تا زمانی که جواب مورد قبول به دست آید.

۲- برون یابی ضعیف

اگر شبکه با یک ورودی تجزیه نشده که خارج از محدوده‌ی آموزش باشد، آزموده شود، ممکن است جوابی بدهد که هیچ تشابهی با مقدار واقعی و مورد انتظار نداشته باشد که این امر می‌تواند خسارات جانی و مالی جبران‌ناپذیری را به‌ویژه در صورت استفاده‌ی نابجا در سامانه‌های کنترل به‌بار آورد. بنابراین در هنگام آموزش شبکه، باید تا حد امکان مجموعه آموزش را غنی‌تر کرد [۶۰].

۳-۱۲- کاربردهای شبکه عصبی مصنوعی

از شبکه‌های عصبی مصنوعی در علوم و صنایع مختلف، به‌طور وسیع استفاده می‌شود که در اینجا تنها به بعضی از این کاربردها اشاره می‌گردد:

۱- صنایع شیمیایی: در سامانه CSTR، برای کنترل PH هیدروکسید سدیم [۳۸]

- ۲- صنایع هوا فضا: هواپیماهای خودران و سامانه‌های کنترل هواپیما
 - ۳- صنایع اتومبیل: سامانه‌های هدایت خودکار اتومبیل
 - ۴- بانکداری: خواننده‌های چک و سایر اسناد بانکی
 - ۵- صنایع نظامی: سلاح‌های هدف‌یاب و سلاح‌های هدایت‌پذیر
 - ۶- صنایع الکترونیک: مدل‌کننده‌های غیرخطی و کنترل فرایند
 - ۷- فرایندهای تولیدی: کنترل فرایندهای تولید و تجزیه، تولید و طراحی محصول
 - ۸- کاربردهای طبی: از بین‌بردن اغتشاش‌های دستگاه ECG که ضربان قلب مریض را نشان می‌دهد.
 - ۹- ساخت ربات: کنترل مسیر ربات
- در اینجا تنها مثال‌هایی از عمومی‌ترین و قابل‌فهم‌ترین کاربردها آورده شده‌اند اما کاربردهای تخصصی در زمینه‌های مختلف بیشمارند [۴۷].

۴-۱- مقدمه

در این تحقیق دانسیته‌ی گروهی از کتون‌ها در محدوده‌ی دمایی K ۳۳۳-۲۷۳ و محدوده‌ی فشار ۲-۳۷۰ MPa با استفاده از روش رگرسیون خطی چندگانه و شبکه عصبی مصنوعی مورد بررسی قرار گرفته است. بخش تجربی این پژوهش شامل مراحل زیر است:

۱- انتخاب سری داده‌ها

۲- محاسبه‌ی توصیف کننده‌ها

۳- مدلسازی به روش رگرسیون خطی چندگانه

۴- مدلسازی به وسیله شبکه عصبی مصنوعی

۴-۲- انتخاب سری داده‌ها

سری داده‌ها از نتایج خواص ترمودینامیکی و اندازه‌گیری‌های حجمی ۶ کتون شامل استون، ۲- پنتانول، ۲- بوتانول، ۴- متیل-۲- پنتانول، ۳- پنتانول و ۲- هگزانول توسط گروه تحقیقاتی R. Malhotra, W. E. Price, and L.A. Woolf به دست آمد [۶۱-۶۴].

این گروه کمیت $k = \frac{V(p)}{V(p_0)}$ را برای هر یک از کتون‌ها در محدوده‌ی فشار ۲-۳۷۰ MPa و در ۶

دمای K ۲۷۳/۱۵ - ۲۸۸/۱۵ - ۲۹۸/۱۵ - ۳۱۳/۱۵ - ۳۲۴/۱۵ و ۳۳۸/۱۳ گزارش نمودند. با

استفاده از مقدار $V(p_0)$ گزارش شده در هر دما برای هر یک از کتون‌ها [۶۱-۶۴]، مقادیر دانسیته‌ی

واقعی برای هر مولکول در ۶ دما و محدوده‌ی فشار معین محاسبه گردید.

در جداول (۴-۱) تا (۴-۶) مقادیر k ، T ، p و دانسیته برای هر ۶ مولکول کتون، شامل استون،

۲-پنتانول، ۲- بوتانول، ۴- متیل-۲- پنتانول، ۳- پنتانول و ۲- هگزانول آورده شده است.

جدول ۴-۱- مقادیر مولکولی ρ ، T ، k و دانسیته‌ی مولکولی استون

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
۱	2/5	278/15	808/96	0/9974
۲	5	278/15	811/08	0/9948
۳	10/36	278/15	815/42	0/9895
۴	15/23	278/15	819/23	0/9849
۵	20/06	278/15	822/82	0/9806
۶	30/51	278/15	830/27	0/9718
۷	41/67	278/15	837/69	0/9632
۸	62/13	278/15	850/31	0/9489
۹	81/52	278/15	860/74	0/9374
۱۰	101/75	278/15	870/77	0/9266
۱۱	126/5	278/15	882/01	0/9148
۱۲	151/2	278/15	892/35	0/9042
۱۳	175/27	278/15	901/62	0/8949
۱۴	198/09	278/15	909/96	0/8867
۱۵	224/76	278/15	918/97	0/8780
۱۶	250/89	278/15	927/32	0/8701
۱۷	275/33	278/15	934/62	0/8633
۱۸	300/21	278/15	941/71	0/8568
۱۹	325/27	278/15	948/69	0/8505
۲۰	349/7	278/15	955/09	0/8448
۲۱	2/5	288/15	797/97	0/9971
۲۲	5	288/15	800/3	0/9942
۲۳	15/24	288/15	809/17	0/9833
۲۴	19/61	288/15	812/81	0/9789
۲۵	24/37	288/15	816/48	0/9745
۲۶	30/71	288/15	821/28	0/9688
۲۷	40/55	288/15	828/29	0/9606

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
۲۸	99/17	288/15	862/41	6
۲۹	125/1	288/15	874/64	7
۳۰	151	288/15	885/84	2
۳۱	175/72	288/15	895/81	2
۳۲	200/57	288/15	905/08	1
۳۳	225/76	288/15	913/82	7
۳۴	250/34	288/15	921/86	1
۳۵	276/14	288/15	929/84	7
۳۶	300/06	288/15	936/84	3
۳۷	327/13	288/15	944/4	5
۳۸	351/2	288/15	950/95	7
۳۹	375/3	288/15	957/24	2
۴۰	388/52	288/15	960/48	4
۴۱	2/5	298/15	786/84	7
۴۲	5	298/15	789/37	5
۴۳	10	298/15	794/17	5
۴۴	15/36	298/15	799/02	5
۴۵	19/86	298/15	802/95	7
۴۶	24/63	298/15	806/91	9
۴۷	60/43	298/15	832/61	9
۴۸	78/42	298/15	843/27	0
۴۹	100/53	298/15	855/22	0

۵۰	125/7	298/15	867/71	8	0/903
۵۱	151/69	298/15	879/39	8	0/891
۵۲	175/88	298/15	889/36	8	0/881
۵۳	198/49	298/15	898/02	3	0/873
۵۴	224/81	298/15	907/37	3	0/864

ادامه‌ی جدول
۱-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k	
۵۵	249/65	298/15	915/74	4	0/856
۵۶	274/47	298/15	923/61	1	0/849
۵۷	300/06	298/15	931/4	0	0/842
۵۸	324/98	298/15	938/65	5	0/835
۵۹	350/9	298/15	945/67	3	0/829
۶۰	374/05	298/15	951/75	0	0/824

جدول ۲-۴- مقادیر مولکولی p ، T ، k و دانسیته‌ی مولکولی ۲- بوتانول

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k	
1	2/32	278/15	822/46	7	0/997
2	5/04	278/15	824/61	1	0/995
3	9/75	278/15	828/27	7	0/990
4	15/13	278/15	832/31	9	0/985
5	18/19	278/15	834/51	3	0/983
6	25/36	278/15	839/46	5	0/977
7	30/66	278/15	842/99	4	0/973

8	40/2	278/15	849/01	5	0/966
9	50/43	278/15	855/12	6	0/959
10	60/47	278/15	860/68	4	0/953
11	70/6	278/15	866/13	4	0/947
12	80/52	278/15	871/19	9	0/941
13	90/25	278/15	875/93	8	0/936
14	100/47	278/15	880/63	8	0/931
15	120/61	278/15	889/6	4	0/922
16	141/69	278/15	898/27	5	0/913

ادامہی
جدول ۲-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k	
17	160/26	278/15	905/41	3	0/906
18	180/77	278/15	912/86	9	0/898
19	200/52	278/15	919/61	3	0/892
20	219/81	278/15	925/84	3	0/886
21	2/59	288/15	812/43	3	0/997
22	5/05	288/15	814/56	7	0/994
23	10/01	288/15	818/59	8	0/989
24	14/94	288/15	822/41	2	0/985
25	20/26	288/15	826/44	4	0/980
26	25/9	288/15	830/5	6	0/975
27	30/61	288/15	833/75	8	0/971
28	39/79	288/15	839/8	8	0/964
29	50/28	288/15	846/38	3	0/957

30	60/06	288/15	852/08	9	0/950
31	70/7	288/15	858/03	3	0/944
32	80/22	288/15	863/06	8	0/938
33	90/6	288/15	868/33	1	0/933
34	101/38	288/15	873/57	5	0/927
35	121/06	288/15	882/61	0	0/918
36	140/14	288/15	890/77	6	0/909
37	160/46	288/15	898/87	4	0/901
38	181/02	288/15	906/61	7	0/893
39	200/22	288/15	913/36	1	0/887
40	220/16	288/15	920	7	0/880
41	2/69	298/15	802/31	9	0/996
42	5/19	298/15	804/49	2	0/994
43	9/99	298/15	808/63	1	0/989

ادامه‌ی
جدول ۲-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k	
44	15/35	298/15	813/07	7	0/983
45	19/97	298/15	816/73	3	0/979
46	26/05	298/15	821/34	8	0/973
47	30/26	298/15	824/47	1	0/970
48	39/94	298/15	831/24	2	0/962
49	50/33	298/15	838/03	4	0/954
50	59/81	298/15	843/78	9	0/947
51	70/9	298/15	850/24	7	0/940

					0/934
52	80/78	298/15	855/7	7	0/929
53	90	298/15	860/85	1	0/923
54	100/67	298/15	865/89	7	0/913
55	120/71	298/15	875/36	7	0/904
56	140/64	298/15	884/07	7	0/896
57	160/66	298/15	892/16	5	0/888
58	180/77	298/15	899/79	9	0/882
59	200/72	298/15	906/83	0	0/875
60	219/56	298/15	913/25	8	0/997
61	2/33	313/15	786/27	0	0/993
62	4/96	313/15	788/8	8	0/987
63	10/21	313/15	793/83	5	0/982
64	15/09	313/15	798/28	0	0/976
65	20	313/15	802/53	8	0/971
66	25/4	313/15	807/07	3	0/966
67	30/76	313/15	811/33	2	0/958
68	39/69	313/15	818/02	3	0/950
69	49/58	313/15	824/99	2	0/942
70	59/91	313/15	831/82	4	

ادامه‌ی
جدول ۲-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k	
				0/935	
71	70/15	313/15	838/05	4	0/928
72	80/78	313/15	844/28	5	0/922
73	90/05	313/15	849/4	9	

					0/916
74	100/72	313/15	854/96	9	0/906
75	120/71	313/15	864/77	5	0/897
76	140/24	313/15	873/63	3	0/888
77	160/46	313/15	882/09	7	0/880
78	180/47	313/15	889/9	9	0/873
79	200/57	313/15	897/23	7	0/867
80	219/96	313/15	903/96	2	0/996
81	2/61	323/15	775/93	4	0/993
82	5	323/15	778/43	2	0/986
83	9/99	323/15	783/48	8	0/980
84	15/24	323/15	788/52	5	0/975
85	19/98	323/15	792/88	1	0/969
86	25/03	323/15	797/38	6	0/964
87	30/21	323/15	801/76	3	0/954
88	40/55	323/15	809/99	5	0/946
89	50/43	323/15	817/19	1	0/938
90	60/21	323/15	823/89	4	0/930
91	70/8	323/15	830/71	7	0/924
92	80/37	323/15	836/46	3	0/927
93	90/65	323/15	842/38	8	0/911
94	101/48	323/15	848/3	4	0/901
95	120/56	323/15	858/09	0	0/891
96	140/74	323/15	867/53	2	

ادامہ ی
جدول ۲-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
				0/882
97	159/91	323/15	875/88	7
				0/874
98	180/12	323/15	884/09	5
				0/867
99	200/07	323/15	891/64	1
				0/860
100	220/01	323/15	898/9	1
				0/996
101	2/51	338/15	759/77	0
				0/992
102	4/76	338/15	762/37	6
				0/985
103	9/87	338/15	768/1	2
				0/978
104	14/99	338/15	773/59	2
				0/972
105	19/95	338/15	778/53	0
				0/966
106	25/02	338/15	783/36	0
				0/961
107	29/25	338/15	787/19	3
				0/950
108	39/34	338/15	795/8	9
				0/941
109	50/03	338/15	804/01	2
				0/932
110	60/26	338/15	811/33	7
				0/925
111	70/2	338/15	817/91	2
				0/917
112	80/62	338/15	824/5	8
				0/911
113	90/3	338/15	830/2	5
				0/905
114	100/22	338/15	835/8	4
				0/893
115	120/61	338/15	846/55	9
				0/883
116	140/34	338/15	856/13	9
				0/874
117	160/26	338/15	864/93	9
				0/866
118	180/57	338/15	873/32	5

119	200/57	338/15	880/94	0	0/859
120	220/41	338/15	888/08	1	0/852

جدول ۴-۳- مقادیر مولکولی p ، T ، k و دانسیته‌ی مولکولی ۲- پنتانون

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
1	2/55	278/15	822/52	0/9977
2	5	278/15	824/34	0/9955
3	9/68	278/15	827/75	0/9914
4	14/78	278/15	831/27	0/9872
5	17/09	278/15	832/87	0/9853
6	23/42	278/15	837/04	0/9804
7	27/4	278/15	839/52	0/9775
8	38/01	278/15	845/92	0/9701
9	58/67	278/15	857/23	0/9573
10	78/78	278/15	867/11	0/9464
11	99/59	278/15	876/46	0/9363
12	124/42	278/15	886/78	0/9254
13	149/1	278/15	896/28	0/9156
14	174/9	278/15	905/47	0/9063
15	199/99	278/15	913/84	0/8980
16	224/78	278/15	921/64	0/8904
17	249/82	278/15	928/95	0/8834
18	275/24	278/15	935/94	0/8768
19	299/7	278/15	942/49	0/8707
20	326/6	278/15	949/36	0/8644
21	2/55	288/15	813/1	0/9975
22	5	288/15	815/06	0/9951
23	7/86	288/15	817/28	0/9924
24	12/92	288/15	821/09	0/9878

25	19/32	288/15	825/68	0/9823
26	23/25	288/15	828/47	0/9790
27	27/8	288/15	831/44	0/9755

ادامه‌ی جدول
۳-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
28	41/84	288/15	840/31	0/9652
29	59/07	288/15	850/09	0/9541
30	79/87	288/15	860/64	0/9424
31	99/57	288/15	869/87	0/9324
32	124/48	288/15	880/45	0/9212
33	149/55	288/15	890/41	0/9109
34	174/87	288/15	899/69	0/9015
35	200/34	288/15	908/46	0/8928
36	225/69	288/15	916/46	0/8850
37	250/38	288/15	923/88	0/8779
38	275/41	288/15	930/87	0/8713
39	300/55	288/15	937/65	0/8650
40	326/74	288/15	944/53	0/8587
41	2/55	298/14	820/07	0/9973
42	5	298/14	805/72	0/9947
43	10/3	298/14	809/95	0/9895
44	15/35	298/14	813/9	0/9847
45	18/48	298/14	816/31	0/9818
46	23/61	298/14	820/07	0/9773
47	29/22	298/14	823/94	0/9727
48	39/72	298/14	830/95	0/9645
49	59/41	298/14	842/74	0/9510
50	80/37	298/14	853/88	0/9386

51	100/84	298/14	863/72	0/9279
52	124/87	298/14	874/28	0/9167
53	150/26	298/14	884/6	0/9060
54	175/41	298/14	894/08	0/8964

اداهى جدول
۳-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
55	200/72	298/14	902/94	0/8876
56	225/92	298/14	911/15	0/8796
57	250/53	298/14	918/57	0/8725
58	275/72	298/14	925/89	0/8656
59	299/96	298/14	932/57	0/8594
60	325/6	298/14	939/35	0/8532
61	2/55	313/14	789/28	0/9969
62	5	313/14	791/58	0/9940
63	10/33	313/14	796/39	0/9880
64	15/02	313/14	800/36	0/9831
65	19/82	313/14	804/28	0/9783
66	24/07	313/14	807/58	0/9743
67	28/14	313/14	810/66	0/9706
68	40/01	313/14	819/1	0/9606
69	59/87	313/14	831/66	0/9461
70	80/11	313/14	842/97	0/9334
71	100/69	313/14	853/49	0/9219
72	125/36	313/14	864/93	0/9097
73	150/2	313/14	875/52	0/8987
74	175/57	313/14	885/37	0/8887
75	200/25	313/14	894/23	0/8799
76	224/77	313/14	902/43	0/8719
77	249/75	313/14	910/37	0/8643

78	275/19	313/14	918/01	0/8571
79	301/02	313/14	925/46	0/8502
80	325/99	313/14	932/15	0/8441
81	2/55	323/14	780/16	0/9967

ادامه‌ی جدول
۳-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
82	5	323/14	782/6	0/9967
83	9/93	323/14	787/35	0/9936
84	14/96	323/14	791/92	0/9876
85	19/55	323/14	795/9	0/9819
86	23/98	323/14	799/58	0/9770
87	30/1	323/14	804/46	0/9725
88	40/13	323/14	811/93	0/9666
89	59/49	323/14	824/77	0/9577
90	79/58	323/14	836/57	0/9428
91	99/94	323/14	847/32	0/9295
92	124/8	323/14	859/22	0/9177
93	149/34	323/14	869/88	0/9050
94	175/09	323/14	880/12	0/8939
95	199/58	323/14	889/18	0/8835
96	224/95	323/14	897/81	0/8745
97	249/02	323/14	905/65	0/8661
98	274/41	323/14	913/52	0/8586
99	300/15	323/14	920/99	0/8512
100	324/06	323/14	927/47	0/8384
101	2/55	338/13	764/77	0/9962
102	5	338/13	767/46	0/9927
103	10/01	338/13	772/76	0/9859
104	14/68	338/13	777/41	0/9800

105	19/67	338/13	782/12	0/9741
106	23/3	338/13	785/34	0/9701
107	27/72	338/13	789/17	0/9654
108	39/05	338/13	798/18	0/9545

ادامه‌ی جدول
۳-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
109	59/92	338/13	812/82	0/9373
110	79/94	338/13	825/06	0/9234
111	100/49	338/13	836/38	0/9109
112	125/77	338/13	848/96	0/8974
113	149/43	338/13	859/79	0/8861
114	175/1	338/13	870/4	0/8753
115	200/32	338/13	879/85	0/8659
116	224/85	338/13	888/47	0/8575
117	250/09	338/13	896/83	0/8495
118	274/98	338/13	904/71	0/8421
119	299/91	338/13	912/19	0/8352
120	325/2	338/13	919/34	0/8287

جدول ۴-۴- مقادیر مولکولی p , T , k و دانسیته‌ی مولکولی ۴-متیل - ۲-پنتانون

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
1	2/55	278/15	816/35	0/9976
2	5	278/15	818/15	0/9954
3	9/72	278/15	821/62	0/9912
4	14/73	278/15	825/12	0/9870
5	19/54	278/15	828/39	0/9831
6	23/79	278/15	831/18	0/9798
7	28/57	278/15	834/25	0/9762

8	39/77	278/15	840/96	0/9684
9	59/01	278/15	851/43	0/9565
10	79/8	278/15	861/7	0/9451
11	100/04	278/15	870/73	0/9353

ادامه‌ی جدول
۴-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
12	124/95	278/15	880/99	0/9244
13	149/85	278/15	890/53	0/9145
14	174/92	278/15	899/38	0/9055
15	200/02	278/15	907/7	0/8972
16	224/99	278/15	915/35	0/8897
17	250/17	278/15	922/61	0/8827
18	274/92	278/15	929/35	0/8763
19	300/9	278/15	936/08	0/8700
20	325/83	278/15	942/47	0/8641
21	2/55	288/15	807/46	0/9947
22	5	288/15	809/33	0/9951
23	9/74	288/15	813	0/9906
24	14/99	288/15	816/88	0/9859
25	20/29	288/15	820/71	0/9813
26	23/66	288/15	822/97	0/9786
27	27/2	288/15	825/33	0/9758
28	39/51	288/15	833/19	0/9666
29	59/31	288/15	844/46	0/9537
30	79/11	288/15	854/58	0/9424
31	100/24	288/15	864/31	0/9318
32	124/76	288/15	874/73	0/9207
33	150/07	288/15	884/62	0/9104
34	174/96	288/15	893/65	0/9012

35	199/88	288/15	901/96	0/8929
36	224/67	288/15	909/81	0/8852
37	249/76	288/15	917/16	0/8781
38	275/26	288/15	924/21	0/8714

ادامه‌ی جدول
۴-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
39	299/93	288/15	930/84	0/8652
40	325/93	288/15	937/45	0/8591
41	2/55	298/14	798/49	0/9972
42	5	298/14	800/49	0/9947
43	9/51	298/14	804/21	0/9901
44	14/6	298/14	808/21	0/9852
45	19/96	298/14	812/25	0/9803
46	23/62	298/14	814/91	0/9771
47	28/4	298/14	818/26	0/9731
48	39/6	298/14	825/73	0/9643
49	59/45	298/14	837/54	0/9507
50	78/81	298/14	847/71	0/9393
51	100/14	298/14	857/94	0/9281
52	125/6	298/14	869/08	0/9162
53	150/54	298/14	879/06	0/9058
54	175/04	298/14	888/18	0/8965
55	200/58	298/14	896/88	0/8878
56	225/02	298/14	904/73	0/8801
57	249/82	298/14	912/19	0/8729
58	274/97	298/14	919/25	0/8662
59	300/46	298/14	926/2	0/8597
60	325/97	298/14	932/81	0/8536
61	2/55	313/14	784/82	0/9969

62	5	313/14	787/11	0/9940
63	9/85	313/14	791/41	0/9886
64	15/25	313/14	796	0/9829
65	19/71	313/14	799/66	0/9784

اداهى جدول
۴-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
66	23/77	313/14	802/86	0/9745
67	28/12	313/14	806/17	0/9705
68	38/48	313/14	813/55	0/9617
69	59/55	313/14	826/88	0/9462
70	79/92	313/14	838/22	0/9334
71	100/08	313/14	848/21	0/9224
72	124/88	313/14	859/49	0/9103
73	149/15	313/14	869/61	0/8997
74	174/59	313/14	879/49	0/8896
75	199/53	313/14	888/37	0/8707
76	225/21	313/14	896/82	0/8724
77	250/39	313/14	904/6	0/8649
78	275/12	313/14	911/77	0/8581
79	299/95	313/14	918/73	0/8516
80	326/13	313/14	925/69	0/8452
81	2/55	323/14	776/14	0/9966
82	4/9	323/14	778/48	0/9936
83	9/87	323/14	783/21	0/9876
84	14/82	323/14	787/68	0/9820
85	19/66	323/14	791/87	0/9768
86	24/01	323/14	795/45	0/9724
87	27/92	323/14	798/49	0/9687
88	39/4	323/14	806/99	0/9585

89	59/33	323/14	819/99	0/9433
90	79/6	323/14	831/63	0/9301
91	100/34	323/14	842/41	0/9182
92	124/49	323/14	853/85	0/9059

ادامه‌ی جدول
۴-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
93	149/67	323/14	864/63	0/8946
94	174/67	323/14	874/41	0/8846
95	200	323/14	883/6	0/8754
96	224/41	323/14	891/75	0/8674
97	250/23	323/14	899/84	0/8596
98	275/58	323/14	907/44	0/8524
99	300/17	323/14	914/52	0/8458
100	326/19	323/14	921/49	0/8394
101	2/55	338/13	761/81	0/9962
102	5	338/13	764/5	0/9927
103	9/73	338/13	769/46	0/9863
104	14/73	338/13	774/41	0/9800
105	19/55	338/13	778/86	0/9744
106	24/32	338/13	783/12	0/9691
107	28/88	338/13	786/93	0/9644
108	43/19	338/13	797/94	0/9511
109	59/66	338/13	809/08	0/9380
110	79/75	338/13	821/16	0/9242
111	100/4	338/13	832/33	0/9118
112	125	338/13	844/46	0/8987
113	149/27	338/13	855/31	0/8873
114	175/06	338/13	865/75	0/8766
115	199/96	338/13	875/04	0/8673

116	224/79	338/13	883/49	0/8590
117	250/19	338/13	891/8	0/8510
118	274/38	338/13	899/19	0/8440
119	301/01	338/13	907/04	0/8367

ادامه‌ی جدول
۴-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
120	325/94	338/13	913/92	0/8304

جدول ۴-۵- مقادیر مولکولی p ، T ، k و دانسیته‌ی مولکولی ۳- پنتانون

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
1	2/55	278/15	830/66	0/9977
2	5	278/15	832/49	0/9955
3	8/53	278/15	835/09	0/9924
4	13/29	278/15	838/39	0/9885
5	20/22	278/15	842/99	0/9831
6	23/37	278/15	844/97	0/9808
7	28/44	278/15	848/08	0/9772
8	39/71	278/15	854/73	0/9696
9	49/07	278/15	859/87	0/9638
10	59/61	278/15	865/44	0/9576
11	68/88	278/15	870/07	0/9525
12	79/2	278/15	875/03	0/9471
13	89/27	278/15	879/68	0/9421
14	99/79	278/15	884/37	0/9371
15	119/6	278/15	892/56	0/9285
16	139/86	278/15	900/32	0/9205
17	160/6	278/15	907/92	0/9128
18	180/35	278/15	914/73	0/9060

19	187/72	278/15	917/06	0/9037
20	2/55	288/15	821/22	0/9974
21	5	288/15	823/28	0/9949

ادامہی جدول
۵-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
22	9/67	288/15	826/85	0/9906
23	14/41	288/15	830/37	0/9864
24	19/36	288/15	833/84	0/9823
25	22/97	288/15	836/22	0/9795
26	27/65	288/15	839/22	0/9760
27	49/35	288/15	851/97	0/9614
28	59/43	288/15	857/5	0/9552
29	69/52	288/15	862/73	0/9494
30	78/84	288/15	867/39	0/9443
31	89/19	288/15	872/48	0/9388
32	100/13	288/15	877/52	0/9334
33	119/97	288/15	886/07	0/9244
34	139/86	288/15	894/09	0/9161
35	159/49	288/15	901/47	0/9086
36	179/47	288/15	908/57	0/9015
37	198/99	288/15	915/17	0/8950
38	219/04	288/15	921/56	0/8888
39	2/73	298/15	811/73	0/9971
40	5/05	298/15	813/69	0/9947
41	9/75	298/15	817/47	0/9901
42	14/72	298/15	821/29	0/9855
43	19/5	298/15	824/8	0/9813
44	24/59	298/15	828/43	0/9770
45	28/29	298/15	830/9	0/9741

46	39/28	298/15	838/04	0/9658
47	49/87	298/15	844/42	0/9585
48	59/43	298/15	849/92	0/9523

ادامہی جدول

۵-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
49	69/58	298/15	855/4	0/9462
50	79/87	298/15	860/77	0/9403
51	89/3	298/15	865/46	0/9352
52	99/87	298/15	870/49	0/9298
53	119/76	298/15	879/38	0/9204
54	139/66	298/15	887/58	0/9119
55	156/96	298/15	894/34	0/9050
56	178/05	298/15	902/02	0/8973
57	198/31	298/15	909/01	0/8904
58	219/12	298/15	915/8	0/8838
59	2/53	313/15	797/26	0/9969
60	5	313/15	799/58	0/9940
61	9/9	313/15	804/03	0/9885
62	14/95	313/15	808/37	0/9832
63	20/33	313/15	812/67	0/9780
64	23/95	313/15	815/42	0/9747
65	29/32	313/15	819/37	0/9700
66	40/05	313/15	826/61	0/9615
67	61/79	313/15	839/8	0/9464
68	79/03	313/15	849/22	0/9359
69	99/76	313/15	859/79	0/9244
70	124/4	313/15	871/29	0/9122
71	149/18	313/15	881/82	0/9013
72	173/77	313/15	891/32	0/8917

73	199/08	313/15	900/3	0/8828
74	224/79	313/15	908/85	0/8745
75	249/97	313/15	916/92	0/8668

ادامہی جدول

۵-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
76	274/55	313/15	924/28	0/8599
77	300/31	313/15	931/65	0/8531
78	321/78	313/15	937/36	0/8479
79	2/55	323/15	787/62	0/9967
80	5	323/15	790	0/9937
81	9/97	323/15	794/64	0/9879
82	15/25	323/15	799/33	0/9821
83	19/67	323/15	803/09	0/9775
84	24/37	323/15	806/89	0/9729
85	28/14	323/15	809/81	0/9694
86	39/65	323/15	818/08	0/9596
87	44/19	323/15	821/24	0/9559
88	59/06	323/15	830/63	0/9451
89	78/61	323/15	841/94	0/9324
90	99/16	323/15	852/64	0/9207
91	123/3	323/15	864/28	0/9083
92	148/38	323/15	875/27	0/8969
93	174/07	323/15	885/53	0/8865
94	198/61	323/15	894/51	0/8776
95	223/69	323/15	902/95	0/8694
96	249/29	323/15	911/23	0/8615
97	274/4	323/15	918/91	0/8543
98	299/71	323/15	926/28	0/8475
99	2/55	338/15	773/12	0/9964

100	5	338/15	775/76	0/9930
101	9/78	338/15	780/72	0/9867
102	14/8	338/15	785/57	0/9806

ادامه‌ی جدول
۵-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
103	17/51	338/15	788/15	0/9774
104	23/85	338/15	793/75	0/9705
105	27/89	338/15	797/2	0/9663
106	39/96	338/15	806/63	0/9550
107	59/45	338/15	820/03	0/9394
108	79/25	338/15	832/07	0/9258
109	100/22	338/15	843/55	0/9132
110	124/72	338/15	855/74	0/9002
111	149/59	338/15	867	0/8885
112	174/45	338/15	877/27	0/8781
113	198/87	338/15	886/56	0/8689
114	223/7	338/15	895/32	0/8604
115	249/84	338/15	903/94	0/8522
116	274/62	338/15	911/64	0/8450
117	299/48	338/15	919/14	0/8381
118	323/72	338/15	926/1	0/8318

جدول ۴-۶- مقادیر مولکولی p , T , k و دانسیته‌ی مولکولی ۲- هگزانون

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
1	2/55	278/15	826/77	0/9978
2	5	278/15	828/51	0/9957
3	10	278/15	831/94	0/9916
4	15/1	278/15	835/39	0/9875
5	20/25	278/15	838/62	۰/۹۸۳۷

6

23/85

278/15

840/93

۰/۹۸۱۰

ادامه‌ی جدول

۶-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
7	29/46	278/15	844/28	۰/۹۷۷۱
8	39/06	278/15	849/76	۰/۹۷۰۸
9	49/23	278/15	855/31	۰/۹۶۴۵
10	59/44	278/15	860/58	۰/۹۵۸۶
11	69/56	278/15	865/45	۰/۹۵۳۲
12	78/89	278/15	869/83	۰/۹۴۸۴
13	89/95	278/15	874/81	۰/۹۴۳۰
14	100/06	278/15	879/1	۰/۹۳۸۴
15	120/26	278/15	887/23	۰/۹۲۹۸
16	140/2	278/15	894/84	۰/۹۲۱۹
17	162/4	278/15	902/67	۰/۹۱۳۹
18	209/81	278/15	917/94	۰/۸۹۸۷
19	209/71	278/15	917/94	۰/۸۹۸۷
20	2/55	288/15	817/84	۰/۹۹۷۶
21	5	288/15	819/73	۰/۹۹۵۳
22	9/56	288/15	823/04	۰/۹۹۱۳
23	14/39	288/15	826/46	۰/۹۸۷۲
24	19/89	288/15	830/24	۰/۹۸۲۷
25	24/17	288/15	833/04	۰/۹۷۹۴
26	29/91	288/15	836/71	۰/۹۷۵۱
27	39/93	288/15	842/76	۰/۹۶۸۱
28	49/58	288/15	848/2	۰/۹۶۱۹
29	59/88	288/15	853/7	۰/۹۵۵۷
30	69/07	288/15	858/37	۰/۹۵۰۵
31	79/97	288/15	863/64	۰/۹۴۴۷

32	89/93	288/15	868/23	۰/۹۳۹۷
33	99/72	288/15	872/6	۰/۹۳۵۰
34	120/11	288/15	881/08	۰/۹۲۶۰
35	139/68	288/15	888/66	۰/۹۱۸۱
36	159/82	288/15	896/08	۰/۹۱۰۵

ادامه‌ی جدول

۶-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
37	179/57	288/15	902/92	۰/۹۰۳۶
38	199/41	288/15	909/46	۰/۸۹۷۱
39	219/63	288/15	915/69	۰/۸۹۱۰
40	2/55	298/14	809	۰/۹۹۷۵
41	5	298/14	810/95	۰/۹۹۵۱
42	9/28	298/14	814/14	۰/۹۹۱۲
43	14/78	298/14	818/19	۰/۹۸۶۳
44	19/95	298/14	821/86	۰/۹۸۱۹
45	23/67	298/14	824/37	۰/۹۷۸۹
46	29/25	298/14	828/01	۰/۹۷۴۶
47	39/83	298/14	834/61	۰/۹۶۶۹
48	60/34	298/14	846/25	۰/۹۵۳۶
49	80/33	298/14	856/48	۰/۹۴۲۲
50	100/92	298/14	866/04	۰/۹۳۱۸
51	125/65	298/14	876/68	۰/۹۲۰۵
52	149/85	298/14	886/11	۰/۹۱۰۷
53	175/41	298/14	895/45	۰/۹۰۱۲
54	200/58	298/14	903/88	۰/۸۹۲۸
55	225/62	298/14	911/74	۰/۸۸۵۱
56	249/98	298/14	918/9	۰/۸۷۸۲
57	275/79	298/14	926/29	۰/۸۷۱۲
58	300/95	298/14	933/03	۰/۸۶۴۹
59	323/42	298/14	938/68	۰/۸۵۹۷
60	2/55	313/14	795/55	۰/۹۹۷۲

90	99/65	323/14	850/66	۰/۹۲۳۰
91	124/84	323/14	862/24	۰/۹۱۰۶
92	149/66	323/14	872/69	۰/۸۹۹۷
93	174/96	323/14	882/4	۰/۸۸۹۸
94	199/84	323/14	891/21	۰/۸۸۱۰
95	224/57	323/14	899/38	۰/۸۷۳۰
96	250/81	323/14	907/59	۰/۸۶۵۱

ادامه‌ی جدول

۶-۴

شماره	p/MPa	T/K	ρ/Kgm^{-3}	k
97	275/43	323/14	914/89	۰/۸۵۸۲
98	300/44	323/14	921/87	۰/۸۵۱۷
99	325/52	323/14	928/63	۰/۸۴۵۵
100	2/55	338/13	773	۰/۹۹۶۵
101	5	338/13	775/49	۰/۹۹۳۳
102	9/41	338/13	779/8	۰/۹۸۷۸
103	15/27	338/13	785/29	۰/۹۸۰۹
104	20/11	338/13	789/47	۰/۹۷۵۷
105	23/94	338/13	792/72	۰/۹۷۱۷
106	28/34	338/13	796/25	۰/۹۶۷۴
107	39/69	338/13	804/65	۰/۹۵۷۳
108	59/93	338/13	817/98	۰/۹۴۱۷
109	79/81	338/13	829/52	۰/۹۲۸۶
110	99/58	338/13	839/92	۰/۹۱۷۱
111	125/9	338/13	852/56	۰/۹۰۳۵
112	150/77	338/13	863/36	۰/۸۹۲۲
113	175/5	338/13	873/15	۰/۸۸۲۲
114	200/53	338/13	882/35	۰/۸۷۳۰
115	225/54	338/13	890/82	۰/۸۶۴۷
116	250/26	338/13	898/72	۰/۸۵۷۱
117	274/34	338/13	906/12	۰/۸۵۰۱
118	300/83	338/13	913/75	۰/۸۴۳۰

۴-۳- محاسبه‌ی توصیف‌کننده‌ها

توصیف‌کننده‌های مولکولی مقادیری هستند که شکل یا ساختار مولکول را توصیف می‌کنند. به عبارتی، توصیف‌کننده‌های مولکولی نتیجه‌ی یک فرایند ریاضی هستند که اطلاعات شیمیایی مربوط به ساختار مولکول را به اعداد تبدیل می‌کنند. از این اعداد برای تفسیر خواص مولکولی و یا ویژگی‌های مولکول استفاده می‌شود. با توجه به موضوع مورد بررسی (بررسی ارتباط کمی ساختار- خاصیت) ابتدا ساختار مولکول‌ها با استفاده از نرم‌افزار Hyperchem رسم و ساختار هندسی آن‌ها با استفاده از روش نیمه تجربی AM1 بهینه گردید و بهینه‌سازی تا زمانی ادامه یافت که گرادیان به $0/001$ کیلو کالری بر مول برسد. سپس ساختارهای بهینه شده به عنوان ورودی به نرم‌افزار Dragon وارد شدند. محاسبه‌ی توصیف‌کننده‌ها توسط این نرم‌افزار صورت گرفت که حدود ۱۴۸۲ توصیف‌کننده از ۱۸ گروه مختلف برای ترکیبات به دست آمد.

۴-۴- انتخاب توصیف‌کننده‌ها با استفاده از روش ورود مرحله‌ای

برای ساختن مدلی که بیانگر ارتباط ساختاری ترکیبات مورد بررسی با دانسیته باشد، از روش رگرسیون خطی چندگانه (MLR) بر پایه‌ی ورود مرحله‌ای استفاده شد.

لازم به تذکر است که از ابتدا داده‌های مربوط به مولکول ۲ - هگزانون کنار گذاشته شد تا بعداً در ارزیابی مدل مورد استفاده قرار گیرد.

در این مرحله توصیف‌کننده‌هایی که بیش از ۹۰٪ مقادیر آنها یکسان بود، حذف شدند. سپس با استفاده از روش رگرسیون مرحله‌ای چهار توصیف‌کننده‌ی خودهمبستگی ساختار توپولوژی (ATS4e)، تابع توزیع شعاعی (RDF040e)، فشار (P) و دما (T) به عنوان توصیف‌کننده‌هایی که بیشترین ارتباط را با دانسیته دارند، انتخاب شدند.

جدول (۷-۴) ماتریس همبستگی متغیرهای مستقل وارد شده را برای سری داده‌ها نمایش

می‌دهد.

جدول ۷-۴- ماتریس همبستگی متغیرهای مستقل وارد شده

	P	T	ATS4e	RDF040e
P	۱	۰/۰۲۶	-۰/۰۹۷	۰/۰۲۲
T	۰/۰۲۶	۱	۰/۱۸۸	۰/۱۱۰
ATS4e	-۰/۰۹۷	۰/۱۸۸	۱	-۰/۰۴۳
RDF040e	۰/۰۲۲	۰/۱۱۰	-۰/۰۴۳	۱

جدول (۸-۴) نام متغیرهای مستقل ، ضرایب رگرسیون و ضرایب استاندارد شده را برای مدل

منتخب این سری از داده‌ها نشان می‌دهد. در زیر این جدول N تعداد داده‌های سری آموزش، F

مقدار مربوط به آماره‌ی F یا آزمون فیشر و R ضریب همبستگی می‌باشد.

جدول ۸-۴- مشخصات مدل منتخب توسط روش رگرسیون خطی چندگانه

توصیف کننده	علامت	ضریب رگرسیون	ضرایب استاندارد شده
فشار	P	۰/۴۵۰	۰/۹۵۶
دما	T	-۰/۷۵۲	-۰/۳۳۱
خود همبستگی ساختار توپولوژی	ATS4e	۱۵۴/۸۷۱	۰/۱۳۷
تابع توزیع شعاعی	RDF040e	-۰/۵۸۳	-۰/۰۱۵

$$N = ۳۲۴$$

$$F = ۲۲۳۶/۸۲۲۰$$

$$R = ۰/۹۹۲$$

۴-۴-۱- مدلسازی به روش MLR

در این مرحله داده‌های مربوط به ۵ کتون (استون، ۲- پنتانول، ۲- بوتانول، ۴- متیل- ۲- پنتانول و ۳- پنتانول) به صورت گزینشی به سه سری آموزش، تست و ارزیابی به ترتیب به نسبت ۶۰٪، ۲۰٪، ۲۰٪ و به ترتیب با تعداد ۳۲۴، ۱۰۷ و ۱۰۷ مولکول تقسیم شدند. سری‌های تست و ارزیابی برای سنجش قدرت پیش‌بینی مدل به‌دست‌آمده از سری آموزش به کار می‌روند. در نهایت با استفاده از سری آموزش و انتخاب روش ورود اجباری، چندین مدل مختلف به‌طور جداگانه به‌دست آمد که عبارتند از:

$$\rho = (1018.470) + (0.439 \times p) + (-0.696 \times T) \quad \text{مدل شماره (۲)}$$

$$\rho = (884.666) + (0.450 \times p) + (-0.757 \times T) + (156.088 \times ATS\ 4e) \quad \text{مدل شماره (۳)}$$

$$\rho = (885.351) + (0.450 \times p) + (-0.752 \times T) + (154.871 \times ATS\ 4e) + (-0.583 \times RDF\ 040e) \quad \text{مدل شماره (۴)}$$

از بین این مدل‌ها، مدلی که دارای بیشترین مقدار R و F ، کمترین مقدار خطای استاندارد برای سری ارزیابی و شامل توصیف‌کننده‌های تا حد امکان قابل توجیه باشد به عنوان بهترین مدل و مدل نهایی انتخاب گردید. مقادیر R ، F و خطای استاندارد برای مدل‌های ۲ تا ۴ در جدول (۴-۹) آمده است.

جدول ۴-۹- مقادیر R ، F و خطای استاندارد برای سری ارزیابی برای مدل‌های ۲ تا ۴

مدل	R	F	خطای استاندارد
۲	۰/۹۷۰۹	۱۷۲۵/۳۸۰۴	۱۲/۰۹۰۳
۳	۰/۹۷۷۳	۲۲۳۶/۸۲۲۰	۱۰/۶۰۲۱
۴	۰/۹۷۷۴	۲۲۴۹/۲۷۱۸	۱۰/۷۷۵۸

با توجه به جدول (۹-۴)، مدل شماره ۴، با ۲ توصیف‌کننده‌ی ساختاری ATS4e و RDF040e و

۲ توصیف‌کننده‌ی T و p، به عنوان مدل نهایی انتخاب گردید

جدول (۱۰-۴) مقادیر متغیرهای مستقل وارد شده در مدل منتخب به غیر از دما و فشار را برای

مولکول‌های مورد بررسی نشان می‌دهد.

جدول ۱۰-۴- مقادیر توصیف‌کننده‌های وارد شده در مدل منتخب

نام ترکیب	ATS04e	RDF040e
استون	۰/۸۹	۰/۱
۲- بوتانول	۰/۹۹	۰/۶۵
۴- متیل - ۲ - بوتانول	۰/۹۶	۰/۶۳
۲- پنتانول	۰/۹۴	۳/۶۴
۳- پنتانول	۱/۰۳	۱/۶۶

اثر این توصیف‌کننده‌ها بر روی دانسیته در انتهای این فصل مورد بررسی قرار می‌گیرد.

۴-۵- مدل‌سازی به روش شبکه عصبی مصنوعی

در این بررسی، مدل شبکه عصبی مصنوعی با استفاده از نرم‌افزار MATLAB، برای پیش‌بینی دانسیته‌ی کتون‌ها طراحی گردید. برای این منظور مدل شبکه عصبی با استفاده از ۴ توصیف‌کننده مورد استفاده در مدل MLR طراحی گردید. در این شبکه به ازای هر توصیف‌کننده یک ورودی وجود دارد. لایه خروجی نیز شامل یک نرون می‌باشد که دانسیته‌ی متناظر با ورودی را نشان می‌دهد. تعداد نرون‌های لایه مخفی نامعلوم است و جزء پارامترهایی است که باید بهینه شوند. علاوه بر تعداد نرون‌های لایه مخفی، تعداد دوره‌های آموزش، سرعت آموزش، نوع تابع انتقال و نوع الگوریتم آموزش نیز باید بهینه شوند. ملاک انتخاب مقدار بهینه برای هر یک از این پارامترها، زمانی است که مقدار میانگین مربع خطا (MSE) مطابق معادله (۳-۴) برای هر دو سری ارزیابی و آموزش، مینیمم مقدار باشد.

$$MSE = \frac{\sum_{i=1}^n (Y - \hat{Y})^2}{n} \quad (3-4)$$

که در این رابطه \hat{Y} خروجی شبکه در n امین تکرار، Y مقدار واقعی دانسیته و n تعداد مولکول‌های سری تست و ارزیابی می‌باشد.

۴-۵-۱- انتخاب تابع انتقال، تابع آموزش، تعداد دوره‌های آموزش و

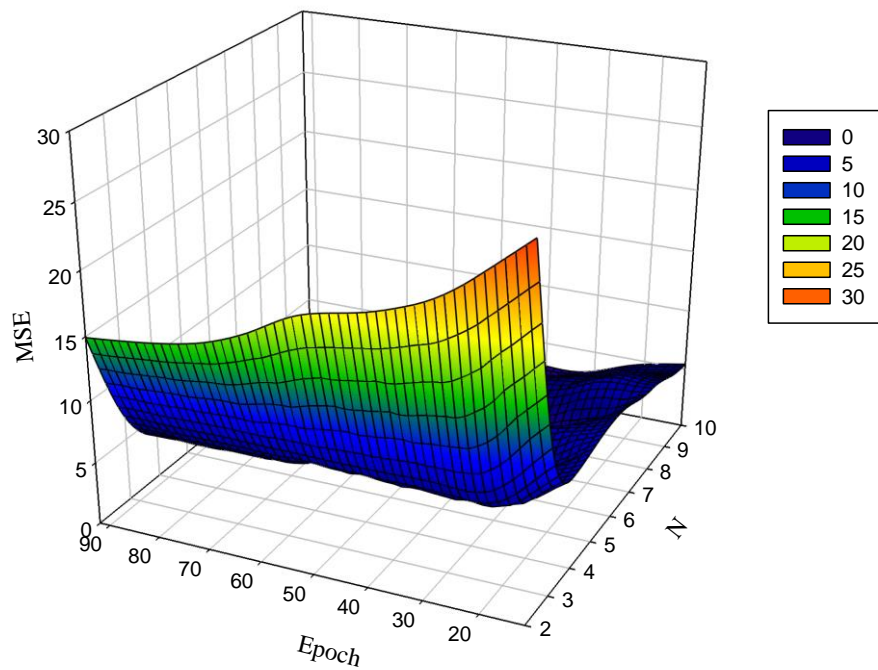
تعداد گره در لایه پنهان

در مدل‌سازی شبکه‌ی عصبی مصنوعی، یک شبکه با استفاده از توصیف‌کننده‌های مدل منتخب MLR به عنوان ورودی طراحی می‌شود. ماتریس‌های بردارهای ورودی p و بردار هدف t به شکل تصادفی مطابق با همان تعداد الگوهای به کاررفته در مدل‌سازی MLR به سه دسته‌ی آموزش، تست و ارزیابی به ترتیب ۶۰٪، ۲۰٪ و ۲۰٪ تقسیم می‌شود. در شروع برنامه، مقدار وزن‌های ورودی و

بهینه می‌شوند. برای آموزش شبکه، دو الگوریتم آموزشی بایزین (trainbr) و لونبرگ - مارکواردت (trainlm) مورد بررسی قرار گرفتند. همچنین برای انتخاب تابع انتقال مناسب دو تابع انتقال تانژانت سیگموئید (tansig) و لگاریتم سیگموئید (logsig) مورد استفاده قرار گرفتند. به این ترتیب با استفاده از دو الگوریتم آموزشی و دو تابع انتقال انتخاب شده، چهار شبکه‌ی عصبی مختلف طراحی شدند. سپس برای هر یک از این شبکه‌ها از روش تغییر همزمان برای تغییر پارامترهای شبکه (تعداد نرون‌های لایه مخفی و تعداد دوره‌های آموزش) جهت آموزش شبکه استفاده گردید. تعداد گره‌های لایه پنهان از ۲ تا ۱۰ در تعداد دوره‌های آموزش مختلف تغییر داده شدند. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطا به عنوان معیار انتخاب گردید. در تمام شبکه‌ها مقدار ضریب ممنتوم (μ) ثابت و برابر ۰/۰۰۵ در نظر گرفته شد. شبکه‌های آموزش یافته برای پیش‌بینی دانسیته سری ارزیابی مورد استفاده قرار گرفت. به این ترتیب در مجموع ۳۶۰ شبکه‌ی سه لایه آموزش داده شد و سپس برای سری ارزیابی مورد استفاده قرار گرفت که نتایج حاصله در جداول (۴-۱۱)، (۴-۱۲)، (۴-۱۳) و (۴-۱۴) و نمودارهای (۴-۱)، (۴-۲)، (۴-۳) و (۴-۴) نشان داده شده است. با توجه به نتایج، شبکه عصبی با تابع تبدیل لگاریتم سیگموئید (logsig) و آموزش با الگوریتم تنظیم بایزین (trainbr) و تعداد دور آموزشی ۴۲ و تعداد ۷ گره در لایه پنهان کمترین مقدار میانگین مربعات خطا را ایجاد می‌کند. لذا شبکه‌ی با این مشخصات برای مطالعات بعدی انتخاب گردید.

جدول ۴-۱۱- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل logsig و تابع آموزش br

		تعداد گره								
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد دوره‌ها		43 8/3905	81 9/3252	15 2/7136	120 7/4773	177 1/5592	294 7/4970	205 0/1777	235 8/0596	917/054 3
	۲	27/ 8806	11/ 3110	7/2 712	7/68 56	6/73 32	7/51 29	6/83 61	6/03 47	5/4064
	۳	24/ 4368	7/1 857	5/7 163	5/17 52	6/52 60	3/94 85	6/46 28	5/21 07	4/5278
	۴	21/ 9111	6/8 423	5/2 525	5/47 14	5/60 93	2/71 05	4/89 34	4/59 72	3/0863
	۵	20/ 6064	6/7 952	5/2 025	3/49 93	5/15 95	2/63 65	4/87 56	3/03 72	3/0236
	۶	19/ 6527	6/7 919	5/1 260	2/98 11	3/08 47	2/65 02	2/94 52	2/97 80	3/0314
	۷	17/ 4896	6/8 271	4/0 093	3/00 74	2/96 65	2/66 16	2/78 49	2/93 24	2/9647
	۸	16/ 0366	6/8 880	3/7 240	2/97 36	2/90 93	2/66 78	2/73 95	2/87 71	2/8925
	۹	15/ 3566	6/8 332	3/5 692	2/94 91	2/87 33	2/67 13	2/70 46	2/83 91	2/8470
	۱۰	14/ 9180	6/8 450	3/4 799	2/93 25	2/84 32	2/67 34	2/70 24	2/79 71	2/8266

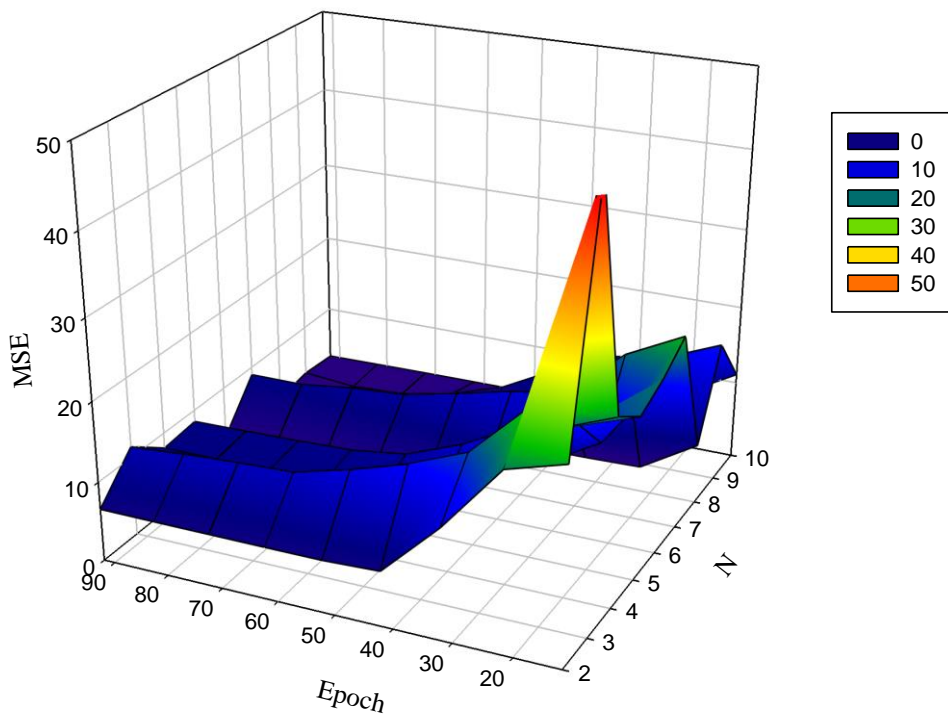


شکل ۴-۱- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی

با تابع تبدیل logsig و تابع آموزش br

جدول ۴-۱۲ - مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل logsig و تابع آموزش lm

		تعداد گره								
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد دوره‌ها		379/ 2242	644/ 3506	323 4/5168	137 0/7866	464/ 8424	7820/ 5188	971 /8962	363 5/1230	466 8/6964
	۲	24/1 241	52/8 330	23/3 546	20/4 000	21/9 759	24/39 75	7/5 740	17/8 073	11/0 949
	۳	22/0 477	26/1 769	20/5 455	17/9 339	12/6 926	20/86 45	3/2 321	14/8 506	7/06 03
	۴	13/7 320	19/3 787	13/6 966	16/3 349	8/57 66	14/54 46	3/0 461	13/7 307	4/06 26
	۵	6/85 49	16/5 534	8/92 49	12/6 762	6/88 01	13/47 11	2/9 943	10/5 684	3/45 80
	۶	6/55 02	14/3 697	8/31 91	9/96 66	5/82 24	11/82 17	2/9 082	6/89 68	3/40 64
	۷	6/53 70	12/5 475	8/17 27	9/04 94	5/00 60	10/31 59	2/8 390	4/58 67	3/37 78
	۸	6/54 01	12/2 609	8/11 81	8/80 60	4/34 76	9/861 9	2/8 204	3/24 03	3/33 58
	۹	6/54 79	11/8 574	8/09 08	8/64 71	3/70 04	9/001 6	2/8 026	2/91 34	3/26 12
	۱۰	6/57 89	11/3 575	8/07 69	8/52 74	3/42 12	8/994 0	2/7 900	3/53 41	3/17 52

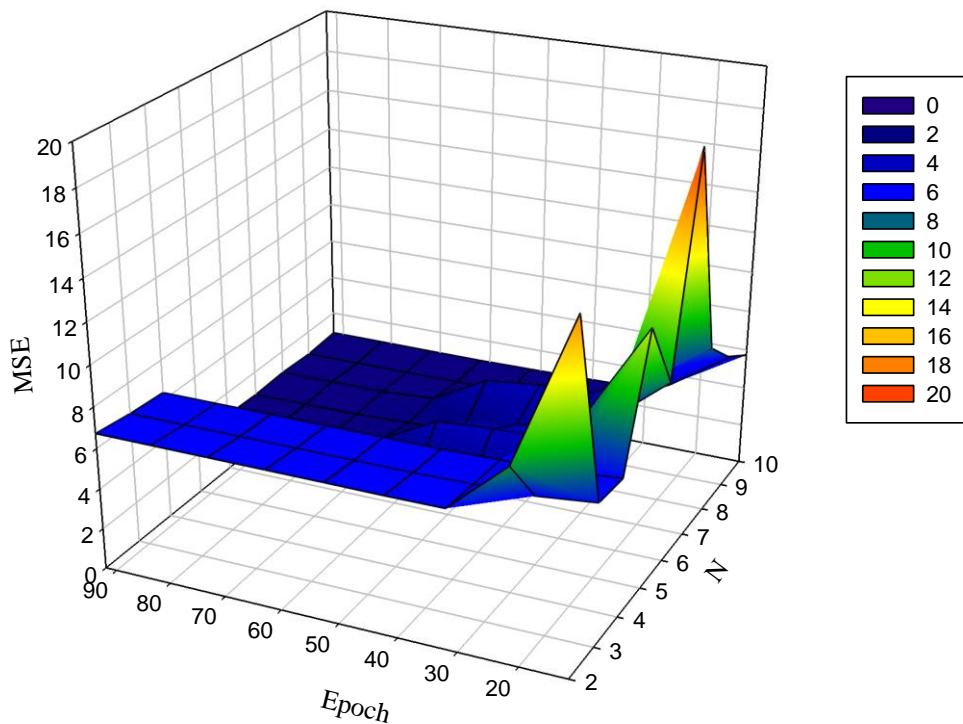


شکل ۴-۲ - نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی

با تابع تبدیل logsig و تابع آموزش lm

جدول ۴-۱۳- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل tansig و تابع آموزش br

		تعداد گره								
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد دوره	۱	1494 /3643	149 7/9486	27 7/8604	2167/ 9553	2691 /7960	2798 /4494	636/ 8350	202 0/3044	108 6/7603
	۲	16/3 507	6/90 14	6/7 401	12/52 05	8/86 79	18/7 325	8/28 42	6/73 55	5/73 05
	۳	9/25 46	6/60 45	6/4 917	8/033 5	6/51 49	6/83 29	6/59 05	4/79 09	4/97 69
	۴	6/73 10	6/54 90	6/4 374	3/815 9	5/01 82	4/53 55	4/95 69	3/16 36	2/98 28
	۵	6/71 30	6/50 55	6/3 697	2/914 0	4/94 42	3/92 07	4/82 82	2/85 28	2/92 02
	۶	6/72 38	6/49 68	6/3 805	2/811 3	4/75 60	3/02 01	4/52 54	2/80 00	2/88 19
	۷	6/72 38	6/49 68	6/3 727	2/796 1	3/32 16	2/90 40	3/02 66	2/78 61	2/84 88
	۸	6/72 38	6/49 68	6/3 721	2/786 0	2/87 99	2/82 30	2/86 59	2/77 54	2/85 47
	۹	6/72 38	6/49 68	6/3 726	2/784 0	2/87 09	2/79 24	2/86 41	2/77 00	2/86 87
	۱۰	6/72 38	6/49 68	6/3 729	2/727 7	2/84 97	2/78 02	2/85 71	2/76 58	2/87 86

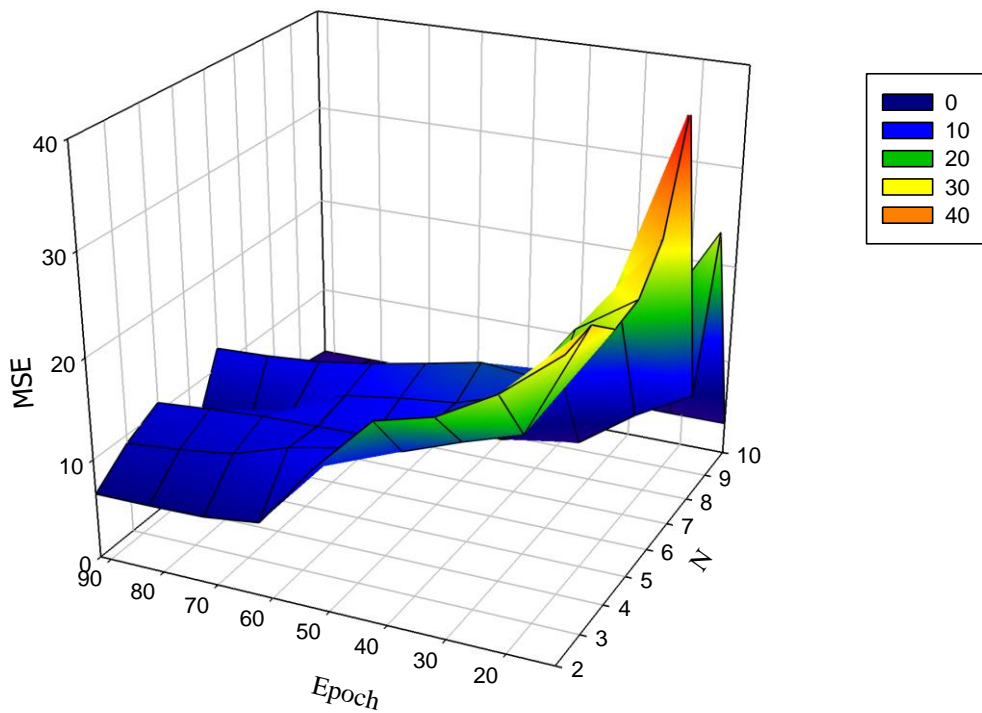


شکل ۴-۳- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی

با تابع تبدیل tansig و تابع آموزش br

جدول ۴-۱۴- مقادیر میانگین مربعات خطا برای شبکه‌هایی با تابع تبدیل tansig و تابع آموزش Im

		تعداد گره								
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد دوره		504/3 717	73 3/5559	457/ 1441	280/ 6314	1444 /8223	4399/ 5711	516 9/6367	5120 /0175	594 /0336
	۲	28/66 71	29/ 0359	26/4 778	27/0 544	30/5 928	40/77 95	11/0 598	25/4 839	3/2 580
	۲	24/15 22	18/ 0790	19/2 450	23/3 090	21/0 826	20/99 05	7/83 59	16/1 082	3/2 394
	۲	20/99 83	16/ 2461	15/9 088	15/9 456	15/2 724	13/71 91	3/77 51	11/4 985	3/2 689
	۲	19/51 71	14/ 1643	14/3 402	14/0 604	15/8 003	5/419 6	3/17 30	7/67 31	3/2 492
	۲	13/94 83	13/ 3422	11/9 768	13/4 245	14/6 514	5/001 3	2/98 33	5/19 59	3/0 906
	۲	7/371 0	12/ 1956	11/8 733	12/5 893	13/5 628	4/477 3	2/93 66	4/98 46	2/8 897
	۲	6/680 7	10/ 4232	11/7 507	9/92 76	12/7 343	3/494 8	2/92 33	4/42 72	2/8 806
	۲	6/560 6	9/6 657	11/3 653	4/37 86	12/2 865	3/464 5	2/81 09	4/19 32	2/8 616
	۲	6/528 8	9/1 765	10/9 307	3/62 06	12/0 504	3/439 9	2/74 43	3/41 63	2/8 426



شکل ۴-۴- نمودار مقادیر میانگین مربعات خطا بر حسب تعداد گره و دوره برای شبکه‌هایی

با تابع تبدیل tansig و تابع آموزش Im

توابع و مقادیر بهینه‌ی پارامترهای تجربی برای شبکه‌ی عصبی در جدول (۴-۱۵) نمایش داده

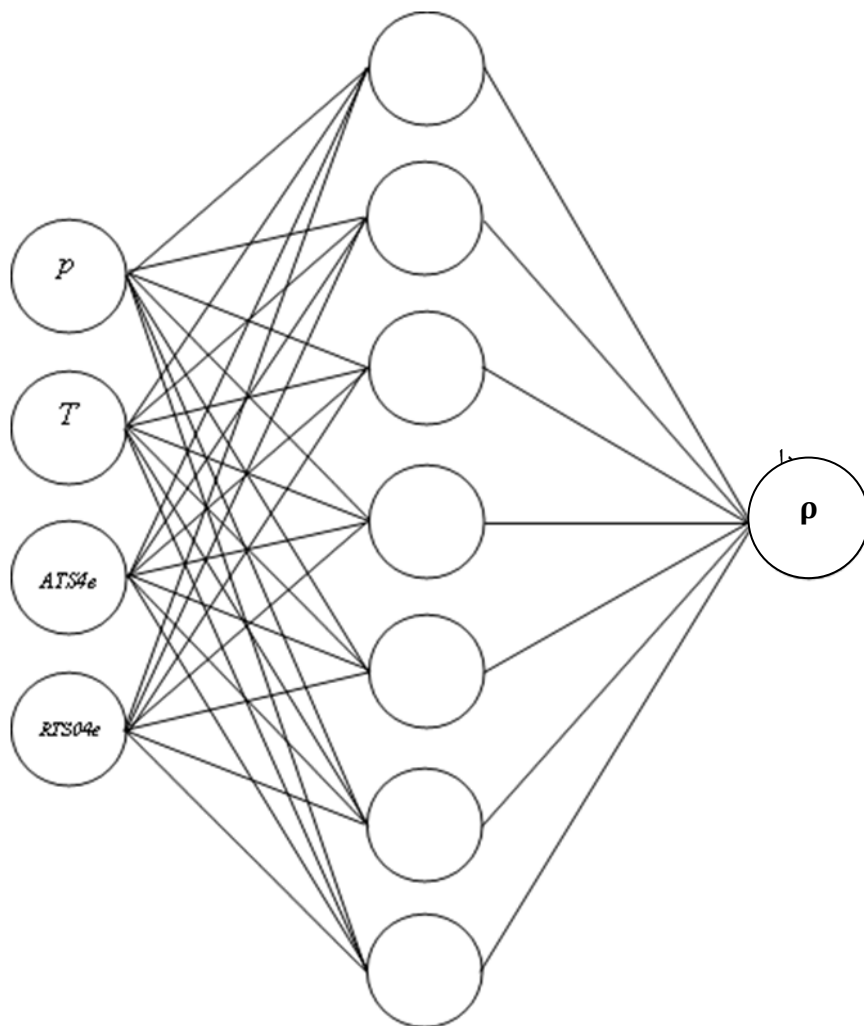
شده است.

جدول ۴-۱۵- توابع و مقادیر پارامترهای مهم شبکه‌ی عصبی

trainbr	تابع آموزش
logsig	تابع انتقال لایه‌ی پنهان
purelin	تابع انتقال لایه‌ی خروجی
MSE	تابع کارآیی
۷	تعداد نورون‌های لایه‌ی پنهان
۴۲	تعداد دوره‌های آموزش
۰/۰۰۵	پارامتر تطبیق مارکواردت (μ)

۴-۶- هندسه‌ی شبکه عصبی مصنوعی بهینه‌شده

با توجه به نتایج حاصل در بهینه‌سازی شبکه، ساختار شبکه‌ی عصبی به‌دست‌آمده پس از دانسیته بهینه‌سازی فاکتورهای بحث شده در بالا به صورت شماتیک در شکل (۴-۵) نشان داده شده است.



لایه ورودی

لایه پنهان

لایه خروجی

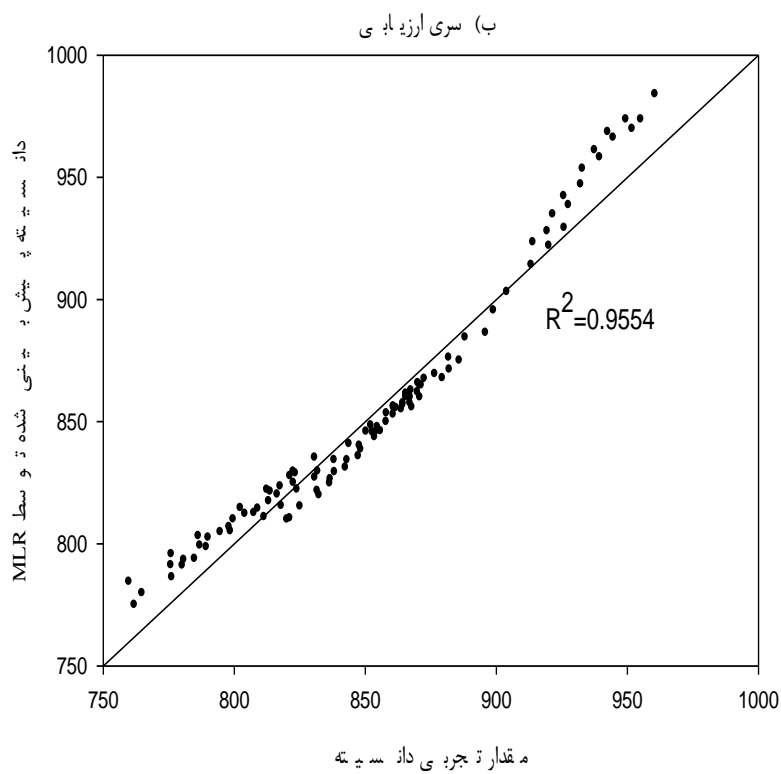
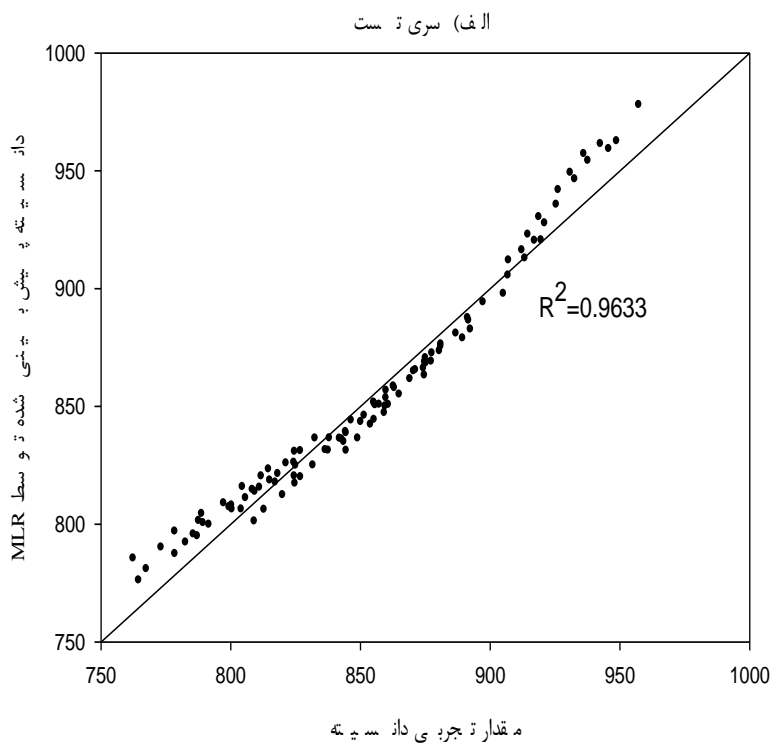
شکل ۴-۵- تصویر شماتیک ساختار شبکه‌ی عصبی مصنوعی به‌دست‌آمده پس از بهینه‌سازی

۷-۴- ارزیابی شبکه و مقایسه‌ی روش‌های رگرسیون خطی چندگانه و شبکه‌ی

عصبی مصنوعی

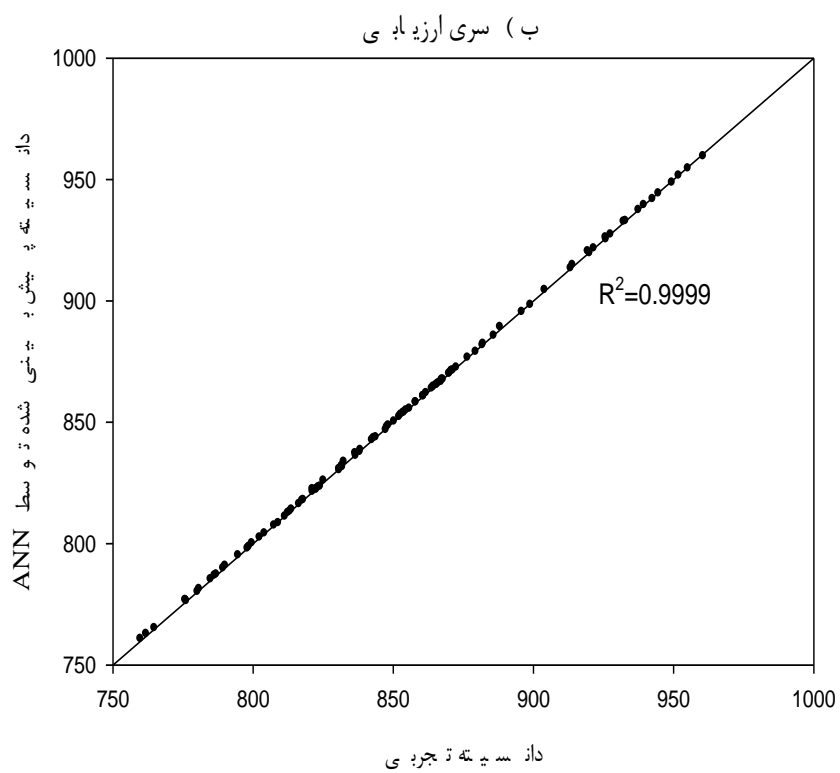
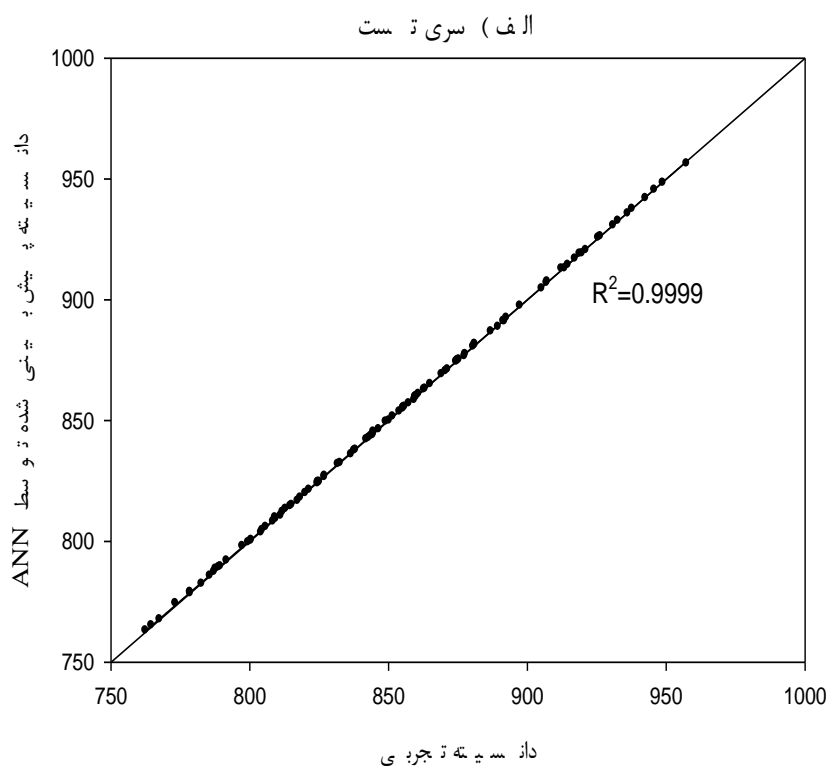
۷-۴-۱- نمودار برگشت

در این نمودار مقادیر پیش‌بینی دانسیته بر حسب مقادیر تجربی دانسیته رسم می‌شود و به کمک مقدار مربع ضریب همبستگی (R^2) به دست آمده از نمودار، پراکندگی نقاط در اطراف خط برگشت تعیین می‌شود. محدوده‌ی تغییرات R^2 بین صفر و یک است: اگر $R^2 = 1$ باشد، همبستگی کاملی بین داده‌ها وجود دارد و اگر $R^2 = 0$ باشد، داده‌ها هیچ‌گونه همبستگی ندارند. بنابراین هر چه مقدار R^2 به دست آمده در نمودار مربوط به مقادیر دانسیته پیش‌بینی شده بر حسب دانسیته تجربی، به ۱ نزدیک‌تر باشد، می‌توان گفت که نتایج پیش‌بینی شده به واقعیت نزدیک‌ترند. شکل (۴-۶) نمودار برگشت دانسیته پیش‌بینی شده بر حسب دانسیته تجربی با استفاده از روش آنالیز رگرسیون چندگانه و شکل (۴-۷) نمودار برگشت دانسیته پیش‌بینی شده بر حسب دانسیته تجربی را با استفاده از مدل ANN بهینه شده برای دو سری تست و ارزیابی، نشان می‌دهد.



شکل ۴-۶- نمودار برگشت دانسیته پیش‌بینی شده توسط MLR بر حسب دانسیته تجربی برای

الف) سری تست و ب) سری ارزیابی



شکل ۴-۷- نمودار برگشت دانسیته پیش‌بینی شده توسط ANN بر حسب دانسیته تجربی برای

الف) سری تست و ب) سری ارزیابی

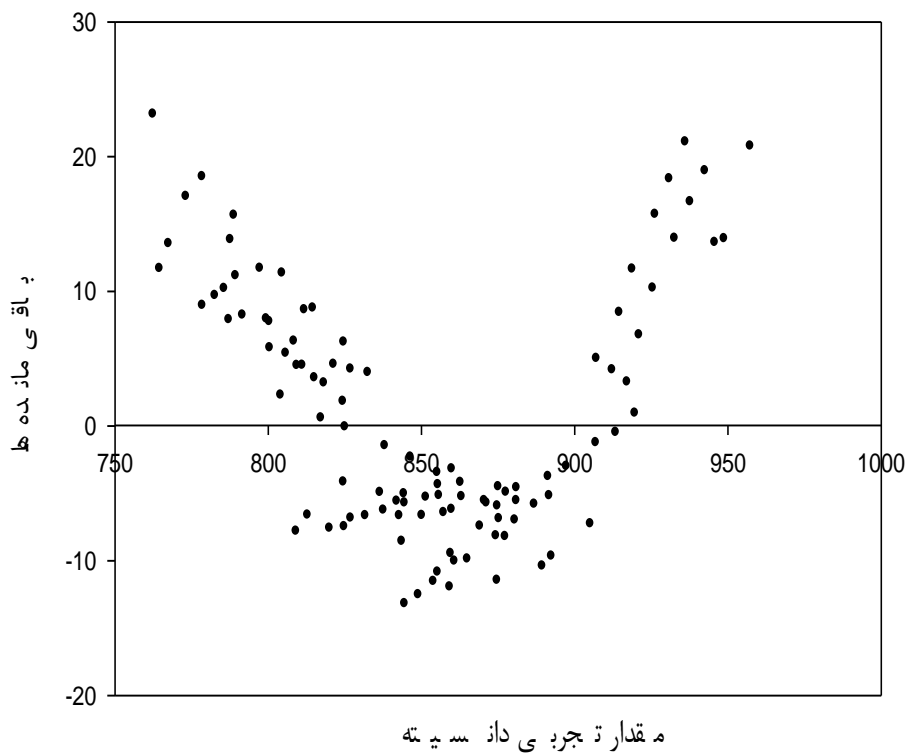
۴-۷-۲- نمودار خطای باقیمانده

منظور از عبارت خطای باقیمانده، تفاوت بین مقادیر پیش‌بینی و مقادیر تجربی است. اگر پراکندگی مقادیر در دو طرف نمودار صفر باشد، این امر نشان‌دهنده‌ی تصادفی بودن خطاهاست. ولی اگر عمده‌ی نقاط، در این نمودار، در یک طرف صفر باشد، این بدان معناست که خطای جهت دار رخ داده است.

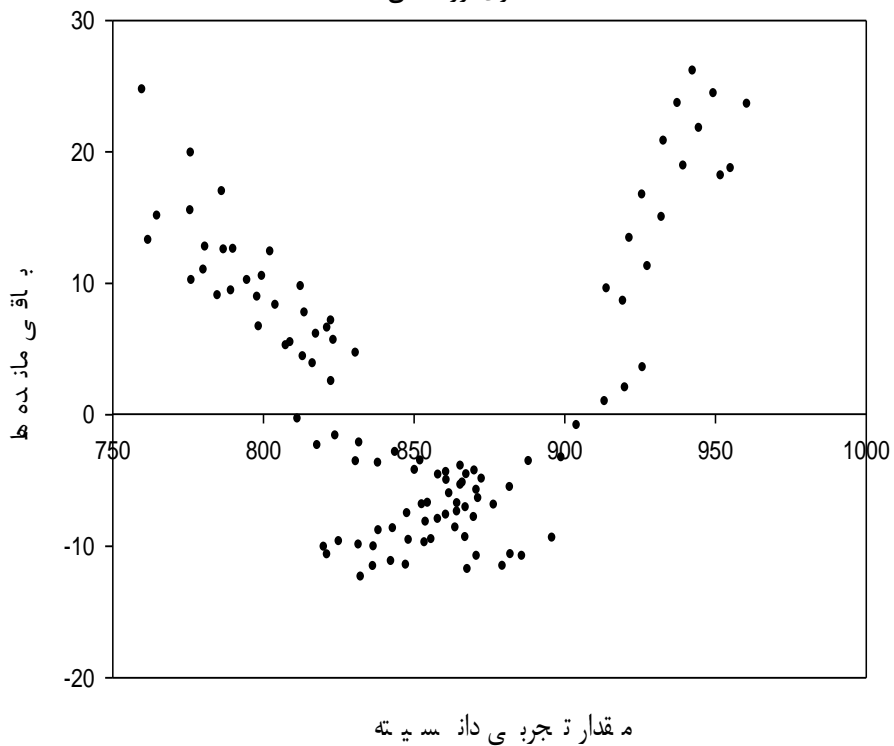
در شکل (۴-۸) نمودار خطای باقیمانده (که از کم کردن مقادیر تجربی از مقادیر نظری محاسبه می‌شود) با استفاده از روش MLR نشان داده شده است.

هم‌چنین در شکل (۴-۹) نمودار خطای باقیمانده با استفاده از روش ANN نشان داده شده است.

الف) سری تست



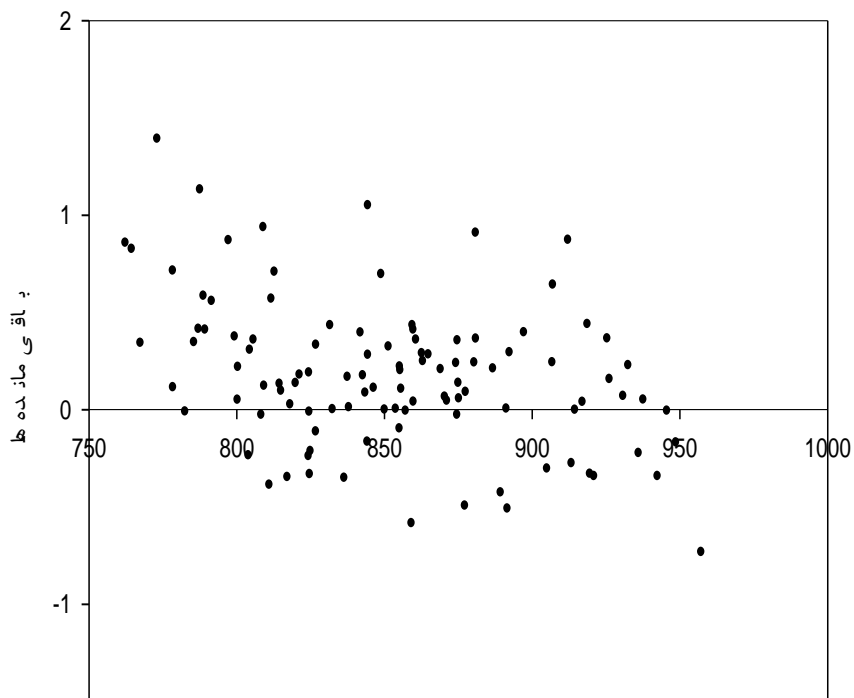
ب) سری ارزیابی



شکل ۴-۸- نمودار خطای باقیمانده نسبت به مقدار تجربی با استفاده از روش MLR برای

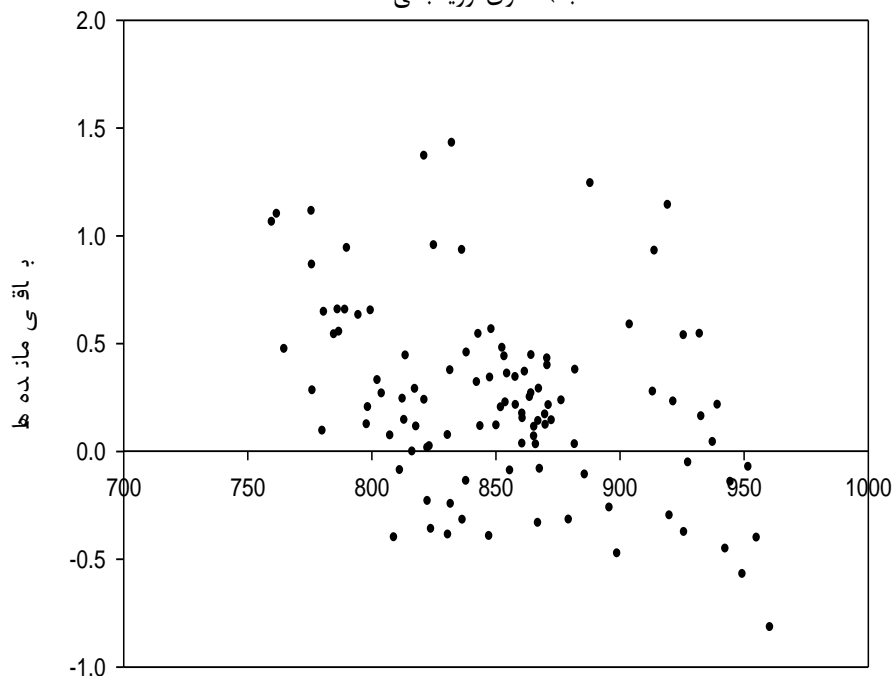
الف) سری تست و ب) سری ارزیابی

الف) سری تست



مقدار تجربی دانسیته

ب) سری ارزیابی



مقدار تجربی دانسیته

شکل ۴-۹- نمودار خطای باقیمانده نسبت به مقدار تجربی با استفاده از روش ANN برای

الف) سری تست و ب) سری ارزیابی

در نمودار های شکل (۷-۴) نقاط در اطراف خط راست نسبت به نمودار های شکل (۶-۴) متراکم تر هستند و این نشان دهنده ی پیش بینی بهتر شبکه ی عصبی مصنوعی نسبت به روش رگرسیون خطی چندگانه است.

همانگونه که در شکل (۹-۴) ملاحظه می شود، در مدل شبکه ی عصبی مصنوعی، نقاط نسبت به نتایج مدل رگرسیون خطی چندگانه (شکل (۸-۴))، به صورت متقارن تری حول خط صفر انتشار یافته اند، که این امر نشان دهنده ی خطای سیستماتیک کمتر در مدل شبکه ی عصبی مصنوعی می باشد. بنابراین فرضیه ی وجود رابطه ی غیر خطی بین دانسیته و توصیف کننده های وارد شده در مدل ، تا حدود زیادی ثابت می شود.

مقادیر تجربی دانسیته و مقادیر پیش بینی شده ی آن برای دو سری تست و ارزیابی، توسط مدل های رگرسیون خطی چندگانه و مدل شبکه ی عصبی مصنوعی بهینه شده در جدول (۴-۱۶) آورده شده است.

جدول ۴-۱۶- مقادیر تجربی دانسیته و مقادیر پیش‌بینی شده توسط تکنیک‌های ANN و MLR

(v: سری ارزیابی، t: سری تست)

ش ماره	دانسیت تهی تجربی	دانسیت‌هی پیش‌بینی	
		ML R	AN N
1 v	/39	/8695	/0723
5	889	878	888
2 t	/36	/9727	/9333
5	945	959	945
9 t	/67	/3069	/6632
6	951	969	951
0 v	/75	/9328	/6765
6	822	829	822
1 v	/46	/5982	/4747
6	824	830	824
2 t	/61	/8467	/5981
6	855	851	855
9 t	/12	/6807	/0223
7	860	856	860
0 v	/68	/2891	/7146
7	866	860	866
1 v	/13	/9388	/1607
7	871	865	871
2 t	/19	/4920	/2347
7	919	920	919
9 t	/61	/5720	/2792
8	925	929	925
0 v	/84	/4262	/4648
8	812	822	812
1 v	/43	/1922	/6724
8	814	823	814
2 t	/56	/3213	/6919
8	846	844	846
9 t	/38	/0819	/4912
9	852	848	852
0 v	/08	/5709	/2832
9	858	853	858
1 v	/03	/4547	/2439
ش	دانسیت	پیش‌بینی	دانسیت‌هی پیش‌بینی
ماره	تهی تجربی	ML R	AN N
9	863	857	863
2 t	/06	/8243	/3081
9	913	912	913
9 t	/36	/9043	/0831
1		922	919
00 v	920	/0568	/7019
1	802	814	802
01 v	/31	/7081	/6387
1	808	814	808
v	/96	/4513	/5603
	811	815	810
2 t	/08	/5988	/6927
9	860	850	861
t	/74	/7215	/0982
1	870	860	871
0 v	/77	/0071	/1679
1	882	871	882
1 v	/01	/3673	/3875
1	892	882	892
2 t	/35	/7046	/6435
1	948	962	948
9 t	/69	/6027	/5199
2	955	973	954
0 v	/09	/8161	/6879
2	797	806	798
1 v	/97	/9213	/0938
2	800	808	800
2 t	/3	/0688	/3496
2	874	863	874
9 t	/64	/1947	/6120
3	885	875	885
0 v	/84	/0828	/7312
3	895	886	895
1 v	/81	/4293	/5483
3	905	897	904
2 t	/08	/8354	/7754
3	957	978	956
9 t	/24	/0365	/5069
4	960	984	959
0 v	/48	/1045	/6634
4	786	799	787
1 v	/84	/3913	/3931
4	789	800	789
2 t	/37	/5388	/7802
4	855	844	855
9 t	/22	/3871	/4399
5	867	855	867
0 v	/71	/9401	/6278
5	879	867	879

شماره	دانشجوی تجربی	دانشجوی پیش‌بینی ML R	دانشجوی پیش‌بینی AN N	شماره	دانشجوی تجربی	دانشجوی پیش‌بینی ML R	دانشجوی پیش‌بینی AN N
1	804	806	803	1	804	815	804
69 _t	/01	/3171	/7735	02 _t	/49	/8556	/7967
1	811	811	811	1	838	836	838
70 _v	/33	/0127	/2413	09 _t	/03	/5748	/0410
1	817	815	818	1	843	840	843
71 _v	/91	/5752	/0240	10 _v	/78	/9262	/8958
1	824	820	824	1	850	846	850
72 _t	/5	/3579	/6899	11 _v	/24	/0165	/3592
1	880	875	881	1	855	850	855
79 _t	/94	/4150	/8481	12 _t	/7	/5514	/8054
1	888	884	889	1	906	905	907
80 _v	/08	/5216	/3224	19 _t	/83	/6038	/0722
1	822	825	822	1	913	914	913
81 _v	/52	/0527	/2888	20 _v	/25	/2514	/5259
1	824	826	824	1	786	803	786
82 _t	/34	/1773	/0996	21 _v	/27	/2478	/9268
				1	788	804	789
				22 _t	/8	/4550	/3838
				1	824	824	824
				29 _t	/99	/9356	/7753
				1	831	829	831
				30 _v	/82	/6771	/5748
				1	838	834	837
				31 _v	/05	/3772	/9117
				1	844	839	844
				32 _t	/28	/2564	/1139
				1	897	894	897
				39 _t	/23	/2400	/6264
				1	903	903	904
				40 _v	/96	/1400	/5474
				1	775	795	776
				41 _v	/93	/8464	/7947
				1	778	796	779
				42 _t	/43	/9434	/1431
				1	817	817	816
				49 _t	/19	/7957	/8423
				1	823	822	823
				50 _v	/89	/2848	/5285
				1	830	827	830
				51 _v	/71	/1456	/3228
				1	836	831	836
				52 _t	/46	/5382	/1084
				1	891	886	891
				59 _t	/64	/4805	/1307
				1	898	895	898
				60 _v	/9	/6330	/4250
				1	759	784	760
				61 _v	/77	/5055	/8327
				1	762	785	763
				62 _t	/37	/5382	/2267

ش ماره	دانسیه تهی تجربی	دانسیتهی پیش‌بینی		21 v	/07	/0003	/0590
		ML R	AN N				
2	767	781	767	2	805	811	806
82 t	/46	/0123	/8021	22 t	/72	/1248	/0789
2	812	806	813	2	842	836	842
89 t	/82	/2206	/5276	29 t	/74	/0990	/9152
2	825	815	826	2	853	845	854
90 v	/06	/4098	/0151	30 v	/88	/7196	/1053
2	836	824	837	2	863	855	863
91 v	/38	/8423	/3130	31 v	/72	/1154	/9706
2	848	836	849	2	874	866	874
92 t	/96	/4458	/6552	32 t	/28	/1451	/5186
2	912	916	913	2	932	946	932
99 t	/19	/3760	/0610	39 t	/57	/5115	/7971
3	919	927	920	2	939	958	939
00 v	/34	/9841	/4816	40 v	/35	/2802	/5650
3	816	820	816	2	789	798	789
01 v	/35	/2378	/3477	41 v	/28	/7053	/9361
3	818	821	818	2	791	799	792
02 t	/15	/3623	/1754	42 t	/58	/8298	/1368
3	851	846	851	2	831	825	832
09 t	/43	/1529	/7533	49 t	/66	/0151	/0931
3	861	855	862	2	842	834	843
10 v	/7	/6955	/0678	50 v	/97	/3053	/5135
3	870	864	871	2	853	843	853
11 v	/73	/9857	/1598	51 v	/49	/7515	/9293
3	880	876	881	2	864	855	865
12 t	/99	/4194	/3532	52 t	/93	/0751	/2126
3	936	957	935	2	925	935	925
19 t	/08	/1804	/8559	59 t	/46	/7030	/8250
3	942	968	942	2	932	947	932
20 v	/47	/6233	/0171	60 v	/15	/1642	/6937
3	807	812	807	2	780	791	780
21 v	/46	/7078	/5323	61 v	/16	/1753	/2546
3	809	813	809	2	782	792	782
22 t	/33	/8323	/4515	62 t	/6	/2998	/5896
3	844	838	844	2	824	817	824
29 t	/46	/7606	/7410	69 t	/77	/3107	/4375
3	854	847	854	2	836	826	836
30 v	/58	/8488	/9393	70 v	/57	/5320	/2510
31 v	/31	/5475	/7547	2	847	835	846
3	874	868	875	71 v	/32	/8773	/9259
32 t	/73	/8022	/0843	2	859	847	858
3	937	961	937	72 t	/22	/2880	/6355
40 v	/45	/1392	/4919	2	920	927	920
3	798	805	798	79 t	/99	/7737	/6474
41 v	/49	/1853	/6940	2	927	938	927
3	800	806	800	80 v	/47	/7484	/4170

ادامه‌ی جدول ۴-۱۶

00 v	/49	/9111	/7201	42 t	/49	/3099	/7080
4	761	775	762	3	837	831	837
01 v	/81	/0728	/9105	49 t	/54	/3024	/7067
4	764	776	765	3	847	840	848
02 t	/5	/1974	/3254	50 v	/71	/1887	/0508
4	809	801	810	3	857	849	858
09 t	/08	/2863	/0169	51 v	/94	/9791	/2831
4	821	810	822	3	869	861	869
10 v	/16	/5076	/5294	52 t	/08	/6653	/2870
4	832	819	833	3	926	941	926
11 v	/33	/9860	/7600	59 t	/2	/9260	/3563
4	844	831	845	3	932	953	932
12 t	/46	/2774	/5091	60 v	/81	/6351	/9711
4	907	912	907	3	784	793	785
19 t	/04	/0660	/6811	61 v	/82	/8903	/3619
4	913	923	914	3	787	795	787
20 v	/92	/5089	/8495	62 t	/11	/0149	/5238
4	830	835	830	3	826	820	827
21 v	/66	/3477	/7342	69 t	/88	/0533	/2123
4	832	836	832	3	838	829	838
22 t	/49	/4723	/4908	70 v	/22	/4031	/6764
4	859	856	859	3	848	838	848
29 t	/87	/7004	/9095	71 v	/21	/6566	/7755
4	865	861	865	3	859	850	859
30 v	/44	/5382	/5085	72 t	/49	/0398	/9223
4	870	865	870	3	918	930	919
31 v	/07	/7932	/1916	79 t	/73	/3969	/1684
4	875	870	875				
32 t	/03	/5301	/1665				
4	917	920	917				
39 t	/06	/3407	/0983				
4	821	827	821				
40 v	/22	/8177	/4576				
4	823	828	823				
41 v	/28	/9423	/3027				
4	826	831	826				
42 t	/85	/0858	/7365				
4	862	858	863				
49 t	/73	/5569	/0190				
4	867	862	867				
50 v	/39	/8348	/6789				
4	872	867	872				
51 v	/48	/5855	/6224				
4	877	872	877				
52 t	/52	/6069	/6098				
4	811	820	812				
59 t	/73	/3703	/2995				

ادامه‌ی جدول ۴-۱۶

ش ماره	دانسیه ته‌ی تجریبی	دانسیتته‌ی پیش‌بینی	
		ML R	AN N
3	925	942	926
80 v	/69	/4135	/2270
3	776	786	776
81 v	/14	/3603	/4218
3	778	787	778
82 t	/48	/4390	/5932
3	819	812	820
89 t	/99	/4223	/1259
3	831	821	832
90 v	/63	/7263	/0043
3	842	831	842
91 v	/41	/2459	/7294
3	853	842	853
92 t	/85	/3308	/8527
3	914	922	914
99 t	/52	/9679	/5185
4	921	934	921

19 t	/12	/1677	/5106
5	775	791	776
20 v	/76	/2923	/8742
5	780	793	781
21 v	/72	/4863	/3653
5	785	795	785
22 t	/57	/7905	/9156
5	843	834	843
29 t	/55	/9982	/6359
5	855	846	855
30 v	/74	/2437	/6498
5		857	866
31 t	867	/6591	/6664
5	877	869	876
32 t	/27	/0698	/7753

ش ماره	دانسیه تهی تجربی	دانسیتتهی پیش‌بینی	
		ML R	AN N
4	813	821	814
60 v	/69	/4352	/1329
4	817	823	817
61 v	/47	/5925	/7587
4	821	825	821
62 t	/29	/8737	/4690
4	855	851	855
69 t	/4	/0545	/6014
4	860	855	860
70 v	/77	/7776	/9218
4	865	860	865
71 v	/46	/1060	/5722
4	870	864	870
72 t	/49	/9576	/5553
4	797	808	798
79 t	/26	/9835	/1297
4	799	810	800
80 v	/58	/1173	/2327
4	804	812	804
81 v	/03	/3664	/2974
4	808	814	808
82 t	/37	/6843	/3425
4	859	853	860
89 t	/79	/6121	/1999
4	871	864	871
90 v	/29	/9219	/5027
4	881	876	881
91 v	/82	/2959	/8514
4	891	887	891
92 t	/32	/5827	/3245
4	787	801	788
99 t	/62	/4627	/7501
5		802	790
00 v	790	/5873	/9420
5	794	804	795
01 v	/64	/8685	/2713
5	799	807	799
02 t	/33	/2920	/7048
5	841	836	842
09 t	/94	/3742	/3352
5	852	845	853
10 v	/64	/8067	/1187
5	864	856	864
11 v	/28	/8870	/5477
5	875	868	875
12 t	/27	/3987	/3264
5	773	790	774

هم‌چنین مقادیر تجربی دانسیته، مقادیر پیش بینی شده و درصد خطا برای مولکول ۲ - هگزانون در جدول (۱۹-۴) گزارش شده است.

جدول ۱۹-۴ - مقادیر دانسیته تجربی، دانسیته پیش بینی شده و درصد خطا برای مولکول ۲- هگزانون

شماره	دانسیته‌ی	دانسیته‌ی	درصد خطا
	تجربی	پیش‌بینی	
1	826/77	830/63	0/47
2	828/51	832/28	0/46
3	831/94	835/58	0/44
4	835/39	838/82	0/41
5	838/62	841/97	0/40
6	840/93	844/11	0/38
7	844/28	847/35	0/36
8	849/76	852/61	0/34
9	855/31	857/85	0/30
10	860/58	862/81	0/26
11	865/45	867/44	0/23
12	869/83	871/52	0/19
13	874/81	876/12	0/15
14	879/1	880/13	0/12
15	887/23	887/71	0/05
16	894/84	894/73	-0/01
17	902/67	902/10	-0/06
18	917/94	916/68	-0/14
19	917/94	916/65	-0/14
20	817/84	822/41	0/56
21	819/73	824/16	0/54
22	823/04	827/34	0/52
23	826/46	830/59	0/50
24	830/24	834/16	0/47

25	833/04	836/84	0/46
26	836/71	840/31	0/43
27	842/76	846/05	0/39
28	848/2	851/22	0/36
29	853/7	856/42	0/32
30	858/37	860/79	0/28

ادامه‌ی جدول ۴-

۱۹

شماره	دانشیته‌ی	دانشیته‌ی	درصد خطا
	تجربی	پیش‌بینی	
31	863/64	865/7	0/24
32	868/23	869/95	0/20
33	872/6	873/95	0/15
34	881/08	881/78	0/08
35	888/66	888/80	0/02
۳۶	896/08	895/63	-0/05
۳۷	902/92	902/02	-0/10
38	909/46	908/19	-0/14
39	915/69	914/25	-0/16
40	809	814/04	0/62
41	810/95	815/9	0/61
42	814/14	819/06	0/60
43	818/19	822/98	0/59
44	821/86	826/51	0/57
45	824/37	828/98	0/56
46	828/01	832/55	0/55
47	834/61	838/95	0/52
48	846/25	850/09	0/45
49	856/48	859/66	0/37
50	866/04	868/49	0/28

51	876/68	878/08	0/16
52	886/11	886/7	0/07
53	895/45	895/21	-0/03
54	903/88	903/17	-0/08
55	911/74	910/76	-0/11
56	918/9	917/85	-0/11
57	926/29	925/07	-0/13
58	933/03	931/8	-0/13
59	938/68	937/51	-0/12
60	795/55	801/05	0/69
61	797/71	803/08	0/67
62	801/82	807/06	0/65
63	805/97	811/06	0/63
64	809/84	814/82	0/61

ادامه‌ی جدول ۴-۴

۱۹

شماره	دانشیته‌ی تجربی	دانشیته‌ی پیش‌بینی	درصد خطا
65	812/58	817/5	0/61
66	816/85	821/64	0/59
67	823/12	827/80	0/57
68	835/34	839/7	0/52
69	845/94	849/76	0/45
70	855/79	859/02	0/38
71	867/21	869/41	0/25
72	877/57	878/62	0/12
73	886/89	887/07	0/02
74	895/7	895/14	-0/06
75	903/55	902/57	-0/11

76	911/34	910/15	-0/13
77	918/51	917/4	-0/12
78	925/15	924/12	-0/11
79	932	931/02	-0/11
80	787/6	792	0/56
81	789/9	794/17	0/54
82	793/89	798/02	0/52
83	798/09	802/05	0/50
84	802/17	806/11	0/49
85	806/04	809/92	0/48
86	810/28	814/12	0/47
87	818/3	822/12	0/47
88	829/63	833/35	0/45
89	840/55	843/95	0/40
90	850/66	853/37	0/32
91	862/24	863/84	0/19
92	872/69	873/07	0/04
93	882/4	881/73	-0/08
94	891/21	889/74	-0/16
95	899/38	897/39	-0/22
96	907/59	905/26	-0/26
97	914/89	912/44	-0/27
98	921/87	919/55	-0/25

ادامه‌ی جدول ۴-

۱۹

شماره	دانشیته‌ی تجربی	دانشیته‌ی پیش‌بینی	درصد خطا
99	928/63	926/46	-0/23
100	773	777/74	0/61

101	775/49	780/1	0/59
102	779/8	784/25	0/57
103	785/29	789/56	0/54
104	789/47	793/78	0/55
105	792/72	797	0/54
106	796/25	800/6	0/55
107	804/65	809/31	0/58
108	817/98	823/05	0/62
109	829/52	834/62	0/61
110	839/92	844/61	0/56
111	852/56	856/11	0/42
113	873/15	874/2	0/12
114	882/35	882/30	-0/01
115	890/82	890/04	-0/09
116	898/72	897/46	-0/14
117	906/12	904/53	-0/18
118	913/75	912/17	-0/17
119	920/52	919/13	-0/15

۴-۷-۵- ارزیابی شبکه با استفاده از آزمون Y - تصادفی

برای حصول اطمینان از تصادفی نبودن نتایج، یک مجموعه اعداد تصادفی در دامنه متغیر هدف تعریف کرده و شبکه با آن آزموده شد. نتایج حاصل از چندین مجموعه تصادفی ایجاد شده، در جدول (۴-۲۰) نشان داده شده است. با توجه به نتایج (ضریب همبستگی پایین) نشان‌دهنده‌ی عدم همبستگی تصادفی در مدل شبکه عصبی مصنوعی ایجاد شده می‌باشد.

جدول ۴-۲۰- مقادیر R^2 برای سری تست و سری ارزیابی بعد از چندین تست Y - تصادفی

تکرار	R^2 سری ارزیابی	R^2 سری تست
۱	۰/۰۱۱۰	۰/۰۰۵۳
۲	۰/۰۰۴۷	۰/۰۰۲۲
۳	۰/۰۰۵۸	۰/۰۰۰۳
۴	۰/۰۰۱۴	۰/۰۰۲۵
۵	۰/۰۰۰۹	۰/۰۰۱۰
۶	۰/۰۰۷۴	۰/۰۰۰۰
۷	۰/۰۰۱۷	۰/۰۰۱۸
۸	۰/۰۰۳۹	۰/۰۰۳۲
۹	۰/۰۰۰۲	۰/۰۱۰۷
۱۰	۰/۰۰۱۱	۰/۰۰۱۶

۴-۸- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با

دانشیته

در این قسمت با توجه به توصیف‌کننده‌های وارد شده در مدل‌های MLR و ANN یک بررسی اجمالی روی اثرات مختلف موجود در دانشیته‌ی ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل چهار توصیف‌کننده است که هر کدام بیان‌گر خصوصیات متفاوتی از مولکول‌های مورد بررسی است. این توصیف‌کننده‌ها عبارتند از: ATS4e، RDF040e، T، P که در ادامه به بررسی مختصر این توصیف‌کننده‌ها می‌پردازیم.

۴-۸-۱- توصیف‌کننده (ATS4e)

خودهمبستگی Moreau - Broto (خودهمبستگی یک ساختار توپولوژیکی، ATS)

$$ATS_d = \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i \cdot w_j)_d = W^T \cdot {}^m B W$$

در رابطه‌ی فوق، w خاصیت اتمی، A تعداد اتم‌ها، d فاصله توپولوژیکی، δ_{ij} دلتای کرونیکر) اگر $\delta_{ij} = 1$ و $d_{ij} = d$ اگر $\delta_{ij} = 1$ و $d_{ij} \neq d$ ، ماتریس دوتایی و W بردار بدون بعد خواص اتمی است.

خودهمبستگی ATS_0 برای طول مسافت صفر به صورت زیر تعریف می‌شود:

$$ATS_0 = \sum_{i=1}^A w_i^2$$

خواص اتمی شامل جرم اتمی، قطبیت، بار و الکترون‌گاتیویته می‌باشد.

برای هر خاصیت اتمی w ، سری آرایش‌های خودهمبسته برای همه‌ی فواصل موجود در نمونه،

به صورت زیر تعریف می‌شود:

$$\langle ATS_0, ATS_1, ATS_2, \dots, ATS_D \rangle_w$$

با توجه به موارد گفته شده، با افزایش جرم اتمی m ، خاصیت اتمی w افزایش پیدا کرده و در نتیجه، ضریب خودهمبستگی ATS_0 افزایش می‌یابد. از طرفی، بر اساس رابطه‌ی $\rho = \left(\frac{m}{v}\right)$ ، با افزایش m ، دانسیته افزایش می‌یابد. لذا با افزایش مقدار توصیف‌کننده‌ی ATS ، دانسیته افزایش می‌یابد.

همچنین با توجه به مدل خطی به‌دست آمده، می‌بینیم ضریب ATS مقداری مثبت است و با افزایش مقدار آن، میزان دانسیته نیز افزایش می‌یابد.

۴-۸-۲- توصیف‌کننده‌ی تابع توزیع شعاعی (RDF040e)

این دسته از توصیف‌کننده‌ها، بر اساس فاصله‌ی توزیع در نمایش هندسی یک مولکول، شکل گرفته‌اند. تابع توزیع شعاعی مجموعه‌ای از اتم‌های A ، به‌صورت احتمال توزیع اتم در اطراف کره‌ای به شعاع R شکل گرفته‌اند. تابع توزیع شعاعی به‌صورت زیر نمایش داده می‌شود:

$$g(R) = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A w_i w_j e^{-\beta(R-r_{ij})^2}$$

در این رابطه، f فاکتور اندازه‌گیری، w خواص اتمی اتم‌های i و j ، r_{ij} فاصله‌ی بین اتمی اتم-های i و j و β عامل دما می‌باشد. $g(R)$ برای یک سری از نقاط مجزا، در فواصل معین، محاسبه می‌گردد.

مقادیر ۱۲۸ کد RDF، با تنظیم پارامتر β در محدوده‌ی 100 \AA^{-2} و 200 \AA^{-2} ، و R حدود 0.2 \AA - 0.1 \AA به‌دست آمده است.

خواص اتمی w ، اتم‌های i و j کد RDF، توانایی تفکیک و تمایز میان هر یک از اتم‌های تشکیل‌دهنده‌ی یک مولکول را دارا می‌باشند.

با توجه به مدل خطی به‌دست آمده، مشاهده می‌شود مقدار ضریب RDF، منفی و رابطه‌ای معکوس با دانسیته دارد یعنی با کاهش مقدار RDF، میزان دانسیته کاهش می‌یابد.

۴-۸-۳- فشار (P)

فشار عبارتست از نیروی وارده بر واحد سطح، این کمیت در گازها نقش عمده را ایفا میکند زیرا یکی از کمیات مشخصه گاز میباشد، از این رو بیشتر قوانین فشار در گازها نهفته شده است.

عوامل موثر بر فشار:

(۱) مقدار نیرو:

فشار با نیروی وارده بر سطح رابطه مستقیم دارد یعنی هر چه نیروی بیشتر باشد فشار بیشتر است.

(۲) مساحت سطح:

مقدار فشار با مساحت سطح رابطه عکس دارد یعنی هر چه سطح بزرگتر باشد، مقدار فشار کم تر است و بر عکس.

برای محاسبه فشار از فرمول زیر استفاده می شود:

$$p = \frac{F}{A}$$

فشار با دانسیته رابطه مستقیم دارد، یعنی هر چه فشار بیشتر باشد دانسیته آن نیز بیشتر تر است. همچنین در مدل خطی حاصله، مشاهده می شود که فشار با دانسیته رابطه مستقیم دارد بدین منظور با افزایش فشار، دانسیته افزایش می یابد.

۴-۸-۴- دما (T)

دما معیاری از میزان سردی و گرمی جسمها را مشخص می کند.

با افزایش دما فاصله مولکولها از یکدیگر افزایش می یابد، بنابراین، حجم آنها زیاد می شود، از طرفی جرم آنها ثابت می ماند و چگالی با حجم رابطه عکس دارد، پس با افزایش دما حجم افزایش و

دانسیتة کاهش می‌یابد.

در مدل خطی حاصل شده نیز، دما رابطه‌ای معکوس با دانسیته را نشان می‌دهد. یعنی با

افزایش دما، میزان دانسیته کاهش می‌یابد.

آینده نگری

با توجه به توانایی شبکه‌ی عصبی (ANN) در پیش‌بینی دانسیته‌ی کتون‌ها، که در این پایان نامه نشان داده شد، می‌توان پیش‌بینی نمود که نه تنها در مورد کتون‌ها، بلکه در ترکیبات مختلف دیگر نیز می‌توان از این روش استفاده نمود.

هم‌چنین علاوه بر پیش‌بینی دانسیته، در پیش‌بینی سایر خواص انتقالی مانند ویسکوزیته و خواص ترمودینامیکی نیز، می‌توان از این روش استفاده کرد.

از آن‌جا که به غیر از روش MLR، روش‌های خطی دیگری از جمله حداقل مربعات جزئی (PLS) و آنالیز اجزای اصلی (PCA) قابل استفاده هستند، پیشنهاد می‌شود از این تکنیک‌ها نیز در این زمینه استفاده شود و نتایج با هم مقایسه گردند.

[1] Parsafar G.A., Kalantar Z., (2005), "Extension of linear isotherm regularity to long chain primary, secondly and tertiary alcohols, ketones and 1-carboxylic acids by group contribution method", *Fluid phase Equilibria*, Vol. 234, pp 11-21.

[۲] نامجو، ن، (۱۳۸۸)، پایان‌نامه ارشد: "مدل سازی شبکه‌ی عصبی موجک با استفاده از روش سهم گروه‌ها برای پیش‌بینی دانسیته‌ی کتون‌ها در محدوده‌ی وسیعی از دما و فشار"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[3] Livingstone D.,(1998), *Data Analysis for Chemists*, Oxford University Press.

[4] Wold S. "chemometrix: what do we mean with it, and what do we want from it?" And Brown, S. D., "Has the 'Chemometricsrevolution' ended? Some views on the past, present and future of Chemometrics" , *papers on chemometrics: Philosophy, History and Directions of InCINC'94*.
<http://WWW.emsl.pnl.gov:2080/docs/hncinic/homepage.html>.

[5] Hovery D.G., Hirsch R.F.,(1983), "chemometrics in the chemistry curriculum" , *J.Chem. Ed.* 60, pp: 656-659.

[6] Zupan J., Gasteiger J., (1991), "Neural networks-a new method for solving chemical problems or just a passing phase", *Anal. Chem. Acta*, 248, pp 1-30.

[7] Andersson G.G., Kaufmann P., (2000), "Development of a generalized neural network", *Chemometr. Intell. Lab. Syst.* 50,pp 101-105.

[8] Hibert D.b.,(1993), "Genetic algorithms in chemistry", *Chemometr. Intell. Lab. Syst.*, 19, pp 277-293.

[9] Trinajstic N. (1992) "*Chemical Graph Theory*", CRC Press, Boca Raton, FL.

[10] Karitzky A.R., Lobanov V.S., Karelson M. (1995), "QSPR: the correlation and quantitative prediction of chemical and physical properties from structure", *Chem. Soc. Rev.* 24, pp 279-287.

[11] Karitzky A.R., Maran U., Lobanov V., Karelson M. (2000), "Structurally diverse quantitative – structure – property relationship correlation of technologically relevant physical properties", *J. Chem. Inf. Comput.Sci.*4, pp 1-18.

[12] Yaffe D. L. (2001), "A Neural Network Approach for Estimating Physicochemical Properties using Quantitative Structure – Property Relationships (QSPRs)", *University of California*, Los Angeles.

[13] Taskinen J., Yliruusi J., 2003. "Prediction of physicochemical properties based on neural network modeling". *Adv. Drug Deliv. Rev.* 55, pp 1163-1183.

[14] Selassie C.D., Mekapati S.B., Verma R.P. (2002), "QSAR: then and now". *Curr. Top. Med. Chem.* 2, pp 1357-1379.

[15] Schneider G. (2004), "*Modeling Structure-Activity Relationship*", Schneider G.; Sung-san So, Landes Bioscience.

[16] Introduction to molecular descriptors, <http://WWW.qsar.pe.kr/descriptors>

[17] Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships (QSAR) Models, *Organisation for Economic Co-operation and Development*, 2007.

[18] Wold S., Sjström M., Eriksson L.,(2001), "PLS-regression: a based tool of Chemometrics". *Chemomet. Intell. Lab. Syst.* 58, pp 109-130.

[19] Geladi P., Kowalski B.R. (1986), "Partial least squares regression: A tutorial" *Anal. Chim. Acta.* 185, pp 1-17

[20] Tropsha A., Gramatica P., Gombar V.K. (2003), "The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models" *QSAR Comb. Sci.* 22(1), pp 69-77.

[21] Osten D.W. (1988), "Selection of Optimal Regression Model via Cross-Validation". *J.Chemom.*, 2, pp 39.

[22] Cruciani G., Baroni M., Clementi S., Costantino G., Rigganelli D., and Skagerberg B. (1992), "Predictive Ability of Regression Models. Part I: Standard Deviation of Prediction Errors (SDEP)" *J.Chemom.*, 6. Pp 335-346.

[23] Gramatica P. (2007), "*Principles of QSAR models validation: internal and external, QSAR*", *Comb. Sci.* 26, pp 694-701

[24] Karelson M. (2000), "*molecular Descriptors in QSAR/QSPR*", Wiley-VCH, New York.

[25] Frank I.E. and Todeschini R. (1994), "The Data Analysis Handbook". *Elsevier*, Amsterdam(The Netherlands), pp 366.

[26] Rusker G., rucker C., (1993), *J.Chem. Inf. Comput. Sci.*, 33, pp 686-695.

[27] Todeschini R., Consonni V. (2000), "*Handbook of molecular descriptors*", in: Mannhold R., Kubinyi H., Timmerman H. (Eds), *Methodes and Principles in Medicinal Chemistry*, weily-VCH, Weinheim.

[28] Karelson M., Lobanov V.S. and Katritzky A.R., (1996), *Chem. Rev.*, Vol. 96, pp 1027.

[29] Darlington R.B., (1990), *Regression and Linear Models*. McGraw-Hill, New York.

[30] *Hyperchem professional Release 7 for windows*, Molecular Modeling System, Hypercube, Inc., 2002. <http://WWW.hyper.com>

[31] Todeschini R., Consonni V., Mauri A., Pavan M., V. 13-20124, Milano, Italy, *Dragon Software version 3.0*.

[32] *SPSS for windows, statistical package for IBM PC*, SPSS Inc.(2005), <http://www.spss.com>

[33] MATLAB, (*The Language of Thechnical Computing*). Version 7.5.0.342(R2007b), The MathWork Inc., 2007.

[34] <http://WWW.mathtools.net/MATLAB/Chemometrics/index.html>

[35] *DRAGON web version 3.0-2003*. is developed by Milano Chemometrics and QSAR Research Group, Dipartimento di Scienze dell'Ambiente e del Territorio Universitadegli Studi di Milano, Bicocca.

[36] Leardi R. (1994), "Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection", *J.Chemomet.* pp 79-865.

[۳۷] کاوه، ع، ثروتی، ه، (۱۳۸۰)، شبکه‌های عصبی مصنوعی در تحلیل و طراحی سازه‌ها، مرکز

تحقیقات ساختمان و مسکن.

[38] Stamations V.K., (1996), *Understanding neural networks and fuzzy logic*, IEEE PRESS,.

[۳۹] منهاج، م، ب، (۱۳۷۹)، هوش محاسباتی، جلد اول، دانشگاه صنعتی امیرکبیر (واحد

تفرش)، چاپ اول.

[۴۰] حبشی نژاد، م، (۱۳۸۲)، پایان نامه ارشد: "پیش بینی گرانروی گازها در محدوده‌ی وسیعی

از دما و فشار با استفاده از شبکه‌های عصبی مصنوعی"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[41] Jalali-Heravi M. and Fatemi M.H.(1998), "prediction of flame ionization detector response factors using artificial neural network", *Journal of chromatography A.*, No. 825, pp 161-169.

[۴۲] حقیقی، س، (۱۳۸۰)، پایان نامه ارشد: "اندازه گیری پدیدات به روش اسپکتروفلوریمتری و آهن به روش اسپکتروفوتومتری"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[۴۳] شاندیزی، ف، (۱۳۷۹)، پایان نامه ارشد: "اندازه گیری اسپکتروفوتومتری منگنز و اندازه گیری همزمان نیکل و کبالت به روش استخراج"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[44] Homer J., Sotos G. and Robson, J.H. (1999), "Artificial neural networks for the prediction of liquid viscosity, density, heat of vaporization, boiling point and pitzer's acentric factor", *Phys. Chem. Chem. Phys.* Vol. 1, No. 17, pp 4075-4082.

[45] McCulloch W.S., Pitts W.H., (1943), "A logical calculus of the ideas immanent in neural nets", *Bull. Math Biophys.* 5, pp 115-133.

[46] Hebb d. O., (1949), "the organization of behavior", Weily, New York.

[47] Hegan M.T., Demuth B., Beale M., (1995), *Neural network design*, PWS.

[48] Hussain A.S., Yu X.Q., Johnson R.D., (1991), *Pharm. Res.*, 8, pp 1248.

[49] Zupan J. and Gasteiger J., (1999), "Neural Network for Chemistry and Drug Design". Weily-VCH Publishers, Weinheim(Germany).

[۵۰] بیل، آر، جکسون، تی، (۱۳۸۰)، آشنایی با شبکه های عصبی، ترجمه البرزی، دانشگاه صنعتی شریف.

[۵۱] سعیدی، مسعود، (۱۳۸۳)، عملیات شبکه های عصبی، جلد دوم، ماهنامه شبکه، شماره ۵۲.

[52] Taylor J.G., Mannion C.L.T., (1992), "Theory and application of neural networks", Springer-Verlag. New York.

[53] Allen D.M., (1971), "Mean Square Error of Prediction as a Criterion for Selecting Variables", *Technometrics*, 13, pp 469-475.

[۵۴] مختاری، م، (۱۳۸۰)، کاربردهای MATLAB و simulink در مهندسی، ترجمه صمدی

بخارایی، چاپ اول، مرکز انتشارات خراسان، ص ۱۸۰.

[55] Rumelhart D.E., Hinton G.E. and Williams R.J., (1986), *"Learning internal representations by error propagation"*, D.E. Rumelhart and J. McClelland, editors, *Parallel Data Processing*, Vol.1, Chapter 8, The M.L.T. Press, Cambridge, MA, pp 318-362.

[56] <http://www.rasekhood.net/article>.

[57] Hagan M.T. and Menhaj M., (1994), *"Training feed-forward networks with the Marquardt algorithm"*, IEEE Transactions on Neural Networks, Vol.5, No. 6, 1994, pp 989-993.

[58] Hirschel K., Schafer M., (2006), *"Bayesian, regularization, neural network for optimizing fluid flow processes"*, Compute. Methodes APPL. Mech. Engrg, 195, pp 481-500.

[59] Demuth H. and Geele M., (2000), *"Neural network toolbox user guide"*, version 4, the Math Work Inc., Natick.

[60] Wasserman Philip D., (1993), *Advanced methods in neural computing*, Van Nostrand Reinhold Co.

[61] Malhotra R., Woolf L.A., (1991), *"Thermodynamic properties of propanone (acetone) at temperatures from 278-323K and pressures up to 400 MPa"*, *J.Chem. Thermodynamics*, 23, pp 867-876.

[62] Malhotra R., Woolf L.A., (1996), *"Volumetric measurements of liquid pentan-2-one, hexan-2-one, and 4-methylpentan-2-one at temperatures from 278.15-338.13K and pressures in the range from 0.1-386 MPa"*, *J.Chem. Thermodynamics*, 28, pp 1411-1421.

[63] Malhotra R., Woolf L.A., (1992), *"Thermodynamic properties of butan-2-one at temperatures from 278-338K and pressures from 0.1-280 MPa; predictions for higher ketones"*, *J.Chem. Thermodynamics*, 24, pp 1207-1217.

[64] Malhotra R., Woolf L.A., and Price W.E., (1993), *"Thermodynamic properties of pentan-3-one at temperatures from 278-338K and pressures from 0.1-380 MPa"*, *J.Chem. Thermodynamics*, 25, pp 361-366.

[65] Fatemi N. H., Gharajani S., (2007), *Bioorj. Med. Chem*, 15, pp 7746-7754.

Abstract

An artificial neural network (ANN) model in quantitative structure property relationship (QSPR) was developed for density prediction of 6 ketones over a wide range of pressure molecular (2-300MPa) and temperatures (273.15-338.15 K). A large number of descriptors were calculated by Dragon software and a subset of calculated descriptors was selected from 18 classes of Dragon descriptors with a stepwise multiple linear regression (MLR) as a feature selection technique. The selected descriptors that appear in multiple linear regression models are: 2D-ATS4e (weighed by atomic sanderson electronegativities) and RDF040e (weighed by atomic sanderson electronegativites). Two calculated and two experimental descriptors contain: pressure and temperature, were selected as the most feasible descriptors in the construction of artificial neural network (ANN) models. The data set was randomly divided into three subsets: training (324 point), validation (107 point) and test set (107 point). After training and optimization of the ANN parameters, the performance of the model was investigated by the test set. The mean squares error (MSE) were 0.9360 and 0.1735 respectively, for the test data set in MLR and ANN methods. The results obtained using ANN were compared with the experimental values as well as with those obtained using regression models and showed the superiority of ANN over linear multiple regression model.

Keywords:Artificial Neural Network(ANN) – multiple linear regression(MLR) –
Descriptors - Ketones



Shahrood University of Technology

Faculty of Science

**QSPR Modeling of Thermodynamic Properties of
Ketones**

Mahboube sabouri

Supervisors:

Dr zahra Kalantar

Dr Mansor Arab Cham Jangali

Consultant:

Dr Hossein Nikoofard

Sep 2010