



دانشگاه صنعتی شاهرود

دانشکده شیمی

گزارش نهایی طرح پژوهشی با عنوان:

مدلسازی QSPR برای پیش‌بینی دانسیته‌ی آلکانها با استفاده از روشهای خطی و غیر خطی

کد طرح: ۲۲۰۵۳

مجری: زهرا کلانتر

عضو هیئت علمی دانشکده شیمی

مدلسازی QSPR برای پیش بینی دانسیته‌ی آلکان‌ها با استفاده از روش‌های خطی و غیرخطی

چکیده:

در این کار، از مدلسازی QSPR با استفاده از روش خطی رگرسیون چند مرحله‌ای و روش غیر خطی شبکه‌ی عصبی مصنوعی، برای پیش‌بینی دانسیته‌ی آلکان‌ها در محدوده‌ی وسیعی از دما و فشار استفاده گردید. با توجه به این که آلکان‌های تا چهار کربنه (اجزاء گاز طبیعی) در شرایط محیط به صورت گاز و بقیه‌ی آلکان‌ها تا ۱۹ کربنه در همین شرایط مایع هستند لذا این مواد به دو گروه اجزاء گاز طبیعی و آلکان‌های ۵ تا ۱۹ کربنه تقسیم شدند. برای هر یک از این گروه‌ها، ۱۴۸۱ توصیف‌کننده توسط نرم افزار دراگون محاسبه گردید و با استفاده از روش رگرسیون چند مرحله‌ای، توصیف‌کننده‌های مناسب از میان آن‌ها انتخاب شد. توصیف‌کننده‌های انتخاب شده عبارتند از Ms و MW برای آلکان‌های ۵ تا ۱۹ کربنه و MATS2m و Mor و 04v برای اجزاء گازهای طبیعی. این توصیف‌کننده‌ها به همراه ۲ متغیر دما و فشار برای ساخت مدل بکار گرفته شدند. داده‌های موجود به صورت اتفاقی به سه دسته‌ی آموزش، ارزیابی و تست تقسیم شدند. پس از آموزش و بهینه کردن پارامترهای شبکه، کارایی مدل با استفاده از سری تست مورد ارزیابی قرار گرفت. متوسط مربعات خطا در پیش‌بینی دانسیته در سری تست، برای آلکان‌های ۵ تا ۱۹ کربنه ۰/۲۳۳ و برای اجزای گازهای طبیعی ۹/۷۰ بدست آمدند. همچنین نتایج به دست آمده از شبکه با نتایج حاصل از مدل خطی مقایسه گردید و مشخص شد که روش ANN توانایی بهتری در پیش‌بینی دانسیته‌ی این ترکیبات دارد.

کلمات کلیدی: دانسیته، آلکانهای خطی، اجزاء گاز طبیعی، شبکه‌ی عصبی مصنوعی، روش رگرسیون چند مرحله‌ای، مدلسازی QSPR

الف

فهرست مطالب

فصل اول: شبکه‌های عصبی مصنوعی

- | | |
|---|--|
| ۱ | 1-1- انگیزه‌های بیولوژیکی |
| ۲ | 2-1- تاریخچه |
| ۴ | 3-1- عنصر پردازش |
| ۵ | 4-1- ساختار لایه‌ای شبکه عصبی مصنوعی |
| ۶ | ۵-۱- انواع شبکه‌ها از نظر ارتباطات بین نرونی |
| ۷ | 6-1- آموزش شبکه عصبی |
| ۹ | 1-6-1- معرفی روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی |

فصل دوم: کمومتریکس

- | | |
|----|--|
| ۹ | مقدمه |
| ۹ | 1-2- تاریخچه |
| ۱۰ | ۲-۲- روش‌های پارامتری |
| ۱۰ | ۲-۲-۱- رگرسیون خطی چندگانه |
| ۱۱ | 2-1-2-2- انتخاب سری داده‌ها |
| ۱۱ | 2-1-2-2- محاسبه توصیف‌کننده‌ها |
| ۱۲ | 2-1-2-3- تجزیه، تحلیل و ارزیابی آماری توصیف‌کننده‌ها |
| ۱۲ | ۲-۲-۱-۴- آنالیز مدل‌های آماری و انتخاب مدل مناسب |

فصل سوم: مدل شبکه‌ی عصبی مصنوعی برای پیش‌بینی دانسیته‌ی آلکان‌های خطی در محدوده‌ی وسیعی از دما و فشار با استفاده از توصیف‌کننده‌های تئوری

۱۴	مقدمه
۱۴	۳-۱- انتخاب سری داده‌ها
۱۵	۳-۲- محاسبه توصیف‌کننده‌ها
۱۶	۳-۳- ارائه مدل خطی برای پیش‌بینی دانسیته n -آلکان با استفاده از روش رگرسیون خطی
۱۷	3-3-1- ارزیابی مدل خطی
۱۷	3-4- مدل‌سازی به روش غیر خطی با استفاده از شبکه عصبی
۱۸	3-4-1- انتخاب ورودی شبکه، انتخاب الگوریتم آموزشی، تعداد گره لایه پنهان و انتخاب تابع انتقال
۲۳	3-4-2- ارزیابی مدل غیر خطی

ب

فصل چهارم: مطالعه‌ی رابطه کمی ساختار – خاصیت (QSPR) برای پیش‌بینی دانسیته‌ی اجزاء گاز طبیعی مایع با استفاده از شبکه عصبی مصنوعی

۲۴	۴-۱- انتخاب سری داده‌ها
۲۴	۴-۲- انجام محاسبات و مدل‌سازی
۲۴	۴-۲-۱- رسم و بهینه‌سازی ساختار مولکول‌ها
۲۵	۴-۲-۲- محاسبه توصیف‌گرهای مولکولی
۲۵	۴-۲-۳- مدل رگرسیون خطی چندگانه برای پیش‌بینی دانسیته گازهای طبیعی
۲۶	۴-۲-۴- ارزیابی مدل خطی
۲۶	۴-۳- مدل شبکه عصبی مصنوعی و بهینه‌سازی آن

۲۶ ۴-۳-۱- انتخاب ورودی شبکه، انتخاب الگوریتم آموزشی، تعداد گره لایه پنهان و انتخاب تابع انتقال

۳۲ ۴-۳-۲- ارزیابی مدل غیر خطی

۳۳ مراجع

فهرست اشکال

- شکل (۱-۱): یک سلول عصبی بیولوژیک ۲
- شکل (۲-۱): ساختمان یک نرون محاسباتی ۴
- شکل (۳-۱): ساختار یک شبکه‌ی تک لایه پیشرو ۵
- شکل (۴-۱): ساختار یک شبکه سه لایه ۶
- شکل (۱-۳): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای آلکان‌های ۵ تا ۱۹ کربنه با استفاده از مدل خطی برای دو سری (الف) تایید و (ب) تست ۱۷
- شکل (۲-۳): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی بایزین (br) ۱۹
- شکل (۳-۳): نمودار مقدار میانگین مربعات خطا برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونبرگ-مارکوارت (lm) ۲۰
- شکل (۴-۳): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی لونبرگ-مارکوارت (lm) ۲۱
- شکل (۵-۳): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br) ۲۲
- شکل (۶-۳): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای آلکان‌های ۵ تا ۱۹ کربنه با استفاده از شبکه‌ی بهینه شده برای دو سری (الف) تایید و (ب) تست ۲۳
- شکل (۱-۴): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای اجزاء گاز طبیعی با استفاده از مدل خطی برای دو سری (الف) تایید و (ب) تست ۲۶
- شکل (۲-۴): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونبرگ-مارکوارت (lm) ۲۸

شکل (۳-۴): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل logsig و آلفوریتم آموزشی بایزین (br)

۲۹

شکل (۴-۴): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و آلفوریتم آموزشی لونبرگ-مارکوارت (lm)

۳۰

شکل (۵-۴): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و آلفوریتم آموزشی بایزین (br)

۳۱

شکل (۶-۴): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای اجزاء گاز طبیعی با استفاده از شبکه‌ی بهینه شده برای دو سری (الف) تایید و (ب) تست

۳۲

ت

فهرست جداول

جدول (۱-۳): محدوده‌ی دما و فشار برای برای آلکان‌های ۵ تا ۱۹ کربنی به همراه مرجع مورد استفاده برای هریک از آنها

۱۵

جدول (۲-۳): آماره‌های چند مدل طراحی شده با استفاده از رگرسیون خطی

۱۶

جدول (۳-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و آلفوریتم آموزشی بایزین (br)

۱۹

جدول (۴-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و آلفوریتم آموزشی لونبرگ-مارکوارت (lm)

۲۰

جدول (۵-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و آلفوریتم آموزشی لونبرگ-مارکوارت (lm)

۲۱

جدول (۳-۶): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br) ۲۲

جدول (۴-۱): محدوده‌ی دما و فشار برای اجزاء گاز طبیعی از C_2-C_4 به همراه مرجع مورد استفاده برای هریک از آنها ۲۴

جدول (۴-۲): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونبرگ-مارکوارت (lm) ۲۸

جدول (۴-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی بایزین (br) ۲۹

جدول (۴-۴): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی لونبرگ-مارکوارت (lm) ۳۰

جدول (۴-۵): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br) ۳۱

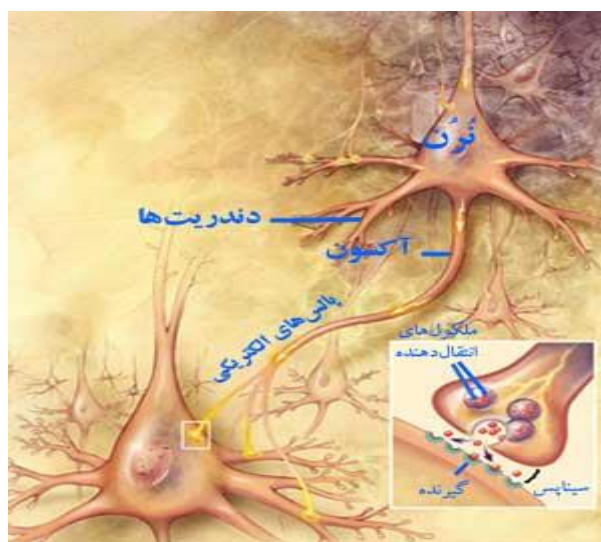
شبکه‌های عصبی مصنوعی

1-1- انگیزه‌های بیولوژیکی

پیاده‌سازی ویژگی‌های شگفانگیز مغز در یک سیستم مصنوعی (سیستم دینامیکی ساخته بشر) همیشه وسوسه انگیز و مطلوب بوده است. محققینی که طی ساها در این زمینه فعالیت کرده‌اند بسیارند. لیکن نتیجه این تلاش‌ها، صرف نظر از یافته‌های ارزشمند، باور هر چه بیشتر این اصل بوده است که مغز بشر دست نیافتنی است. تحقیقات و علاقه‌مندی به شبکه‌های عصبی از زمانی شروع شد که مغز به عنوان یک سیستم دینامیکی با ساختار موازی و پردازشگرهای کاملاً مغایر با پردازشگرهای متداول شناخته شد. نگرش نوین در مورد کارکرد مغز نتیجه تفکراتی بود که در اوایل قرن بیستم در مورد ساختار مغز به عنوان اجتماعی از اجزای محاسباتی کوچک به نام نرون شکل گرفت. در این بخش، هدف آن است که به طور مختصر آن دسته از ویژگی‌ها، کارکرد و ساختار مغز را که انگیزه‌ای جهت توسعه شبکه‌های عصبی مصنوعی به حساب می‌آیند، تشریح شوند [1].

مغز به عنوان یک سیستم پردازش اطلاعات با ساختار موازی، از 10^{11} نرون به هم مرتبط با تعداد 10^{16} ارتباط تشکیل شده است. نرون‌ها ساده‌ترین واحد ساختاری سیستم‌های عصبی هستند. بافت‌هایی که عصب نامیده می‌شوند، اجتماعی از نرون‌ها هستند که اطلاعات و پیام‌ها را از یک قسمت بدن به قسمت دیگر منتقل می‌کنند. میلیون‌ها نرون در بدن انسان وجود دارند. حتی ساده‌ترین کارهای روزمره انسان از قبیل پلک زدن تنها از همکاری همه جانبه این نرون‌ها میسر است. بیشترین تعداد نرون‌ها در مغز و بقیه در نخاع و سیستم‌های عصبی جانبی تمرکز یافته‌اند. گرچه همه نرون‌ها کارکرد یکسانی دارند، ولی اندازه و شکل آنها بستگی به محل استقرارشان در سیستم عصبی دارد. با وجود این همه تنوع، بیشتر نرون‌ها از سه قسمت اساسی بدنه سلول، دندریت و آکسون تشکیل شده‌اند که دو قسمت آخر عناصر ارتباطی نرون را تشکیل می‌دهند. شکل (۱-۱) ساختمان سلول عصبی را نشان می‌دهد.

دندریت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی، شبکه‌هایی تشکیل شده از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بی‌شمار می‌باشند، به همین علت آن‌ها را شبکه‌های دریافتی درخت گونه گویند. دندریت‌ها سیگنال‌های الکتریکی را به هسته سلول منتقل می‌کنند. بدنه سلول، انرژی لازم را برای فعالیت نرون فراهم کرده و بر روی سیگنال‌های دریافتی عمل می‌کند. آکسون برخلاف دندریت‌ها از سطحی هموارتر (تعداد شاخه‌های کمتر) برخوردار می‌باشد. آکسون طول بیشتری داشته و سیگنال الکتروشیمیایی دریافتی از هسته سلول را به نرون‌های دیگر منتقل می‌کند. محل تلاقی یک آکسون از یک سلول به دندریت‌های سلول دیگر را سیناپس می‌گویند. سیناپس‌ها واحد ساختاری کوچک تابعی هستند که ارتباط بین نرون‌ها را برقرار می‌سازند. سیناپس‌ها انواع مختلفی دارند که یکی از مهم‌ترین آنها، سیناپس‌های شیمیایی است.



شکل (۱-۱): یک سلول عصبی بیولوژیک

پیام‌های عصبی تنها به صورت یک طرفه حرکت می‌کنند، یعنی از دندریت‌ها به بدنه سلول و سپس به آکسون. سال‌ها محققان بیولوژی فکر می‌کردند که سیگنال‌های عصبی تنها از نوع الکتریکی می‌باشند اما بعدها دریافتند که این پالس‌های عصبی توسط یک فرآیند الکتروشیمیایی به انجام می‌رسند [2].

2-1- تاریخچه

از قرن نوزدهم به طور همزمان اما جداگانه نروفیزولوژیست‌ها سعی کردند سیستم یادگیری و تجزیه و تحلیل مغز را کشف کنند و از سویی دیگر ریاضیدانان تلاش کردند تا مدل ریاضی بسازند که قابلیت فراگیری و تجزیه و تحلیل عمومی مسائل را دارا باشد. اولین کوشش‌ها در شبیه‌سازی در سال ۱۹۴۳ با استفاده از یک مدل منطقی توسط مک کلوت^۱ و والتر پیتز^۲ انجام شد که امروزه بلوک اصلی سازنده‌ی اکثر شبکه‌های عصبی مصنوعی است. این مدل فرضیه‌هایی در مورد عملکرد نرون‌ها ارائه می‌کرد اما محدودیت‌های تکنولوژی در آن زمان اجازه کار بیشتر به آنها را نداد. نه تنها نروفیزولوژیست‌ها، بلکه روانشناسان و مهندسان نیز در پیشرفت شبیه‌سازی شبکه‌های عصبی تاثیر داشتند.

در سال ۱۹۵۸ شبکه پرسپترون^۳ توسط روزنبلات^۴ معرفی گردید. پرسپترون دارای سه لایه می‌باشد که لایه‌ی میانی به عنوان لایه‌ی پیوند شناخته می‌شود. سیستم دیگر مدل آدالاین می‌باشد که در سال ۱۹۶۰ توسط ویدرو^۵ و هاف^۶ به وجود آمد که اولین شبکه‌های عصبی به کار گرفته شده در مسائل واقعی بودند [3].

در سال ۱۹۶۹ مینسکی^۷ و پاپرت^۸ کتابی نوشتند که در آن محدودیت‌های سیستم تک لایه و چند لایه پرسپترون را تشریح می‌کردند. نتیجه‌ی این کتاب، پیش‌داوری و قطع سرمایه‌گذاری برای تحقیقات در زمینه شبیه‌سازی شبکه‌های عصبی بود. آنها با طرح این که پرسپترون قادر به حل هیچ مسئله‌ی جالبی نمی‌باشد تحقیقات در این زمینه را برای مدت چندین سال متوقف کردند [2].

1-Waren Mc Culloch
4-Frank Rosenblat
7-Minsky
10- Feedback

2-Walter Pitts
5-Bernurd Widrow
8-Papert
11- David Rummelhart

3-perceptron
6-Ted Hoff
9- John Hopfield
12- James Mcland

با وجودی که اشتیاق عمومی و سرمایه‌گذاری موجود به حداقل رسیده بود، برخی محققان تحقیقات خود را برای ساخت ماشین‌هایی که توانایی حل مسائلی از قبیل تشخیص الگو را داشته باشند ادامه دادند. پیشرفت‌های شبکه عصبی تا دهه‌ی هشتاد قرن بیستم ادامه یافت. با رشد تکنولوژی میکروپرسورها در دهه‌ی هشتاد، ایده‌ی به کارگیری شبکه‌ها در زمینه‌های جدیدی مطرح شد. در سال ۱۹۸۲ هاپفیلد^۹ با معرفی شبکه‌های چند لایه و الگوریتم یادگیری دارای بازخورد^{۱۱}، راه‌حلی را برای حل مسائل غیرخطی ارائه کرد و سپس الگوریتم پس انتشار خطا در سال ۱۹۸۶ توسط دیوید راملهارت^{۱۲} و جیمز مک‌کلند^{۱۳} مطرح گردید. از آن زمان تاکنون، این زمینه‌ی علمی روند رشد و پیشرفت را ادامه داده است، که البته بیشتر این پیشرفت‌ها به ساختارهای نوین و روش‌های یادگیری جدید مربوط می‌شود [۳].

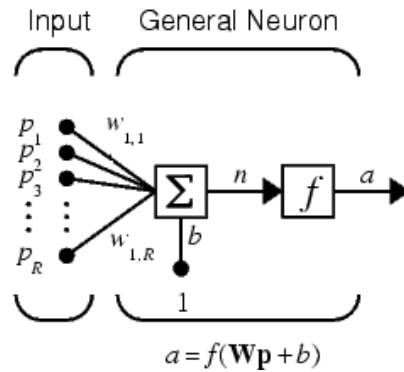
3-1- عنصر پردازش

مهم‌ترین جزء تشکیل دهنده‌ی هر شبکه عصبی، نرون می‌باشد. نرون که یک واحد محاسباتی - پردازشی است در حقیقت مدل ساده شده‌ای از نرون طبیعی است. نرون از تعدادی ورودی، یک خروجی، وزن‌های مربوط به ایجاد سیگنال تحریک و تابع تبدیل^۱ یا تابع تحریک تشکیل شده است. ارتباط بین خروجی و سیگنال تحریک در قسمتی به نام تابع تحریک به وجود می‌آید. بنابراین یک نرون از دو بخش تشکیل شده است: بخش اول وظیفه‌ی دریافت ورودی‌ها، وزن دهی آن‌ها و محاسبه‌ی مجموع آن‌ها به عنوان سیگنال تحریک را به عهده دارد و بخش دوم وظیفه‌ی اعمال تابع تحریک روی سیگنال تحریک و ساختن پیغام خروجی نرون به لایه‌ی بعد و یا خروجی شبکه را به عهده دارد. شکل (۱-۲) ساختمان یک نرون محاسباتی را نشان می‌دهد. بیان رابطه‌ی ورودی‌ها و خروجی‌های شبکه در یک مدل ریاضی مطابق فرمولاسیون شکل (۱-۲)، پرسپترون^۲ نامیده می‌شود.

^۱ - Transfer function

2- Perceptron

3- Bais



شکل (2-1): ساختمان یک نرون محاسباتی

اگر به شکل دقت کنید، علاوه بر تمام ورودی‌های وزن دار، ورودی دیگری نیز در نظر گرفته شده و آن عدد یک است. این عدد یک، در پیش قدر^۳ ضرب شده و با حاصلضرب ورودی‌ها در وزن‌ها جمع می‌شود تا ورودی خالص را ایجاد کند.

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

سپس تابع محرک روی این ورودی خالص n اثر می‌کند و خروجی نرون را ایجاد می‌کند.

$$(1-2)$$

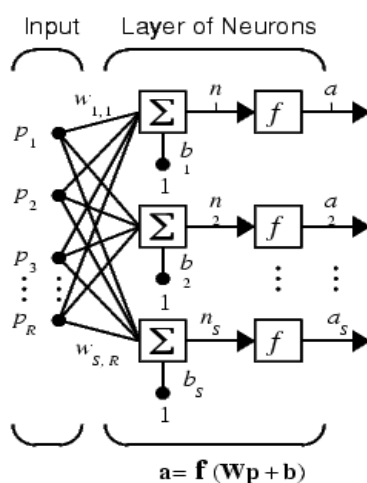
$$a = f(n) = f(\mathbf{Wp} + b)$$

این پیش قدر شبیه یک وزن است با این تفاوت که روی ورودی ثابت یک عمل می‌کند و مزیت آن افزودن یک متغیر اضافی به شبکه است. از این رو، شبکه‌ای که پیش قدر دارد قوی‌تر از شبکه‌ی فاقد پیش قدر می‌باشد. به علاوه، یک نرون بدون پیش قدر همیشه در اثر ورودی صفر، خروجی را صفر می‌دهد که این حالت مطلوب نیست و با اعمال پیش قدر برطرف می‌شود [4].

۱-۴- ساختار لایه‌ای شبکه عصبی مصنوعی

در این تقسیم‌بندی شبکه‌های عصبی مصنوعی، به دو صورت تک لایه و چند لایه تفکیک می‌شوند. در شبکه‌های تک لایه، نرون‌های ورودی و خروجی یکی هستند بنابراین به تعداد خروجی، نرون در شبکه وجود

دارد که همگی در یک لایه و به موازات هم کار می‌کنند. شکل (3-1) ساختار یک شبکه‌ی تک لایه را نشان می‌دهد.

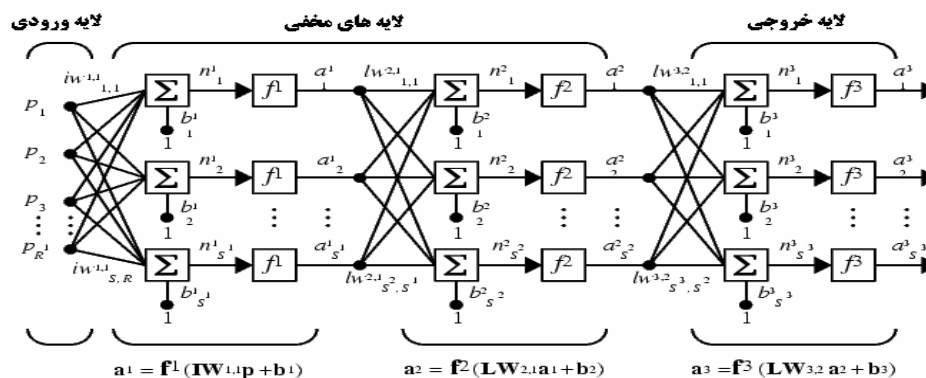


شکل (۳-۱): ساختار یک شبکه‌ی تک لایه پیشرو

گاهی دینامیک و قوانین شبکه‌های عصبی آنقدر پیچیده است که به پارامترهای قابل تنظیم بیشتر نیاز است، در این موارد از شبکه‌های چند لایه استفاده می‌شود. شبکه‌های چند لایه شامل لایه ورودی^۱، لایه خروجی^۲ و یک یا چند لایه پنهان^۳ یا لایه میانی می‌باشند. 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 لایه‌ی خروجی نام دارند و بقیه‌ی نرون‌ها که از دید پنهان هستند لایه‌های پنهان را تشکیل می‌دهند [5]. در یک شبکه‌ی چند لایه هر لایه، ماتریس وزن، بردار پیش قدر، بردار ورودی خالص و بردار خروجی مخصوص به خود را دارد، که برای مشخص کردن

1-Input layer
2-Output layer
3-Hidden layer

لایه‌های مختلف از بالانویس^۲ استفاده می‌شود. در شکل (۴-۱) یک شبکه سه لایه‌ای نشان داده شده است [4].



شکل (۴-۱): ساختار یک شبکه سه لایه

5-1- انواع شبکه‌ها از نظر ارتباطات بین نرونی

ارتباطات بین نرونی در یک شبکه عصبی از اهمیت بسیاری برخوردار می‌باشد و به نوعی قدرت شبکه را تعیین می‌کند. در این نوع تقسیم بندی دو نوع شبکه وجود دارد.

- ۱- شبکه‌های پیش‌خور^۳
- ۲- شبکه‌های پس‌خور^۴

در شبکه‌های پیش‌خور، اتصالات بین نرونی به گونه‌ای است که هر نرون لایه میانی، می‌تواند ورودی خود را از هر یک از نرون‌های لایه‌های قبلی خود دریافت کند اما هیچ سیگنالی از نرون به خودش یا به نرون لایه‌ی قبل برنمی‌گردد. بدین ترتیب در یک شبکه عصبی، سیگنال‌ها به تدریج از یک لایه نرونی به لایه‌های بالاتر حرکت می‌کنند و در نهایت به لایه آخر یا خروجی شبکه می‌رسند.

^۲- Superscript

^۳- Feed forward

^۴- Feed Back

نوع دیگری از ارتباط بین نرونی در شبکه‌های عصبی ارتباط بازخوردی است که در شبکه‌های پس‌خور وجود دارد. در این نوع از ارتباطات، خروجی یک نرون به همان نرون یا نرون‌های همان لایه یا به نرون‌های لایه‌ی قبلی اتصال داده می‌شود. ارتباطات بین نرونی در شبکه‌های عصبی از اهمیت بسیار زیادی برخوردار هستند و به نوعی قدرت یک شبکه عصبی را تعیین می‌کنند [۵].

1-6- آموزش شبکه عصبی

شبکه‌های عصبی می‌توانند بر اساس طراحی خود، سیگنال‌های ورودی را پردازش کنند و به سیگنال‌های خروجی مورد نظر تبدیل نمایند. به طور معمول پس از آنکه شبکه‌های عصبی طراحی و پیاده‌سازی شدند، باید پارامترهای پیش‌قدر و وزن‌ها (W و b) به ازای مجموعه‌هایی از سیگنال‌های ورودی، به گونه‌ای تنظیم شوند که سیگنال‌های خروجی شبکه خروجی مطلوب را تشکیل دهند. چنین فرایندی را آموزش شبکه عصبی می‌نامند. به طور کلی برای آموزش شبکه دو روش وجود دارد:

۱- یادگیری با ناظر¹ ۲- یادگیری بدون ناظر²

روش یادگیری با ناظر، شامل مراحل است که در آن سیستم از طریق در دست داشتن جواب واقعی آموزش می‌بیند. اما در روش بدون ناظر، شبکه عصبی باید خود سامان ده باشد و بدون در دست داشتن جواب واقعی، بتواند کار آموزش را انجام دهد. در این نوع یادگیری، شبکه می‌آموزد که الگوهای ورودی را به تعداد متنهای از گروه‌ها تقسیم‌بندی کند و عبارتی عمل خوشه‌بندی را انجام می‌دهد. باید توجه داشت در این حالت فرد طراح است که مقصد و هدف نهایی یعنی چیزی را که باید به آن رسید مشخص می‌کند. رفتار سیستم‌های یادگیر توسط الگوریتم‌های بازگشتی بیان می‌شود که به آنها قوانین یادگیری می‌گویند. انواع مختلفی از قوانین یادگیری وجود دارد که یادگیری عملکردی³ یکی از آنهاست. قانون یادگیری در اینجا روندی است که توسط آن ماتریس وزن‌ها و بردارهای پیش قدر شبکه عصبی به نحوی

1-Supervised learning
2- Unsupervised learning
3- Performance learning

تنظیم می‌شوند که عملکرد شبکه بهینه شود. منظور از بهینه شدن عملکرد شبکه، حداقل شدن خطایی است که بین مقادیر تجربی و پاسخ شبکه وجود دارد. برای بهینه‌سازی عملکرد، شبکه باید پارامترهایش را به گونه‌ای تنظیم کند که شاخص عملکرد^۵ کمترین مقدار را داشته باشد. شاخص عملکرد، معیاری است که به طور کمی، عملکرد شبکه را بیان می‌کند و رابطه معکوسی با نحوه‌ی عملکرد شبکه دارد، یعنی هر چه مقدار این شاخص کوچکتر باشد، عملکرد شبکه بهتر خواهد بود و بالعکس. به شاخص عملکرد اصطلاحاً تابع هدف یا تابع خطا می‌گویند و در بیشتر الگوریتم‌ها، متوسط مربعات خطا^۶ (MSE) می‌باشد [۴].

1-6-1- معرفی روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی

در تمام الگوریتم‌های بهینه‌سازی یک مقدار اولیه برای تابع حدس زده می‌شود و بعد این مقدار اولیه طی مراحل‌ی که از قاعده‌ی زیر پیروی می‌کند بهینه می‌شود [6].

$$X_{K+1} = X_K + \alpha P_K \quad (3-1)$$

که در آن X_K تخمین نقطه مینیمم تابع در مرحله K ام است و X_{K+1} حدس مرحله‌ی بعدی نقطه مینیمم تابع می‌باشد. P_K نیز بردار جستجو می‌باشد، بدین معنی که از روی آن، تخمین مرحله‌ی بعد به دست می‌آید. α یک مقدار اسکالر بین صفر و یک بوده و سرعت یادگیری^۷ نامیده می‌شود. سرعت یادگیری، طول قدم در هر تکرار را در هر مسیر P_K تعیین می‌کند. انتخاب P_K و α روش‌های مختلف مینیمم سازی را از هم متمایز می‌کند که بعضی از آنها عبارتند از:

۱- روش نزول گرادیانی^۸

۲- الگوریتم پس انتشار خطا^۹

^۵- Performance

^۶- Mean square error

^۷- Learning rate

^۸- Gradient descent algorithm

^۹- Error back propagation algorithm

۳- روش لونیبرگ - مارکوارت^{۱۰}

۴- الگوریتم تنظیم بایزین^{۱۱}

فصل دوم: کمومتریکس

مقدمه

برای درک مکانیسم فرآیندهای مختلف شیمیایی، کشف و توسعه مواد جدید، حفظ محیط زیست و زمینه های دیگر شیمی، هنوز توانایی حل مسائل به طور کامل وجود ندارد و برای عملی کردن بعضی از مسائل، نیاز به سیستم های بسیار پیچیده ای می باشد که انجام آنها در گرو صرف هزینه های بسیار و مطالعات گسترده است. در جهت حل این مشکل، روشهای محاسباتی کمومتریکس می توانند مفید باشند. تجزیه و تحلیل آماری و ریاضی داده های شیمیایی معمولاً تحت عنوان کمومتریکس یاد می شود [7].

1-2- تاریخچه

کمومتریکس اولین بار توسط دانشمند سوئدی به نام ولد^{۱۲} در سال 1972 به کار گرفته شد. توسط کووالسکی^{۱۳} توسعه داده شد [8]. در سال ۱۹۷۴ انجمن بین المللی کمومتریکس تأسیس گردید و از سال 1980 دانش کمومتریکس خیلی سریع توسعه یافت [9]. کمومتریکس به عنوان یک شاخه علمی جوان در دو دهه اخیر به سرعت توسعه پیدا کرده است. این رشد سریع مدیون پیشرفت سریع دستگاه های هوشمند و خودکار آزمایشگاهی و همچنین امکان استفاده از کامپیوترهای قدرتمند و نرم افزارهای ساده است.

^{۱۰} - Levenberg-Marquardt

^{۱۱} - Bayesian regularization

^{۱۲} - Wold

^{۱۳} - B. R. Kowalski

کمومتریکس در شاخه‌های مختلف شیمی مورد استفاده قرار می‌گیرد که برخی از این کاربردها شامل کنترل فرایندها، تجزیه و تحلیل و شناخت الگوها، پردازش علائم و بهینه کردن شرایط می باشد یکی از زمینه‌های مهم کاربرد کمومتریکس در مطالعاتی است که خواص مولکول‌ها را به ویژگی‌های ساختاری آن‌ها نسبت می‌دهد. از نظر شیمی‌دانان فعالیت و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است [10]. این نوع از مطالعات به بررسی کمی ارتباط ساختمان با فعالیت^{۱۴} و همچنین بررسی کمی ارتباط ساختمان با ویژگی^{۱۵}، معروف می‌باشد. امروزه در کمومتریکس از روش‌های مختلف ریاضی و آماری مانند رگرسیون خطی چندگانه^{۱۶} حداقل مربعات جزئی^{۱۷}، آنالیز اجزاء اصلی^{۱۸} استفاده شده است. همچنین در پیشرفتهای دو دهه اخیر از شبکه‌های عصبی مصنوعی، منطق فازی و الگوریتم ژنتیکی نیز استفاده شده است.

پس به طور خلاصه می‌توان گفت کمومتریکس علم استفاده از ریاضی، آمار و کامپیوتر در شیمی می باشد. حال پس از دانستن تعریف و تاریخچه کمومتریکس به شرح یکی از پرکاربردترین زمینه‌های کاربرد آن یعنی مطالعات کمی ارتباط ساختمان با فعالیت و بررسی کمی ارتباط ساختمان با خاصیت می‌پردازیم. به مجموعه ابزارها و روش‌هایی که به این منظور مورد استفاده قرار می‌گیرند روش‌های پارامتری گویند که در اینجا به شرح این روش‌ها می‌پردازیم.

۲-۲- روش‌های پارامتری

روش‌های پارامتری، مجموعه تکنیک‌هایی هستند که برای مدل‌سازی و پیش‌بینی خاصیت‌های ملکولی مورد استفاده قرار می‌گیرند. با استفاده از این روش‌ها می‌توان ارتباط منطقی بین فعالیت یا خاصیت موردنظر با

۱۴- Quantitative structure–activity relationship

۱۵- Quantitative structure–activity relationship

۱۶- Multiple linear Regression

۱۷- Partial Least Square

۱۸- Principle Component Analysis

تعدادی از توصیف‌کننده‌های ملکولی برقرار نمود. توصیف‌کننده‌های ملکولی، متغیرهای مستقلی هستند که ویژگی‌های ملکول را به طور کمی بیان می‌کنند. معمولاً از روش رگرسیون خطی چندگانه جهت ایجاد این روابط استفاده می‌شود.

۲-۲-۱- رگرسیون خطی چندگانه

در رگرسیون خطی چندگانه، ارتباط بین توصیف‌کننده‌ها به عنوان متغیر مستقل و خاصیت به صورت یک متغیر وابسته در قالب یک معادله ریاضی طرح‌ریزی می‌شود. روش‌های آماری برای ایجاد ارتباط بین متغیرهای وابسته و مستقل کاربرد وسیعی دارند و نرم افزارهای آماری متفاوتی جهت تسریع و سهولت انجام عملیات محاسباتی مورد استفاده قرار می‌گیرند. بعد از معرفی متغیرهای مستقل و وابسته، براساس روش‌های رگرسیون خطی چندگانه، معادلات خطی ارتباط دهنده بین این دو متغیر ایجاد می‌گردد [7]. قابل ذکر است که متغیرهای مستقل براساس اصول علمی که انتظار می‌رود با متغیر وابسته در ارتباط باشند، انتخاب می‌شوند. مراحل کلی مدل سازی به روش پارامتری به ترتیب زیر می‌باشد:

۱- انتخاب سری داده ها

۲- انتخاب و محاسبه توصیف کننده ها

۳- تجزیه و تحلیل آماری توصیف کننده ها

۴- آنالیز مدل های آماری و انتخاب مدل مناسب

۵- ارزیابی اعتبار مدل های انتخاب شده

برای آشنایی بیشتر توضیحاتی در مورد هر یک از مراحل فوق ارائه می‌گردد.

2-2-1- انتخاب سری داده ها

اولین مرحله ، جمع آوری و انتخاب یک سری ملکولی است که مقادیر تجربی خاصیت موردنظر آنها در دسترس باشد. سپس سری داده ها به سه قسمت سری آموزشی و سری آزمون و سری ارزیابی تقسیم می‌شوند. عملیات مدل سازی بر روی سری آموزشی که اکثر ملکول ها را در بر می گیرد، انجام می‌شود. سری آزمون و سری ارزیابی به نحوی انتخاب می‌شوند که نماینده جمعیت ملکول های سری آموزشی باشند. لازم به ذکر

است که ملکول‌های سری ارزیابی در هیچ یک از مراحل مدل‌سازی شرکت ندارند. در واقع سری ارزیابی برای ارزیابی اعتبار مدل‌های حاصله است. هر قدر که سری داده‌ها بزرگتر و متنوع‌تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش بینی بالاتری خواهد داشت.

2-1-2-2- محاسبه توصیف کننده‌ها

توصیف کننده‌ها پارامترهای عددی هستند که ویژگی‌های متفاوتی از ملکول را بیان می کنند. به عبارت دیگر، هر یک از این توصیف کننده‌ها، اطلاعات خاصی از ملکول را در اختیار می گذارند [11]. تعیین و انتخاب توصیف کننده‌ها از اهمیت به سزایی برخوردار است. در صورتی که توصیف کننده‌های انتخاب شده مناسب نباشند، عمل مدل‌سازی با مشکل مواجه شده و مدل مناسب به دست نخواهد آمد. توصیف کننده‌ها از نظر بیان چگونگی خصوصیات ملکول به دسته‌های متفاوتی تقسیم می شوند.

2-1-2-3- تجزیه، تحلیل و ارزیابی آماری توصیف کننده‌ها

تا این مرحله انتخاب توصیف کننده‌ها براساس احتمالات و تشخیص نظری استوار است. بنابراین احتمال اینکه برخی از توصیف کننده‌ها برای مدل‌سازی مناسب نباشند وجود دارد و باید آنها را حذف کرد [12]. توصیف کننده‌هایی که حذف می شوند معمولاً یک یا چند ویژگی زیر را دارند:

- ۱- توصیف کننده‌هایی که کمتر از ۱۰٪ مقادیر غیرصفر دارند یا دارای بیش از ۹۰٪ مقادیر یکسان باشند.
- ۲- توصیف کننده‌هایی که با توصیف کننده دیگر همبستگی بالایی دارند ($r > 0.9$)
- ۳- توصیف کننده‌هایی که با متغیر وابسته، همبستگی کمی دارند و محاسبه آنها مشکل است.

با توجه به موارد فوق، تعدادی از توصیف کننده‌ها حذف می شوند و از پیچیدگی محاسبات جلوگیری می‌شود.

۲-۱-۲-۲- آنالیز مدل های آماری و انتخاب مدل مناسب

در این مرحله می توان با استفاده از روشهای آماری مختلف به جستجوی مدل پرداخت. مدل، در واقع یک رابطه ریاضی است که بیان کننده رابطه بین متغیر وابسته و مستقل می باشد و به کمک آن می‌توان با داشتن مقادیر متغیرهای مستقل، متغیر وابسته را ارزیابی کرد. برای این کار چند روش رگرسیون متفاوت وجود دارد که در ادامه به اختصار به آنها پرداخته می‌شود. [7]

الف) روش ورود اجباری:

در این روش اجباراً همه متغیرهای مستقل وارد مدل می‌شوند. در این روش اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می‌شود. بنابراین روش مطلوبی نیست.

ب) روش انتخاب جلوبرنده:

در این روش ابتدا متغیری که بیشترین همبستگی را با متغیر وابسته دارد وارد مدل می‌شود و ابتدا یک خط یک متغیره ایجاد می‌شود. با انجام آزمون F روی مدل و معنی دار بودن آزمون، آن متغیر وارد شده در مدل باقی می‌ماند. بقیه متغیرهایی که بزرگترین آماره F را داشته و معنی دار باشند وارد مدل می‌شوند. این کار تا زمانی که دیگر آماره F حاصل از متغیرهای خارج از مدل معنی دار نباشند، تکرار می‌شود و در نهایت، مدل نهایی بدست می‌آید.

ج) روش حذف عقب برنده:

در این روش ابتدا یک آماره F برای هر متغیر موجود در مدل تعیین می‌شود. با توجه به اثر متغیر بر روی F کلی و ضریب خطی بودن مدل، متغیر حذف یا نگه داشته می‌شود. حذف شدن متغیرها تا زمانی که دیگر متغیری قابل حذف نباشد تکرار می‌شود و نهایتاً مدل باقی مانده به عنوان مدل مناسب انتخاب می‌شود.

د) روش رگرسیون چند مرحله‌ای:

این روش مشابه روش انتخاب جلوبرنده است، با این تفاوت که بر روی متغیرهای وارد شده به مدل در مرحله قبل، آزمون مجدد انجام می‌گیرد. در این روش در هر مرحله برای متغیر وارد شده در مدل، آماره F محاسبه می‌شود و متغیر با کوچکترین آماره F بدون معنی حذف می‌شود و مدل مناسب بدست می‌آید.

۲-۱-۵-۲-۲ ارزیابی مدل بدست آمده

برای اطمینان از اینکه مدل بدست آمده، مدلی مناسب می‌باشد باید یک سری پارامترها را در نظر

بگیریم، به طور کلی می‌توان گفت مدلی مناسب‌تر است که دارای ویژگی‌های زیر باشد [13]:

- تعداد متغیرهای مستقل آن از سایر مدل‌های دیگر کمتر باشد.

- محاسبه متغیرهای مستقل آن ساده‌تر از سایر مدل‌ها باشد.

- حتی الامکان توصیف کننده هایی که مقدار آنها باید به روش تجربی اندازه گیری شود در مدل وارد نشده باشد.
- ضریب همبستگی بین توصیف کننده هایی که در مدل شرکت دارد کم باشد.
- خطای استاندارد آن کم باشد.
- آماره F بزرگ بوده و ضریب همبستگی آن به یک نزدیک باشد.
- قدرت پیش بینی بالایی داشته باشد.

فصل سوم: مدل شبکه‌ی عصبی مصنوعی برای

پیش‌بینی دانسیته‌ی آلکان‌های خطی در

محدوده‌ی وسیعی از دما و فشار با استفاده از

توصیف کننده‌های تئوری

مقدمه

همانطور که در فصل قبل اشاره شد شبکه‌های عصبی مصنوعی یکی از روش‌های نوین پردازش اطلاعات می‌باشند که قدرت یادگیری و محاسباتی بالایی را دارا می‌باشند و در علوم مختلف استفاده می‌شوند. به همین دلیل در این کار از شبکه‌ی عصبی مصنوعی استفاده گردید. آنچه در ادامه می‌آید چگونگی استفاده از شبکه‌ی عصبی مصنوعی برای پیش‌بینی دانسیته‌ی آلکان‌های خطی (C_5-C_{19}) در محدوده وسیعی از دما و فشار با استفاده از توصیف کننده‌های تئوری می‌باشد.

۳-۱- انتخاب سری داده‌ها

در این کار چون هدف پیش‌بینی دانسیته‌ی آلکان‌ها از ۵ کربنی تا ۱۹ کربنی می‌باشد، لذا داده‌های دانسیته برای این ترکیبات در محدوده‌ای وسیعی از دما و فشار جمع‌آوری گردید. نام ترکیبات، به همراه محدوده‌ی دما و فشار و مرجع مورد استفاده برای هر ترکیب در جدول (۱-۳) آمده است. لازم به تذکر است که این داده‌ها به صورت اتفاقی به سه سری آموزش، تائید و تست تقسیم شدند که در سری آموزش ۴۲۷، سری تایید ۱۲۲ و سری تست ۱۰۹ داده قرار گرفتند.

جدول (۱-۳): محدوده‌ی دما و فشار برای آلکان‌های ۵ تا ۱۹ کربنی به همراه مرجع مورد استفاده برای هریک از آنها

مرجع	Δp (MPa)	ΔT (K)	نام ترکیب
[۱۴]	۵-۱۷۶	۳۰۳-۳۲۳	پنتان
[۱۴] و [۱۵]	۰/۱-۲۵۰	۲۹۸/۱۵-۳۴۸/۱۵	هگزان
[۱۶]	۲/۱-۳۸/۶	۲۹۸/۱۵-۳۲۸/۱۵	هپتان
[۱۴]	۰/۱-۲۴۷/۴۱	۳۰۳/۱۵-۳۴۸/۱۵	اکتان
[۱۷]	۲۰-۲۶۰	۳۰۳-۴۲۳	نونان
[۱۴]	۰/۱-۲۵۴/۳۷	۳۰۳-۳۴۸	دکان
[۱۸]	۰/۱-۱۴۰	۲۹۳-۴۳۳	دودکان
[۱۹]	۲۰-۱۴۹/۵۵	۲۹۳-۳۸۳	پنتادکان
[۱۹]	۲۰-۱۴۹/۵۵	۳۲۳-۳۸۳	هپتادکان
[۲۰]	۲۰-۱۴۹/۵۵	۳۱۳-۳۸۳	اکتادکان
[۲۰]	۲۰-۱۴۹/۵۵	۳۱۳-۳۸۳	نونادکان

۳-۲- محاسبه‌ی توصیف‌کننده‌ها

توصیف‌کننده‌ها مقادیر عددی هستند که بیانگر ویژگی خاصی از مولکول می‌باشند و انتخاب توصیف‌کننده مناسب برای برقراری ارتباط بین ساختار مولکولی و دانسیته، در ایجاد مدل اهمیت بسیاری دارد. از آنجایی که دانسیته‌ی هر ترکیب به دما و فشار بستگی دارد لذا این دو ویژگی به عنوان اولین توصیف‌کننده‌ها انتخاب شدند. برای انتخاب سایر توصیف‌کننده‌ها با توجه به روش مورد بررسی مطابق زیر عمل گردید:

- ۱- ابتدا ساختار مولکولها توسط نرم افزار Hyperchem به صورت تقریبی رسم شد و سپس با کمک روش نیمه تجربی AM1 ساختار هر ترکیب بهینه گردید.
- ۲- محاسبه توصیف کننده ها توسط نرم افزار دراگون صورت می گیرد. برای این منظور ساختارهای بهینه رسم شده در Hyperchem به عنوان ورودی برای نرم افزار دراگون استفاده شدند. توسط این نرم افزار ۱۸ گروه مختلف جمعاً ۱۴۸۱ توصیف کننده برای هر ترکیب بدست می آید.
- ۳- با توجه به تعداد زیاد توصیف کننده ها احتمال برهم کنش ها آن ها با هم بالا می رود و همچنین حجم محاسبات نیز افزایش می یابد لذا تعدادی از این توصیف کننده ها که ۹۰٪ مقادیر یکسان را دارند، حذف شدند.
- ۴- در مرحله بعد با استفاده از نرم افزار SPSS ضریب همبستگی میان توصیف کننده ها محاسبه گردید و توصیف کننده هایی که دارای ضریب همبستگی بیشتر از ۰/۹ با هم بودند، مشخص شده و یکی از آنها حذف شد.

۳-۳- ارائه مدل خطی برای پیش بینی دانسیته ی آلکان های خطی با

استفاده از روش رگرسیون خطی

پس از انتخاب توصیف کننده های مناسب، مدل رگرسیون خطی چندگانه برای پیش بینی دانسیته ی آلکان ها طراحی شد. به این صورت که توصیف کننده های انتخاب شده در قسمت قبل به علاوه دما و فشار به عنوان متغیر مستقل و دانسیته به عنوان متغیر وابسته به نرم افزار SPSS داده شدند و سپس با استفاده از روش آنالیز و انتخاب رگرسیون مرحله ای، چندین مدل بدست آمد که بهترین مدل با توجه به خصوصیات آماری، مدلی است که دارای بیشترین مقدار R ، F و کمترین مقدار خطای استاندارد باشد. جدول (۳-۲) آماره های چند مدل طراحی شده با استفاده از رگرسیون خطی را نشان می دهد.

جدول (۳-۲): آماره های چند مدل طراحی شده با استفاده از رگرسیون خطی

Model	F	R ²
Constant+p	422/029	0/391
Constant+p+MW	1682/735	0/837
Constant+p+MW+T	2226/021	0/911
Constant+p+MW+T+Ms	2789/724	0/945

با استفاده از اطلاعات این جدول مدل چهارم به عنوان بهترین مدل انتخاب گردید. سپس این توصیف‌کننده‌ها در طراحی و ساخت مدل خطی برای داده‌های سری آموزش به کار برده شدند. با انتخاب رگرسیون خطی و روش اجباری ضرایب توصیف‌کننده‌ها در مدل استخراج گردید و معادله‌ی (۱-۳) برای پیش‌بینی دانسیته‌ی آلکان‌ها بدست آمد.

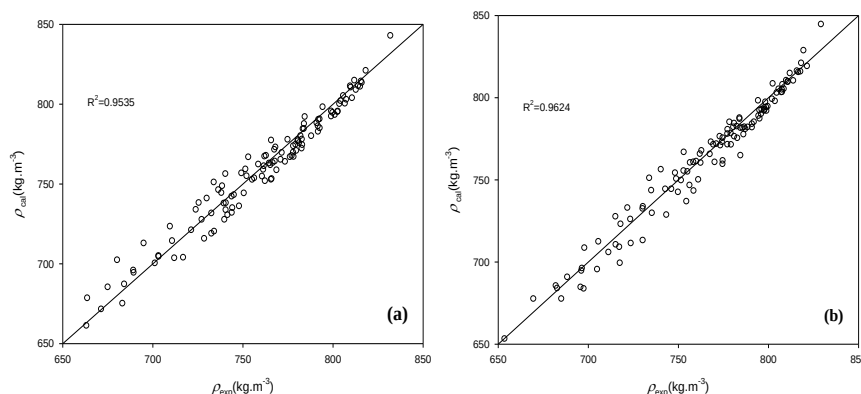
$$D = 2022.052 - 727.245 Ms + 0.199 MW + 0.418 P - 0.526 T$$

که در آن T دما، p فشار، MW وزن ملکولی و Ms توصیف‌کننده‌ی زیر ساختاری می‌باشد.

1-3-3- ارزیابی مدل خطی

برای ارزیابی مدل رگرسیون خطی بدست آمده برای پیش‌بینی دانسیته، داده‌های سری تایید و تست در مدل خطی بدست آمده قرار داده شد و مقادیر MSE و متوسط درصد خطا برای این دو سری محاسبه گردید که مقادیر MSE و متوسط درصد خطا به ترتیب برای سری تایید ۸۵/۵۸ و ۰/۹۴۵ و برای سری تست ۷۰/۰۶ و ۰/۸۱۶ بدست آمد.

نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده نیز برای این مدل برای دو سری تایید و تست در شکل (۱-۳) رسم شده است. این شکل نشان می‌دهد که یک ارتباط غیر خطی بین توصیف‌کننده‌ها و دانسیته وجود دارد.



شکل (۳-۱): نمودار مقادیر تجربی بر حسب مقادیر پیش‌بینی شده برای آلکان‌های ۵ تا ۱۹ کربنه با استفاده از مدل خطی برای دو سری (الف) تایید و (ب) تست

3-4- مدل سازی به روش غیر خطی با استفاده از شبکه‌ی عصبی

در این بررسی مدل شبکه‌ی عصبی مصنوعی با استفاده از نرم‌افزار MATLAB برای پیش‌بینی دانسیته‌ی آلکان‌ها طراحی گردید. برای این منظور مدل شبکه‌ی عصبی مصنوعی با استفاده از ۴ توصیف‌کننده مورد استفاده در مدل MLR طراحی گردید. در این شبکه به ازای هر توصیف‌کننده یک ورودی وجود دارد. لایه‌ی خروجی نیز شامل یک نرون می‌باشد که دانسیته‌ی متناظر با ورودی‌ها را نشان می‌دهد. تعداد نرون‌های لایه مخفی نامعلوم است و جزء پارامترهایی است که باید بهینه شود. علاوه بر تعداد نرون لایه مخفی، تعداد چرخه‌های آموزش، سرعت آموزش، ممنت، نوع تابع انتقال و نوع الگوریتم آموزشی نیز باید بهینه شوند. مقدار بهینه برای هر یک از این کمیتها زمانی است که میانگین مربع خطا برای سری تایید مینیمم شود.

3-4-1- انتخاب ورودی شبکه، الگوریتم آموزشی، تعداد گره لایه پنهان و تابع انتقال

در مدل سازی شبکه‌ی عصبی مصنوعی، یک شبکه به روش پیش‌خور برگشتی با استفاده از توصیف‌کننده‌های مدل منتخب MLR به عنوان ورودی طراحی می‌شود. در شروع برنامه، مقدار وزنه‌های ورودی و خروجی به صورت تصادفی از اعداد بین صفر و یک انتخاب می‌شوند و بعد این مقادیر حین آموزش بهینه می‌شوند. برای آموزش شبکه نیز دو الگوریتم آموزشی بایزین و لوبنبرگ-مارکوارت مورد بررسی قرار گرفتند. همچنین برای انتخاب تابع انتقال مناسب دو تابع انتقال تانژانت سیگموئید (tansig) و لگاریتم سیگموئید (logsig)

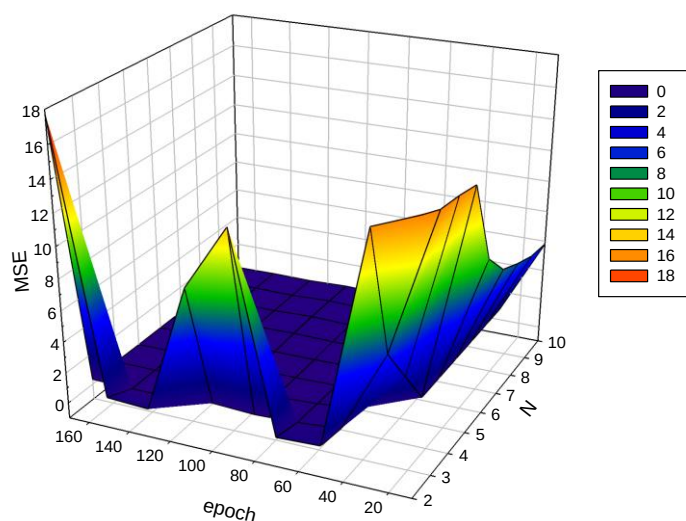
مورد استفاده قرار گرفتند. به این ترتیب با استفاده از دو الگوریتم آموزشی و دو تابع انتقال انتخاب شده چهار شبکه‌ی عصبی مصنوعی مختلف طراحی شدند. سپس برای هر یک از این شبکه‌ها، از روش تغییر همزمان برای تغییر پارامترهای شبکه (تعداد نرونها، لایه‌ی مخفی و تعداد دور آموزش) جهت آموزش شبکه استفاده گردید و به ازای همه‌ی ترکیبهای ممکن از این دو پارامتر شبکه‌های طراحی شده بهینه‌سازی شدند. نتایج حاصل از این بررسی‌ها برای سری تایید در جداول (۳-۳) تا (۶-۳) و نمودارهای (۳-۲) تا (۵-۳) نشان داده شده است.

با توجه به نتایج بدست آمده، شبکه با الگوریتم آموزشی بایزین و تابع انتقال لگاریتم سیگموئید با تعداد گره ۷ در لایه‌ی پنهان و تعداد دور آموزشی ۹۰ دارای کمترین مقدار خطا می‌باشد. از آنجا که سرعت آموزش با این الگوریتم بالاست، با تغییر مقدار سرعت آموزش تغییری در خطای شبکه ایجاد نمی‌شود. بنابراین نیازی به بهینه کردن سرعت آموزش نیست.

جدول (۳-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی بایزین (br)

تعداد دور	تعداد گره								
	2	3	4	5	6	7	8	9	10
10	38/5064	39/433 9	0/986 7	0/880 9	13/03 8	8/9930	0/831 7	0/973 0	17/72 8
30	34/8567	5/9917	0/711 2	0/480 5	0/380 6	0/5160	0/277 6	0/257 9	0/319 7
50	34//670 4	3/3388	0/704 1	0/386 0	0/243 6	0/2940	0/269 5	0/266 2	0/262 2
70	34/3592	0/7208	0/553 1	0/303 5	0/253 3	0/2555	0/271 1	0/270 5	0/248 1

90	8/9644	0/7211	0/542 7	0/290 8	0/244 8	0/233 0	0/269 5	0/272 0	0/256 8
11 0	7/2540	0/7211	0/542 4	0/287 6	0/284 0	0/3282	0/270 2	0/272 5	0/254 5
13 0	6/5683	0/7211	0/543 4	0/284 7	0/278 4	0/3616	0/257 5	0/272 9	0/248 9
15 0	6/0073	0/7211	0/543 8	0/281 5	0/280 8	0/3624	0/252 3	0/273 5	0/240 3
17 0	5/7978	0/7211	0/543 9	0/278 0	0/281 7	0/3421	0/251 7	0/274 7	0/238 9
19 0	5/5435	0/7211	0/543 9	0/274 7	0/283 2	0/3352	0/253 8	0/277 1	0/235 2



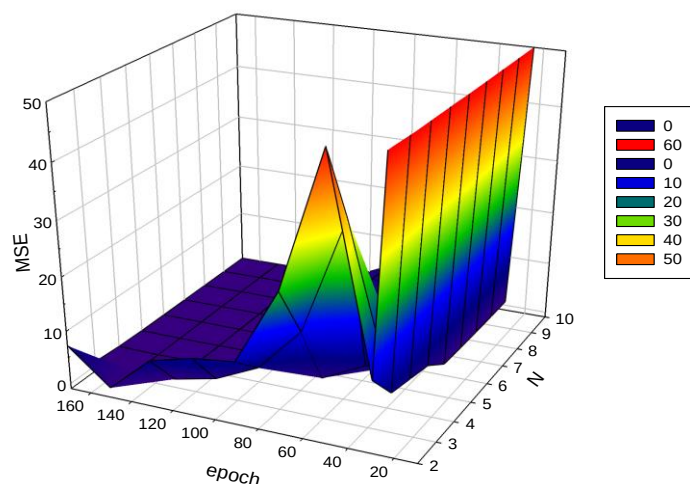
شکل (۳-۲): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل

logsig و الگوریتم آموزشی بایزین (br)

جدول (۳-۴): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونیگ-مارکوارت (lm)

	تعداد گره
--	-----------

تعداد دور	2	3	4	5	6	7	8	9	10
10	182/891 9	12/188 0	49/276 6	24/548 9	10/642 9	9/632 5	7/601 8	0/8206	7/251 6
30	182/891 2	6/3427	35/301 5	14/228 1	6/0480	2/370 5	0/615 2	0/2941	0/757 4
50	177/765 0	5/3632	6/2325	2/2328	5/4785	1/359 6	0/469 3	0/2853	0/421 5
70	177/787 6	4/3767	2/2460	2/0567	4/8638	0/393 5	0/301 5	0/242 1	0/292 6
90	177/774 5	1/2048	1/8410	1/8809	4/6826	0/335 5	0/301 7	0/3011	0/284 1
110	177/816 4	0/9987	1/8360	1/8352	4/4560	0/325 1	0/315 5	0/3262	02725
130	177/871 7	0/8514	1/8253	1/7930	4/3695	0/321 5	0/338 8	0/3416	0/269 5
150	177/782 2	0/7572	1/8224	1/7816	4/2801	0/319 2	0/343 4	0/3421	0/269 1
170	177/785 0	0/7316	1/8220	1/7710	3/9858	0/309 3	0/341 6	0/3253	0/270 0
190	177/785 3	0/7178	1/8200	1/7615	3/4899	0/282 8	0/482 2	0/3180	0/271 6

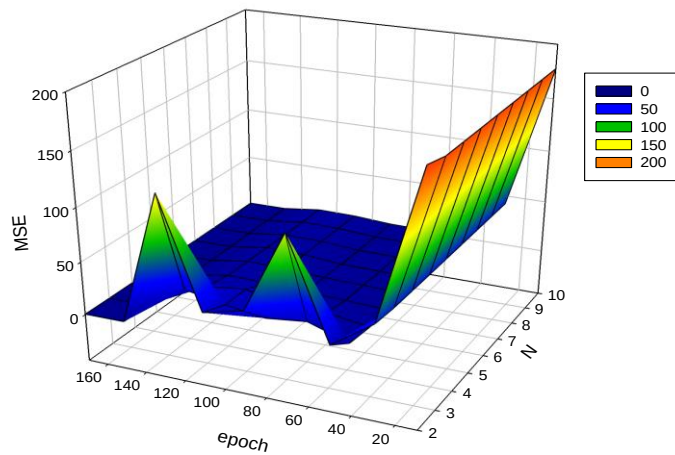


شکل (۳-۳): نمودار مقدار میانگین مربعات خطا برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونیبرگ-مارکوارت (lm)

جدول (۳-۵): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی لونیبرگ-

مارکوارت (lm)

تعداد دور	تعداد گره								
	2	3	4	5	6	7	8	9	10
10	182/6902	44/0232	17/1668	109/5643	33/3384	24/2848	125/2395	0/7665	0/5230
30	177/7873	42/4137	2/1410	14/8041	10/4540	15/6362	23/1403	0/6207	0/2802
50	177/7909	42/2550	1/6438	7/0125	5/3395	15/2946	10/9610	0/6087	0/2584
70	177/7715	42/1433	1/4467	3/4081	4/1810	6/1102	6/7703	0/5958	0/2474
90	177/7820	42/1620	1/4087	3/0179	3/5731	5/4358	6/2658	0/5265	0/2445
110	177/7791	42/1708	1/3449	1/9324	2/3592	5/1350	5/8996	0/5022	0/2400
130	177/7806	42/1911	1/2764	1/8668	2/2154	4/9186	5/3084	0/4850	0/2531
150	177/7844	42/1959	1/2681	1/8142	2/1867	4/7938	5/1703	0/4795	0/2415
170	177/7847	42/1970	1/2633	1/7926	2/1714	4/6199	5/0829	0/4786	0/2397
190	177/7843	42/1975	1/2598	1/6849	2/2039	4/5854	5/0471	0/4771	0/2460

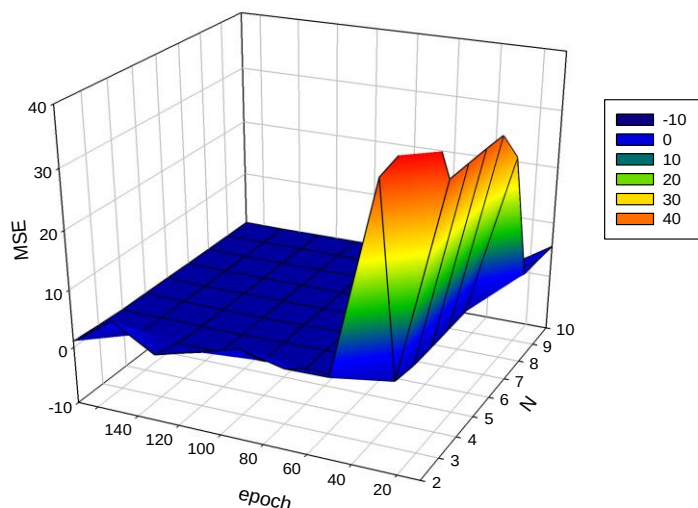


شکل (۳-۴): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل

tansig و الگوریتم آموزشی لونیگ-مارکوارت (lm)

جدول (۳-۶): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br)

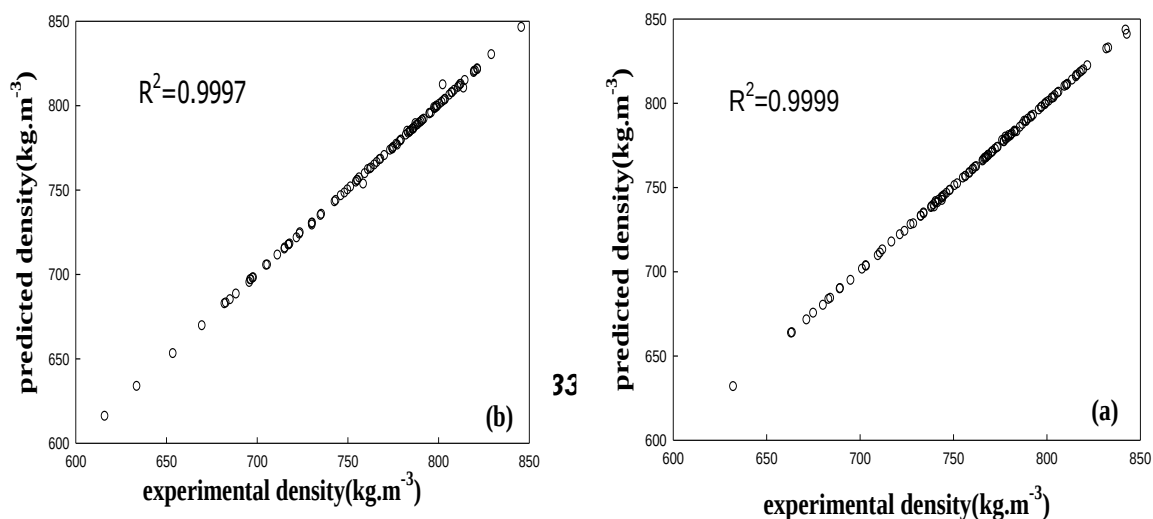
تعداد دور	تعداد گره								
	2	3	4	5	6	7	8	9	10
10	45/8730	36/4881	4/0885	3/9710	5/7255	3/7234	1/6891	5/9760	1/0813
30	34/9100	1/6144	0/4890	0/5884	0/4317	0/4466	0/4754	0/5514	0/7533
50	34/8451	0/9160	0/4648	0/5109	0/4140	0/2657	0/3321	0/4696	0/5055
70	34/7708	0/8232	0/4548	0/4861	0/3206	0/2342	0/2369	0/3879	0/4987
90	34/5827	0/8261	0/4489	0/3469	0/2745	0/2355	0/2394	0/2943	0/4542
110	28/8153	0/8269	0/4448	0/2993	0/2550	0/2392	0/2346	0/2573	0/3889
130	6/2136	0/8269	0/4416	0/2896	0/2412	0/2443	0/2403	0/2355	0/3541
150	5/9817	0/8269	0/4390	0/2880	0/2352	0/2513	0/2489	0/2344	0/3332
170	5/6416	0/8269	0/4264	0/2877	0/2374	0/2592	0/2582	0/2368	0/3183
190	5/5338	0/8269	0/4456	0/2877	0/2529	0/2667	0/2653	0/2480	0/3038



شکل (۳-۵): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br)

3-4-2- ارزیابی مدل غیر خطی

شبکه بهینه شده برای پیش‌بینی دانسیته‌ی سری تایید و تست ترکیبات فوق استفاده گردید. مقادیر MSE و RMAE برای ۰ برای سری تایید به ترتیب ۰/۲۳۳ و ۰/۳۷ و برای سری تست ۱/۴۲۲ و ۰/۰۵۷ می‌باشد. این مقادیر توانایی بسیار خوب شبکه در پیش‌بینی دانسیته‌ی آلکان‌های ۵ تا ۱۹ کربنه را نشان می‌دهند. نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده نیز برای این شبکه برای دو سری تایید و تست در شکل (۳-۶) رسم شده است. مقدار R^2 نزدیک به یک نشان‌دهنده‌ی کارایی بالای شبکه ارائه شده در پیش‌بینی دانسیته می‌باشد.



شکل (۳-۶): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای آلکان‌های ۵ تا ۱۹ کربنه با استفاده از شبکه‌ی بهینه شده برای دو سری (الف) تایید و (ب) تست

فصل چهارم: مطالعه‌ی رابطه کمی ساختار -

خاصیت (QSPR) برای پیش‌بینی دانسیته‌ی اجزاء

گاز طبیعی مایع با استفاده از شبکه عصبی مصنوعی

۴-۱- انتخاب سری داده‌ها:

برای دستیابی به مدلی مناسب، اولین مرحله جمع‌آوری مقادیر تجربی دانسیته‌ی اجزاء گاز طبیعی مایع از C₂-C₄ در محدوده‌ی وسیعی از دما و فشار می‌باشد. نام ترکیبات، به همراه محدوده‌ی دما و فشار و مرجع مورد استفاده برای هر ترکیب در جدول (۴-۱) آمده است. داده‌های بدست آمده به سه سری آموزش، تایید و تست تقسیم شدند که سری آموزش ۵۴۳، تایید ۱۵۹ و تست ۱۴۳ داده را شامل می‌شدند.

جدول (۴-۱): محدوده‌ی دما و فشار برای اجزاء گاز طبیعی از C₂-C₄ به همراه مرجع مورد استفاده برای هریک از آنها

مرجع	Δp (MPa)	ΔT (K)	نام ترکیب
[۲۱]	۵ - ۷۰	۹۵ - ۳	اتان
[21]	۵ - ۱	۹۰ - ۳۶۵	پروپان
[22]	۱ - ۲۰۰	۲۸۰ - ۳۸۰	نرمال بوتان
[23]	۱ - ۲۰۰	۲۸۰ - ۳۸۰	ایزو بوتان

۴-۲- انجام محاسبات و مدل سازی

۴-۲-۱- رسم و بهینه سازی ساختار مولکول ها

دومین مرحله در مطالعات QSPR رسم و بهینه سازی ساختار مولکول های مورد بررسی می باشد. بدین منظور مولکولها توسط نرم افزار Hyperchem به صورت تقریبی رسم گردید و سپس با کمک روش نیمه تجربی AM1 ساختار هر ترکیب بهینه شد.

۴-۲-۲- محاسبه توصیف گرهای مولکولی

یکی از مراحل مهم در هر مطالعه ی QSPR تعیین و انتخاب توصیف گرهای مناسب می باشد. با توجه به این که توصیف کننده های زائد کار پردازش را طولانی می کنند لذا از توصیف کننده هایی استفاده می شود که با پارامتر مورد نظر ارتباط نزدیکی داشته باشند. از آنجایی که دانسیته ی هر ترکیب به دما و فشار بستگی دارد لذا این دو ویژگی به عنوان اولین توصیف کننده ها انتخاب گردیدند. علاوه بر دما و فشار که دانسیته به آنها وابسته می باشد، سایر توصیف کننده ها مطابق روش زیر انتخاب شدند:

ساختارهای بهینه شده توسط نرم افزار Hyperchem به عنوان ورودی برای نرم افزار دراگون استفاده شد و ۱۴۸۱ توصیف کننده به عنوان خروجی این برنامه محاسبه گردید.

از آنجا که توصیف کننده های بدست آمده از نرم افزار دراگون بسیار زیاد می باشد در نتیجه توصیف کننده هایی که دارای ۹۰٪ مقادیر یکسان بودند و نیز یکی از توصیف کننده ها از میان جفت توصیف کننده هایی که دارای ضریب همبستگی بیش از ۰/۹ بودند، از لیست توصیف کننده ها حذف شدند.

۴-۲-۳- مدل رگرسیون خطی چندگانه برای پیش بینی دانسیته ی اجزاء گاز طبیعی

پس از انتخاب توصیف کننده های مناسب، مدل رگرسیون خطی چندگانه برای پیش بینی دانسیته اجزاء گاز طبیعی طراحی شد. برای این منظور از روش رگرسیون مرحله به مرحله استفاده گردید به این صورت که متغیرها یکی پس از دیگری وارد مدل شدند. سپس با ورود هر متغیر کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معنی داری خود را از دست داده بودند حذف شدند.

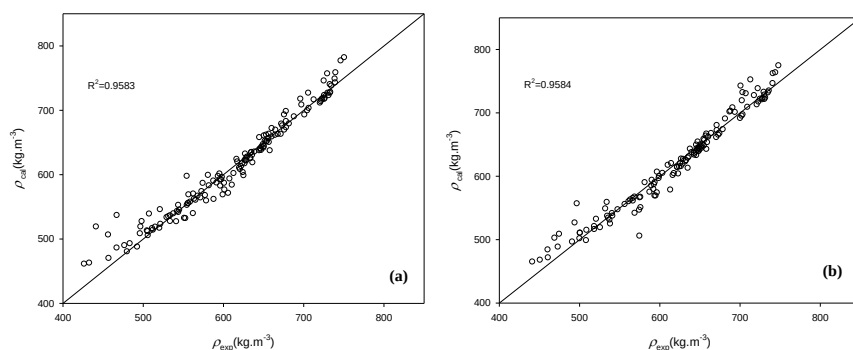
برای انجام این کار توصیف‌کننده‌های انتخاب شده در قسمت قبل به علاوه دما و فشار به عنوان متغیر مستقل و دانسیته به عنوان متغیر وابسته به نرم افزار SPSS داده شدند و سپس با استفاده از روش آنالیز و انتخاب رگرسیون مرحله‌ای، چندین مدل بدست آمد که بهترین مدل با توجه به خصوصیات آماری، مدلی است که دارای بیشترین مقدار R ، F و کمترین مقدار خطای استاندارد باشد. با استفاده از این اطلاعات بهترین مدل انتخاب گردید. سپس توصیف‌کننده‌ها در بهترین مدل را برای داده‌های سری آموزش به کار برده شد و با انتخاب رگرسیون خطی و روش اجباری ضرایب توصیف‌کننده‌ها در مدل استخراج گردید. معادله‌ی (1-4) برای پیش‌بینی دانسیته‌ی اجزاء گاز طبیعی بدست آمد.

$$D = 2957.863 - 2140.3254 \text{ MATS2m} + 439.608 \text{ Mor04V} + 0.735 P - 1.002 T \quad (1-4)$$

۴-۲-۴- ارزیابی مدل خطی

برای ارزیابی مدل رگرسیون خطی بدست آمده برای دانسیته، داده‌های سری تایید و تست در مدل خطی بدست آمده برای پیش‌بینی دانسیته، قرار داده شد و مقادیر MSE و متوسط درصد خطا برای این دو سری محاسبه گردید. مقادیر MSE و متوسط درصد خطا به ترتیب برای سری تایید 196/46 و 1/777 و برای سری تست 200/55 و 1/76 به دست آمد.

نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده نیز برای این مدل برای دو سری تایید و تست در شکل (۱-۴) رسم شده است. این شکل نشان می‌دهد که یک ارتباط غیر خطی بین توصیف‌کننده‌ها و دانسیته وجود دارد.



شکل (۱-۴): نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده برای اجزاء گاز طبیعی با استفاده از مدل خطی برای دو

سری (الف) تایید و (ب) تست

۴-۳- مدل شبکه عصبی مصنوعی و بهینه‌سازی آن

در این تحقیق هر چند مدل رگرسیون خطی به خوبی توانست دانسیته‌ی اجزاء گاز طبیعی را پیش‌بینی کند اما برای بهتر شدن نتایج از مدل غیر خطی شبکه عصبی مصنوعی نیز استفاده گردید. برای بهینه‌سازی عملکرد شبکه باید پارامترهای شبکه بهینه شوند که برخی از آنها عبارتند از:

۱- نوع الگوریتم آموزشی

۲- تعداد نرون لایه مخفی

۳- تعداد چرخه‌های آموزش

۴- سرعت آموزش

۵- ممنتوم

۶- نوع تابع انتقال

معیار بهینه‌سازی پارامترها به حداقل رساندن میانگین مربعات خطا در سری تایید می‌باشد.

۴-۳-۱- انتخاب ورودی شبکه، الگوریتم آموزشی، تعداد گره لایه‌ی پنهان و تابع انتقال:

در مدل‌سازی شبکه‌ی عصبی مصنوعی، شبکه با استفاده از توصیف‌کننده‌های مدل منتخب MLR به عنوان ورودی طراحی گردید. در شروع برنامه، مقدار وزنهای ورودی و خروجی به صورت تصادفی از اعداد بین صفر و یک انتخاب شدند و بعد از این مقادیر حین آموزش بهینه شدند.

برای آموزش شبکه دو الگوریتم آموزشی بایزین و لونیبرگ-مارکوارت مورد بررسی قرار گرفتند. مرحله‌ی بعد انتخاب تابع انتقال مناسب برای لایه پنهان شبکه می‌باشد. برای این منظور دو تابع انتقال تانژانت سیگموئید (tansig)، لگاریتم سیگموئید (logsig) در برنامه استفاده شدند. سپس چهار شبکه‌ی طراحی شده با دو الگوریتم آموزشی لونیبرگ-مارکوارت و تنظیم بایزین و دو تابع انتقال مختلف، به ازاء همه‌ی ترکیبهای ممکن از مقادیر مختلف دو پارامتر دیگر شبکه یعنی؛ تعداد نرونهای لایه‌ی مخفی و تعداد دور آموزش، آموزش داده شد. نتایج حاصله برای سری تایید در جداول (۴-۲) تا (۴-۵) و نمودارهای (۴-۲) تا (۴-۵) نشان داده شده است.

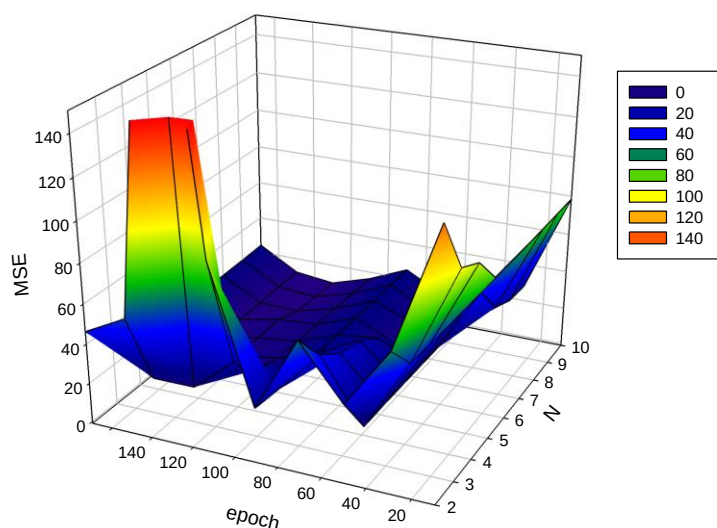
با توجه به نتایج بدست آمده، شبکه با الگوریتم آموزشی لونیگ-مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره ۸ در لایه پنهان و تعداد دور آموزشی ۱۷۰ دارای کمترین مقدار خطا برای سری تایید می‌باشد. از آنجا که سرعت آموزش با این الگوریتم بالاست و نیز با تغییر مقدار سرعت آموزش تغییری در خطای شبکه ایجاد نمی‌شود بنابراین نیازی به بهینه کردن سرعت آموزش نیست.

جدول (۴-۲): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی لونیگ-

مارکوارت (lm)

تعداد دور	تعداد گره								
	2	3	4	5	6	7	8	9	10
10	128/24 3	67/62 1	38/09 0	63/97 5	26/30 3	92/78 5	280/16 6	56/96 6	46/35 6
30	100/62 3	52/53 5	16/48 5	47/43 5	25/02 0	34/80 9	17/349	17/26 5	29/01 0
50	94/648	51/76 7	16/83 7	38/58 7	24/84 9	21/15 2	14/854	14/96 9	28/25 9
70	79/359	51/23 3	16/78 8	35/55 8	24/77 6	17/29 2	14/536	10/38 7	27/15 6
90	79/004	51/11 8	16/75 4	27/39 0	24/76 5	17/08 1	14/501	10/59 9	23/10 5
110	78/574	50/81 4	16/70 3	14/90 0	24/80 8	16/86 3	15/378	11/23 5	22/48 2
130	78/425	33/92 9	16/75 2	17/18 4	24/92 0	16/71 5	10/986	11/56 6	23/22 1
150	78/425	28/82 7	16/73 0	14/06 7	25/08 7	16/30 4	10/257	12/08 7	23/53 4

170	78/347	28/02 5	16/73 1	14/05 1	25/17 6	16/13 5	9/703	12/11 2	24/03 6
190	78/307	27/30 7	16/71 1	14/05 0	25/16 7	16/00 9	9/829	11/88 4	23/75 1



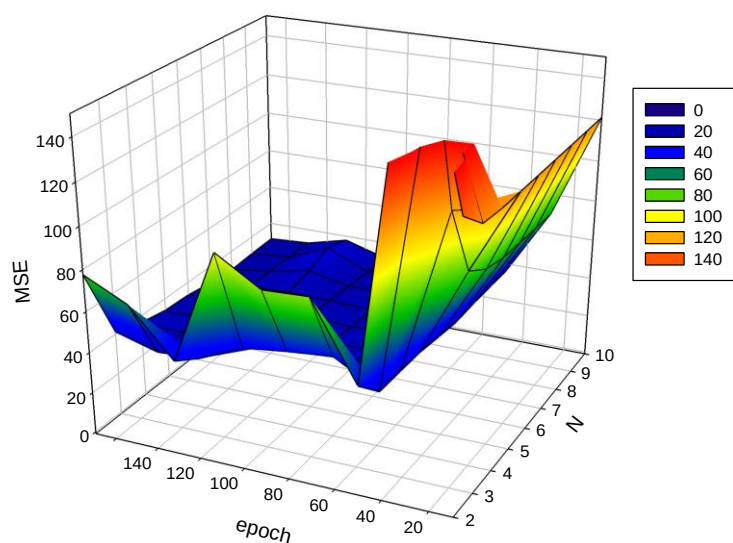
شکل (۴-۲): نمودار مقدار میانگین مربعات خطا در سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل

logsig و الگوریتم آموزشی لونیگ-مارکوارت (lm)

جدول (۴-۳): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل logsig و الگوریتم آموزشی بایزین (br)

تعداد دور	تعداد گره								
	2	3	4	5	6	7	8	9	10
10	133/01 2	224/40 8	51/19 0	87/47 2	86/51 1	99/45 8	44/72 8	67/47 2	77/647
30	122/38 6	187/22 6	38/49 2	50/85 9	48/75 8	46/13 6	36/55 7	35/21 3	41/101
50	121/39 6	116/86 7	38/42 0	39/45 3	39/01 5	37/38 3	33/59 0	35/85 0	36/663
70	121/34 9	81/562	38/15 5	34/34 8	34/64 9	35/08 4	32/00 7	33/54 5	34/161

90	121/33 7	74/168	36/49 7	35/90 3	33/16 0	34/55 3	30/69 2	33/47 7	33/512
110	121/33 1	71/572	35/00 5	33/77 6	33/24 2	33/91 5	34/18 2	33/37 6	31/391
130	121/33 0	70/897	35/66 5	34/40 0	33/28 9	32/90 2	37/28 4	33/15 0	30/14 7
150	121/33 0	70/383	35/71 1	34/41 9	33/23 3	32/13 2	41/55 6	32/83 7	30/883
170	121/33 0	70/080	35/71 5	34/38 7	33/10 9	31/66 6	37/81 0	32/55 7	31/279
190	121/33 0	69/992	35/71 6	34/00 1	32/69 3	31/35 9	37/55 1	32/39 2	31/367



شکل (۴-۳): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع

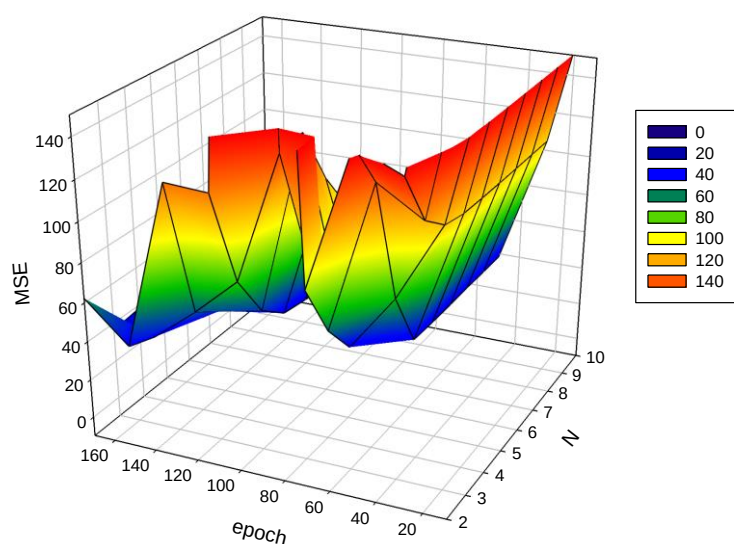
تبدیل $\log\text{sig}$ و الگوریتم آموزشی بایزین (br)

جدول (۴-۴): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی لونبرگ-

مارکوارت (lm)

	تعداد گره
--	-----------

تعداد دور	2	3	4	5	6	7	8	9	10
10	258/504	146/018	155/696	87/427	582/787	124/564	125/931	42/216	62/238
30	226/996	118/911	132/606	58/489	138/818	74/139	54/715	37/364	36/447
50	203/392	108/847	69/4150	40/497	141/098	50/406	45/860	37/783	38/216
70	203/118	106/173	39/1230	39/050	112/439	39/6830	42/7820	28/6360	39/9390
90	202/959	105/617	38/5740	37/836	91/030	37/472	38/964	27/324	40/887
110	202/884	105/340	38/4440	36/008	74/344	35/432	38/102	26/782	41/569
130	202/839	105/165	38/6100	33/044	66/3780	33/790	38/231	29/491	41/780
150	202/833	105/044	38/3040	31/007	67/754	32/725	40/441	32/086	41/927
170	202/831	104/958	38/2620	30/165	67/5260	31/154	40/600	32/443	42/069
190	202/828	104/893	38/2310	27/941	67/139	30/689	40/608	32/752	42/205

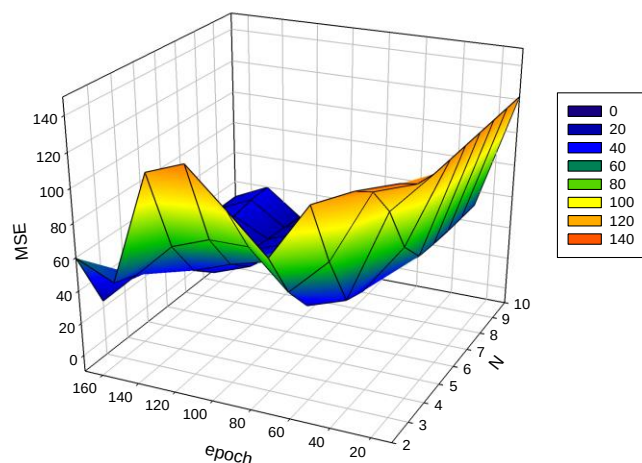


شکل (۴-۴): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع

تبدیل tansig و الگوریتم آموزشی لونبرگ-مارکوارت (lm)

جدول (۴-۵): مقدار میانگین مربعات خطا در سری تایید برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br)

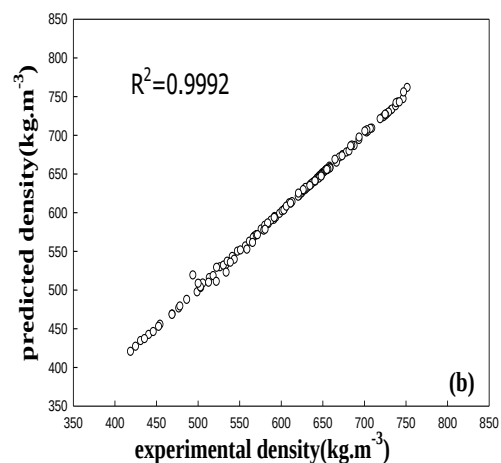
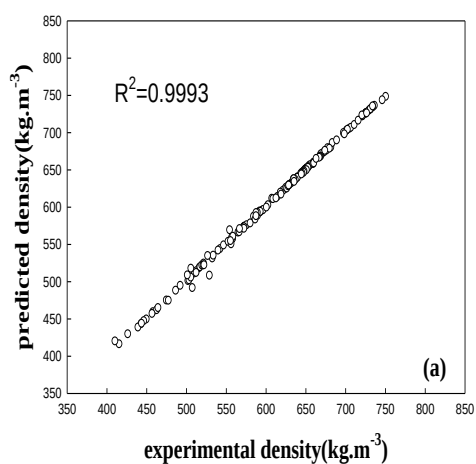
تعداد د	تعداد گروه								
	2	3	4	5	6	7	8	9	10
10	132/70 8	125/00 9	114/19 4	81/329	103/30 0	124/56 4	116/33 0	49/50 1	59/698
30	124/45 6	116/53 3	74/407	52/391	74/364 0	74/139	65/904	44/64 9	23/907
50	121/77 1	97/635	42/301	34/399	56/372 0	50/406	42/171	45/06 8	25/676 0
70	121/70 8	68/658	38/052	32/952	54/925	39/683	31/448	35/92 1	27/399
90	121/70 7	54/049	37/668	31/738	53/711	37/472	29/237	34/60 9	28/347
110	121/70 7	52/506	36/998	29/910	51/883 0	35/432	27/197	34/06 7	29/029
130	121/70 7	51/476	36/024	26/946	48/919	33/790	25/555	36/77 6	29/240
150	121/70 7	51/112	35/524	24/909	46/882	32/725	24/490	39/37 1	29/387
170	121/70 7	50/816	35/356	24/067	46/040	31/154	22/919	39/72 8	29/529
190	121/70 7	50/929	35/281	21/84 3	43/816	30/689	22/454	40/03 7	29/665



شکل (۴-۵): نمودار مقدار میانگین مربعات خطا برای سری تایید برحسب تعداد گره و دور آموزش برای شبکه‌هایی با تابع تبدیل tansig و الگوریتم آموزشی بایزین (br)

۴-۳-۲- ارزیابی مدل غیر خطی

شبکه بهینه شده برای پیش‌بینی دانسیته سری تایید و تست ترکیبات فوق استفاده گردید. مقادیر MSE و RMAE به ترتیب برای سری تایید ۹/۷۰ و ۲۸۰۱/ و برای سری ارزیابی ۱۰/۳۱ و ۰/۲۸۳۸ می‌باشد. این مقادیر توانایی بسیار خوب شبکه در پیش‌بینی دانسیته‌ی اجزاء گاز طبیعی را نشان می‌دهند. نمودار مقادیر تجربی برحسب مقادیر پیش‌بینی شده نیز برای این شبکه برای دو سری تایید و تست در شکل (۴-۶) رسم شده است. مقدار R^2 نزدیک به یک نشان‌دهنده‌ی کارایی بالای شبکه ارائه شده در پیش‌بینی دانسیته می‌باشد.



شکل (۴-۶): نمودار مقادیر تجربی بر حسب مقادیر پیش‌بینی شده برای اجزاء گاز طبیعی با استفاده از شبکه‌ی بهینه شده برای دو سری (الف) تایید و (ب) تست

مراجع:

- ۱[- <http://www.neshatsoft.com>]
- ۲[- <http://www.daneshnameh.roshd.ir/mavara>]
- ۳[- Marini F., Bucci R., Magri A. L., Magri A. D., "Artificial neural network in chemometrics: History, examples and perspectives" *MICROCHEMICAL JOURNAL*, Vol. 88, pp. 178-185, 2008.
- ۴[- منهاج، محمدباقر، "هوش محاسباتی"، جلد اول، دانشگاه صنعتی امیرکبیر (واحد تفرش)، چاپ اول، ۱۳۷۹.
- ۵[- سعیدی، مسعود، "عملیات شبکه‌های عصبی (قسمت دوم)"، ماهنامه شبکه، شماره ۵۲، ۱۳۸۳.

- [۶] - شیرپور، الهه، " کاربرد مطالعات رابطه کمی ساختار - فعالیت برای پیش‌بینی و مدل‌سازی فعالیت دارویی برخی از داروهای سنتز شده جدید " پایان‌نامه کارشناسی ارشد، دانشکده شیمی، دانشگاه صنعتی شاهرود، ۱۳۸۸.
- [۷] - یامی، فاطمه، " مدل سازی QSPR اندیس بازداري هیدروکربن‌های آروماتیک چند حلقه‌ای با روش رگرسیون خطی چندگانه و شبکه عصبی مصنوعی " پایان‌نامه کارشناسی ارشد، دانشکده شیمی، دانشگاه صنعتی شاهرود، ۱۳۸۸.
- [۸] - <http://www.emsl.pnl.gov:2008/docs/incinc/homepage.html>.
- [۹] - Howery D.G., Hirsch R.F., "chemometrics in the chemistry curriculum" *J. Chem. Ed.*, 60, pp:656-659, 1983.
- [۱۰] - Hilbert D.B., "Genetic algorithms in chemistry" *chemometr. intel. Lab. Syst.*, 19 pp:277-293, (1993).
- [۱۱] - Karelson M., "Molecular descriptors in QSPR/QSAR" Wiley-VCH, Newyork, 2000.
- [۱۲] - Arab Chamjangali M., Beglari M., Bagherian G., " Prediction of cytotoxicity data of anti-HIV 1-[5-chlorophenyl sulphonyl] - 1H-Pyrrol Derivatives using Calculated molecular descriptors and Levenberg-Marquardt artificial neural network" *Chem. Biol Derug desm*, 2009.
- [۱۳] - <http://hassangolmohamadi.persianblog.ir>.
- [۱۴] - Oliveira. C. M. B. P., Wakeham. W. A., "The Viscosity of Five Liquid Hydrocarbons at Pressures Up to 250 MPa" *International Journal of Thermophysics*, Vol. 13, 773-790, 1992.
- [۱۵] - Dymond. J. H., Young. K. J., Isdale. J. D., "p, T behaviour for n-hexane+ n-hexadecane in the range 298 to 373 K and 0.1 to 500 MPa" *J. Chem. Thermodynamics.*, Vol. 11, pp. 887-895, 1979.
- [۱۶] - Morávková. L., Linek. J., "Excess molar volumes of (heptane + 1-chlorobutane) at elevated temperatures and high pressures" *J. Chem. Thermodynamics*, Vol. 34, pp. 1397-1405, 2002.
- [۱۷] - Grindley. T., Lind. John E., "The residual energy and entropy of hydrocarbon liquids and their relation to the liquid structure" *J. Chem. Phys*, 68, 11, pp. 5046-5052, 1978.
- [۱۸] - Khasanshin. T. S., Shchamaliou. A. P., Poddubskij. O. G., "Thermodynamics properties of heavy n-alkanes in the liquid state: n-Dodecane" *International Journal of Thermophysics*, 24, 5, pp. 1277-1289, 2003.
- [۱۹] - Daridon. J. L., Carrier. H., Lagourette. B., "pressure dependence of thermophysical properties of n-pentadecane and n-heptadecane" *International Journal of Thermophysics*, Vol. 23, pp. 697-708, 2002.
- [۲۰] - Dutour. S., Daridon. J. L., Lagourette. B., "pressure and temperature dependence of the speed of sound and related properties in normal Octadecane and Nonadecane" *International Journal of Thermophysics*, Vol. 21, pp. 173-180, 2000.
- [۲۱] - <http://webbook.nist.gov>.
- [۲۲] - H. Miyamoto, M. Uematsu, *J. Chem. Thermodynamics*, Vol 38, pp. 360, 2006.
- [۲۳] - H. Miyamoto, M. Uematsu, *J. Chem. Thermodynamics*, Vol 39, pp. 588, 2007.