
بسم الله الرحمن الرحيم

دانشکده: شیمی

گروه شیمی فیزیک

مطالعه‌ی کمی ساختار –خاصیت ضریب فعالیت در رقت بی‌نهایت ترکیبات آلی و آب در محیط مایع یونی ۱-بوتیل ۱-متیل پیرولیدینیوم تریسیانومتانید

دانشجو: **مریم طاهرزاده**

استاد راهنما:

دکتر زهرا کلانتر

استاد مشاور:

دکتر ناصر گودرزی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۹۳

سپاس بی‌کران پروردگار یکتا را که هستی‌مان بخشید و به طریق علم و دانش رهنمونمان شد و به همنشینی رهروان علم و دانش مفتخرمان نمود و خوشه‌چینی از علم و معرفت را روزیمان ساخت

تقدیم به :

مقدس‌ترین واژه‌ها در لغت نامه‌ی دلم:

مادر مهربانم که سجده‌ی ایثارش گل محبت را در وجودم پروراند و دامان گهربارش لحظه‌های مهربانی را به من آموخت.

پدر عزیزم که مهر آسمانی‌اش آرام بخش آلام زمینی‌ام است.

خواهرانم، همراهان همیشگی و پشتوانه‌های زندگیم.

تشکر و قدردانی

اکنون که در پرتو الطاف الهی خداوند سبحان این مقطع تحصیلی را به پایان رساندم، وظیفه می‌دانم از زحمات خانواده عزیزم که با صبر و تحمل مشکلات مسیر را برایم تسهیل نمودند تشکر کنم.

با تقدیر و تشکر شایسته از استاد فرهیخته و فرزانه سرکار خانم دکتر زهرا کلانتر که با نکته‌های دلاویز و گفته‌های بلند، صحیفه‌های سخن را علم پرور نمود و همواره راهنما و راه گشای نگارنده در اتمام واکمال پایان نامه بوده است.

از جناب آقای دکتر ناصر گودرزی به خاطر نظرات ارزنده به عنوان استاد مشاور نهایت قدردانی را دارم و سلامتی و موفقیت همیشگی این بزرگوران را از درگاه یزدان پاک خواستارم.

از زحمات بی‌دریغ خانم دکتر اشرفی، خانم دکتر دوستی، خانم عجم،خانم شهابی، ابراهیمی، خانم کشتکار و آقای مهندس سلیمی، نیز در طول انجام پایان‌نامه صمیمانه تشکر می‌کنم.

چکیده

در این تحقیق مطالعه‌ی ارتباط کمی ساختار- خاصیت (QSPR)، برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت ۵۸ ترکیب آلی و آب در محیط مایع یونی، ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید [BMPYR][TCM]، در ۶ دمای مختلف انجام شده‌است. تعداد زیادی توصیف‌کننده شامل ۱۸ دسته‌ی مختلف، توسط نرم افزار Dragon محاسبه شدند و سپس برای انتخاب توصیف‌کننده‌های مهم از دو روش رگرسیون مرحله‌ای (SR) و الگوریتم ژنتیک بر اساس آنالیز حداقل مربعات (GA-PLS) استفاده گردید. تعداد ۱۲ توصیف‌کننده توسط روش SR و ۱۰ توصیف‌کننده توسط روش GA-PLS انتخاب شدند. توصیف‌کننده‌های انتخاب شده توسط این دو روش برای مدل‌سازی و پیش‌بینی ضریب فعالیت در رقت بی‌نهایت این ترکیبات به عنوان ورودی به شبکه‌ی عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) داده شدند. عملکرد هر مدل توسط سری تست ارزیابی مورد بررسی قرار گرفت. نتایج به دست آمده نشان از برتری روش SR-ANN نسبت به دیگر روش‌های به کار برده شده جهت پیش‌بینی ضریب فعالیت در رقت بی‌نهایت ترکیبات مورد مطالعه دارد. میانگین مربعات خطا (MSE) و میانگین درصد انحراف مطلق (AAD) سری تست برای روش‌های SR-ANN، GA-ANN، SR-SVM و GA-SVM به ترتیب برابر ۰/۰۱۷۹، ۲/۴۳۰٪ و ۰/۰۲۰۱، ۳/۴۸۳٪ و ۰/۰۶۰۱ ، ۱۰/۱۳۵٪و ۴/۰۳۱ و ۲۸/۸۷۴٪ می‌باشد.

کلمات کلیدی: ضریب فعالیت در رقت بی‌نهایت، رگرسیون مرحله‌ای، الگوریتم ژنتیک، شبکه‌ی عصبی مصنوعی، ماشین برداری پشتیبان.

نتایج حاصل از این پایان‌نامه در دو پوستر تحت عناوین

“Prediction of activity coefficients at infinite dilution for organic solutes in ionic liquids [BMPYR] [TCM] by artificial neural network”

“Prediction of activity coefficient at infinite dilution for some organic solutes in ionic liquid of [BMPYR] [TCM] by using GA-SVM and SR-SVM”

در هفدهمین کنفرانس شیمی فیزیک خواجه نصیر طوسی ارائه گردید.

فهرست مطالب

فصل اول: مقدمه

فصل دوم: کمومتریکس

فهرست اشکال

فهرست جدول‌ها

فصل اول

مقدمه

مقدمه

ضریب فعالیت در رقت بی‌نهایت[۱] که با γ_1^{∞} نشان داده می‌شود پارامتر مهمی در صنعت استخراج است که اطلاعاتی در مورد برهمکنش حلال-حل شونده در اختیار ما قرار می‌دهد. از این ضریب برای تفکیک گزینش‌پذیر حل شونده‌ها استفاده می‌شود. بنابراین اندازه‌گیری این خاصیت کلیدی بسیار حائز اهمیت می‌باشد. مقادیر ضریب فعالیت در رقت بی‌نهایت به نوع حلال بستگی دارد [۱].

دسته جدیدی از حلال‌ها که در چند سال اخیر در مراکز تحقیقاتی و صنایع شیمیایی انقلابی به پا کرده‌اند مایعات یونی می‌باشند. این ترکیبات جزء مواد شیمیایی سبز هستند که به عنوان حلال، نقش بسیار مهمی را در کاهش استفاده از حلال‌های متدوال آلی که، سمی و آسیب زننده به محیط زیست هستند، دارا می‌باشند. مایعات یونی دارای خواص مهمی همچون فشار بخار بسیار ناچیز(عدم آلایندگی)، پایداری حرارتی در دماهای بالا، هدایت یونی زیاد و توانایی برای انحلال ترکیباتی با قطبیت های گوناگون هستند. همچنین مایعات یونی در فرآیندهای تفکیک، استخراج و تقطیر نقش مهمی را ایفا می‌کنند [۲].

اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت، توسط کروماتوگرافی گاز-مایع انجام می‌شود که در ادامه به توضیح مختصری راجع به کروماتوگرافی، انواع آن و مبانی تئوری استخراج این ضریب از روی اطلاعات کروماتوگرافی خواهیم پرداخت.

۱-۱- کروماتوگرافی[۲]

در سال ۱۹۰۶ دانشمند روسی بنام تسوت[۳] گزارشی مبنی بر جداسازی اجزای رنگی موجود در عصاره‌ی برگهای گیاهی با عبور دادن آن از ستونی حاوی کربنات کلسیم و آلومینا، بود ارایه داد. او بر

۱- Activity coefficient at infinite dilution
۲- Choromatography
۳- Tsvet

همین اساس واژه‌ی کروماتوگرافی را انتخاب نمود که از لغات یونانی کروم، به معنی رنگ و گرافی، به معنای نوشتن گرفته شده بود [۳].

واژه‌ی کروماتوگرافی امروزه به دسته‌ای از روش‌ها اطلاق می‌شود که در آنها جداسازی بر مبنای تمایل نسبی هر جزء به فاز ساکن است. گونه‌ای که تمایل بیشتری به فاز متحرک دارد، با سرعت بیشتری حرکت می کند و بالعکس گونه‌ای که به فاز ساکن تمایل بیشتری دارد، با سرعت کمتری در طول ستون حرکت می کند. با روش‌های کروماتوگرافی می توان جداسازی‌هایی را که به روش‌های دیگر خیلی مشکل می باشند، انجام داد. یکی از مزایای برجسته‌ی روش‌های کروماتوگرافی این است که احتمال تجزیه‌ی مواد جداشونده به وسیله‌ی این روش‌ها در مقایسه با سایر روش‌ها کمتر است. مزیت دیگر روش‌های کروماتوگرافی، در این است که تنها مقدار بسیار کمی از مخلوط برای تجزیه لازم است، به همین دلیل روش‌های تجزیه‌ای مربوط به جداسازی مواد کروماتوگرافی می‌توانند در مقیاس میکرو و نیمه میکرو انجام گیرند. روش‌های کروماتوگرافی، ساده، سریع و وسایل مورد نیاز آنها ارزان هستند. مخلوط‌های پیچیده را می‌توان تقریباً به آسانی به وسیله این روش‌ها جدا کرد.

۱-۱-۱- کروماتوگرافی گازی[۱]

کروماتوگرافی گازی، یکی از روش‌های کروماتوگرافی است که برای بررسی و جداسازی مواد فرار بدون تجزیه شدن آنها، به کار می‌رود. در این روش، فاز گازی یک فاز بی‌اثر مثل هلیوم، نیتروژن، آرگون یا دی‌اکسیدکربن است که به آن فاز متحرک یا گاز حامل نیز می‌گویند. فاز ساکن یک جسم جامد جاذب و یا لایه نازکی از یک مایع غیر فرار است که به دیواره‌ی داخلی ستون یا به صورت پوششی روی سطح گلوله‌های شیشه‌ای یا فلزی قرار داده می‌شود. در صورتیکه فاز ساکن، جسم جامد جاذب باشد، به آن

۱- Gas Chromatography

کروماتوگرافی گازی (GC) و اگر فاز ساکن، مایع غیر فرار باشد، به آن کروماتوگرافی گاز-مایع[۱] (GLC) می‌گویند. اما هر دو به کروماتوگرافی گازی معروف هستند. در کروماتوگرافی گازی، جداسازی اجزای یک مخلوط متناسب با میزان توزیع اجزای تشکیل‌دهنده‌ی مخلوط بین فاز متحرک گازی و فاز ساکن جامد یا مایع صورت می‌گیرد. در این روش گاز حامل، مخلوط را درون ستون حرکت می‌دهد و در حالت تعادل، اجزای تشکیل‌دهنده‌ی مخلوط بین دو فاز متحرک و ساکن توزیع می‌شوند. بنابراین، فاز متحرک اجزای تشکیل‌دهنده‌ی نمونه را به طرف بیرون ستون حرکت می‌دهد. هر مولکولی که با ارتباط ضعیف‌تری جذب ستون شده، زودتر و جزیی که قدرت جذب بیشتری در ستون دارد، دیرتر از ستون خارج می‌شود و بدین‌ترتیب اجزای مخلوط از یکدیگر جدا می‌شوند. به همین دلیل، از کروماتوگرافی گازی می‌توان برای جداسازی و شناسایی اجزای تشکیل‌دهنده‌ی یک مخلوط و تجزیه‌ی کمی آنها استفاده کرد.

۱-۲- تعیین ضریب فعالیت در رقت بی‌نهایت با استفاده از مبانی تئوری GLC

یکی از روش‌های تعیین ضریب فعالیت در رقت بی‌نهایت برای حل‌شونده‌های مختلف در حلال‌های غیرفرار، کروماتوگرافی گاز-مایع است [۴]. در این روش، ضریب فعالیت حل‌شونده در رقت بی‌نهایت از رابطه‌ای که بین حل‌شونده، گاز حامل و حلال مایع غیر فرار وجود دارد به دست می‌آید. این رابطه که اولین بار توسط اورت[۲] و کرویک شانگ[۳] ارائه شد به صورت زیر است: [۱]

$$\ln \gamma_{13}^{\infty} = \ln(\frac{n_3 RT}{V_N p_1^*}) - \frac{p_1^*(B_{11} - V_1^*)}{RT} + \frac{p_0 J_2^3 (2B_{13} - V_1^{\infty})}{RT} \qquad (۱-۱)$$

که در آن γ_{13}^{∞}، ضریب فعالیت در رقت بی‌نهایت، n_3 تعداد مول حلال در ستون، R ثابت گازها، T دمای ستون، V_N حجم خالص باقیمانده از حل‌شونده، p_1^* فشار بخار حل‌شونده‌ی خالص در دمای T، B_{11} ضریب

۱- Gas- Liquid Chromatography
۲- Evertt
۳- Cruickshunk

دوم ویریال حل‌شونده‌ی خالص، V_1^* حجم مولی حل‌شونده‌ی خالص، p_0 فشار خروجی ستون، $p_0 J_2^3$ فشار متوسط ستون، B_{13} ضریب دوم ویریال مخلوط حل‌شونده و گاز حامل، V_1^∞ حجم مولی جزئی حل‌شونده در رقت بی‌نهایت در حلال می‌باشد. لازم به تذکر است که در نماد γ_{13}^∞، علامت ∞ به رقت بی‌نهایت محلول، عدد ۱ به جزء حل‌شونده در محلول و عدد ۳ به حلال اشاره می‌کند و بنابراین، γ_{13}^∞ به معنی ضریب حل‌شونده‌ی جزء ۱ در حلال۳ در رقت بی‌نهایت است[۱].

حجم خالص باقیمانده‌ی حل‌شونده، V_N، نیز از رابطه‌ی (۱-۲) به دست می‌آید:

$$V_N = (J_2^3)^{-1} u_0 (t_R - t_G) \qquad (۱-۲)$$

که t_R و t_G به ترتیب زمان بازداری حل‌شونده در ستون[۱] و زمان بازداری ترکیبی که در داخل ستون بازداری نمی‌شود[۲] و u_0 سرعت جریان خروجی ستون است که برای فشار بخار آب به صورت رابطه‌ی (۱-۳) تصحیح می‌شود.

$$u_0 = u(1 - \frac{p_w}{p_0})\frac{T}{T_f} \qquad (۱-۳)$$

T_f دمای خروجی ستون، p_w فشار بخار آب در T_f و u سرعت جریان گاز است که با جریان سنج[۳] گاز اندازه‌گیری می‌شود.

J_2^3 در رابطه‌ی (۱-۲)، جمله‌ی تصحیح فشار است، و از عبارت زیر به دست می‌آید:

$$J_2^3 = \frac{2}{3}\frac{(p_i/p_0)^3 - 1}{(p_i/p_0)^2 - 1} \qquad (۱-۴)$$

که p_i فشار ورودی ستون، و p_0 فشار خروجی ستون است.

۱- Retention time for a nonretained compound
۲- Retention time for a retained compound
۳- Flowmeter

۱-۲-۱- مبانی تئوری برای استخراج معادله‌ی اورت

برای اولین بار، اورت و همکارانش معادلات دقیقی را برای محاسبه‌ی ضریب فعالیت در رقت بی‌نهایت از اندازه‌گیری‌های فشار بخار ایستا به دست آوردند. برای یک محلول دوتایی متشکل از دو جزء ۱ و ۲ در حضور یک گاز ۳ که غیر قابل حل شدن در محلول است، ضریب فعالیت جزء ۱ در فشار کل p چنین تعریف می‌شود: [۴]

$$\mu_1^l(T,p)-\mu_1^{\text{perf.}}(T,p)=RT\ln\gamma_1(T,p) \qquad (۵-۱)$$

که $\mu_1^l(T,p)$، پتانسیل شیمیایی جزء ۱ در محلول حقیقی و $\mu_1^{\text{perf.}}(T,p)$، پتانسیل شیمیایی جزء ۱ در محلول کامل است و به این صورت تعریف می‌شود:

$$\mu_1^{\text{perf.}}(T,p)=\mu_1^{\circ,l}(T,p)+RT\ln x_1^l \qquad (۶-۱)$$

که $\mu_1^{\circ,l}(T,p)$، پتانسیل شیمیایی مایع خالص ۱ در فشار p می‌باشد. اگر فشار استاندارد $p^\dagger$ انتخاب شود، $\mu_1^{\circ,l}(T,p)$ می‌تواند این طور نوشته شود:

$$\mu_1^{\circ,l}(T,p)=\mu_1^{\circ,l}(T,p^\dagger)+(p-p^\dagger)v_1^\circ \qquad (۷-۱)$$

که در آن فرض شده است که v_1°، حجم مولی خالص جزء ۱، با فشار تغییر چندانی نمی‌کند. از تلفیق روابط (۱-۶) و (۱-۷) و سپس جایگذاری نتیجه‌ی حاصل در رابطه‌ی (۱-۵)، می‌توان عبارت زیر را برای $\mu_1^l(T,p)$ به دست آورد:

$$\mu_1^l(T,p)=\mu_1^{\circ,l}(T,p^\dagger)+(p-p^\dagger)v_1^\circ+RT\ln\gamma_1(T,p)\cdot x_1^l \qquad (۸-۱)$$

اگر فشار جزئی، جزء ۱ در فاز بخار p_1 باشد، نتیجه می‌شود:

$$\mu_1^g(T,p)=\mu_1^{\dagger,g}(T)+RT\ln\left(\frac{p_1}{p^\dagger}\right)+p\left\{B_{11}-(1-x_1^g)^2(B_{11}-2B_{13}+B_{33})\right\} \qquad (۹-۱)$$

معادله‌ی (۱-۹) از این فرض نتیجه می‌شود که معادله‌ی حالت یک مخلوط گازی که به طور جزئی ناکامل

است می‌تواند به این صورت نوشته شود:

$$p = \frac{n^g RT}{V} + \frac{RT}{V^2}\sum_{ij} B_{ij} x_i x_j \qquad (۱۰-۱)$$

این تعریف برای ضریب دوم ویریال، B_{ij}، بر این فرض دلالت دارد که نیروهای بین مولکولی عمدتاً نیرو-های پراکندگی لاندن هستند و مولکول‌ها دارای تقارن کروی هستند. فرضی که در واقع درست نیست.

با یکسان فرض کردن پتانسیل شیمیایی جزء ۱ در حالت‌های مایع و گازی و بازآرایی آن عبارت زیر به دست می‌آید:

$$\mu_1^{\dagger,g}(T) - \mu_1^{\dagger,l}(T) = (p - p^{\dagger})v_1^{\circ} + RT\ln\gamma_1(T,p)\cdot x_1^l - RT\ln\left(\frac{p_1}{p^{\dagger}}\right) + p\left\{B_{11} - (1-x_1^g)^2(B_{11} - 2B_{13} + B_{33})\right\} \qquad (۱۱-۱)$$

با در نظر گرفتن این شرط که اگر $x_1^l \to 1$، آنگاه $\gamma_1^l \to 1$ و این که اگر غلظت گاز بی‌اثر به صفر کاهش یابد $x_1^g \to 1$ و $p \to p_1^{\circ}$، به دست می‌آید:

$$\mu_1^{\dagger,g}(T) - \mu_1^{\dagger,l}(T) = (p_1^{\circ} - p^{\dagger})v_1^{\circ} - RT\ln\left(\frac{p_1^{\circ}}{p^{\dagger}}\right) + p_1^{\circ}B_{11} \qquad (۱۲-۱)$$

با جایگذاری معادله‌ی (۱-۱۲) در (۱-۱۱) می‌توان به معادله‌ی (۱-۱۳) رسید.

$$RT\ln\gamma_1(T,p) = RT\ln\left(\frac{p_1}{p_1^{\circ}x_1^l}\right) + p_1^{\circ}(v_1^{\circ} - B_{11}) + p\left\{(B_{11} - v_1^{\circ}) - (1-x_1^g)^2(B_{11} - 2B_{13} + B_{33})\right\} \qquad (۱۳-۱)$$

ضریب فعالیت در فشار استاندارد $p^{\dagger}$ با رابطه‌ی زیر به فشار p مرتبط است:

$$RT\ln\gamma_1(T,p) = RT\ln\gamma_1(T,p^{\dagger}) + (v_1 - v_1^{\circ})(p - p^{\dagger}) \qquad (۱۴-۱)$$

که v_1 حجم مولی جزئی جزء ۱ در محلول است. با جانشینی معادله‌ی (۱-۱۳) در این معادله و بازآرایی آن به دست می‌آید:

$$RT\ln\gamma_1(T,p^{\dagger}) = RT\ln\left(\frac{p_1}{p_1^{\circ}x_1^{l}}\right) + p_1^{\circ}(v_1^{\circ} - B_{11}) + p\Big\{(B_{11} - v_1^{\circ}) - (1-x_1^{g})^2(B_{11} - 2B_{13} + B_{33})\Big\} + (v_1^{\circ} - v_1)(p - p^{\dagger}) \qquad (۱۵-۱)$$

برای به دست آوردن ضریب فعالیت در رقت بی‌نهایت، γ_1^{∞}، با استفاده از G.L.C، $p^{\dagger}$ صفر انتخاب می‌شود در نتیجه معادله‌ی (۱-۱۶) به این صورت در می‌آید:

$$\ln\gamma_1^{\infty}(T,0) = \ln\gamma_1^{\infty,*} - \frac{(B_{11} - v_1^{\circ})p_1^{\circ}}{RT} + \frac{(2B_{13} - B_{33} - v_1^{\infty})p}{RT} \qquad (۱۶-۱)$$

که $\gamma_1^{\infty,*} = \lim_{x_1^{l}\to 0}(p_1 / p_1^{\circ}x_1^{l})$ و v_1^{∞} حجم مولی جزیی جزء ۱ در رقت بی‌نهایت می‌باشد. رابطه‌ی فوق نشان می‌دهد که فقط جملات اول و سوم در سمت راست معادله‌ی فوق به فشار وابسته هستند. p_1° فشار بخار خالص جزء ۱ در غیاب هر گاز بی‌اثر است بنابراین، p_1° مستقل از p می‌باشد.

رفتار ستون GLC در اندازه‌های نمونه‌های کوچک توسط ضریب توزیع تعیین می‌شود، که چنین تعریف می‌شود:

$$k = \lim_{x_1^{l}\to 0}\frac{n_1^{l}}{V^{l}}\frac{V^{g}}{n_1^{g}} \qquad (۱۷-۱)$$

که n_1^{l} و n_1^{g} به ترتیب تعداد مول‌های جزء ۱ در حجم‌های V^{l} و V^{g} از مایع و گاز هستند. با توجه به این که $n_1^{g} = x_1^{g}n^{g}$، که n^{g} تعداد کل مول‌های گازی اشغال کننده‌ی حجم V^{g} است و رابطه‌ی مشابهی نیز برای فاز مایع وجود دارد می‌توان رابطه‌ی فوق را به این صورت نوشت:

$$k = \lim_{x_1^{l}\to 0}\left[\frac{x_1^{l}n^{l}}{V^{l}}\frac{V^{g}}{x_1^{g}n^{g}}\right] = \lim_{x_1^{l}\to 0}\left[\frac{x_1^{l}}{x_1^{g}}\frac{n^{l}v^{g}}{V^{l}}\right] \qquad (۱۸-۱)$$

v^{g}حجم مولی فاز گازی است.

در حد رقت بی‌نهایت جزء ۱ در فاز گازی، از معادله‌ی (۱-۱۰) نتیجه می‌شود

$$v^{g} = \left(\frac{RT}{p}\right) + B_{33} \qquad (۱۹-۱)$$

با استفاده از این معادله در (۱-۱۸) به دست می‌آید:

$$k = \lim_{x_1^l \to 0} \left\{ \frac{x_1^l n^l}{x_1^g V^l} \left[\frac{RT}{p} + B_{33} \right] \right\} \qquad (۱-۲۰)$$

یا

$$\lim_{x_1^l \to 0} \left[\frac{x_1^g p}{x_1^l p_1^\circ} \right] = \gamma_1^{\infty,*} = \frac{n^l RT}{kV^l p_1^\circ} \left[1 + \frac{B_{33} p}{RT} \right] \qquad (۱-۲۱)$$

که نشان می‌دهد k تابعی از فشار کل p می‌باشد.

بنابراین، با جایگذاری (۱-۲۱) در (۱-۱۶) و با تقریب زدن $\ln(1+\frac{B_{33}p}{RT})$ به $(\frac{B_{33}p}{RT})$،به دست می‌آید:

$$\ln \gamma_1^\infty(T,0) = \ln \frac{n^l RT}{kV^l p_1^\circ} - \frac{(B_{11} - v_1^\circ) p_1^\circ}{RT} + \frac{(2B_{13} - v_1^\infty) p}{RT} \qquad (۱-۲۲)$$

سپس دستی[۱] و همکارانش، با مساوی قرار دادن p در رابطه‌ی فوق با فشار متوسط ستون، $\bar{p}$، رابطه‌ی فوق را برای نتایج GLC به کار بردند. $\bar{p}$، را چنین تعریف می‌شود:

$$\bar{p} = \frac{2}{3} p_\circ \left[\frac{\left(\frac{p_i}{p_\circ} \right)^3 - 1}{\left(\frac{p_i}{p_\circ} \right)^2 - 1} \right] = J_2^3 p_\circ \qquad (۱-۲۳)$$

که p_i، و $p_\circ$، فشارهای ورودی و خروجی ستون هستند و J_2^3 عدد محضی است که برای تصحیح فشار به کار می‌رود. با استفاده از $p = \bar{p}$ در رابطه‌ی (۱-۱۸) به دست می‌آید:

$$\ln \gamma_1^\infty(T,0) = \ln \frac{n^l RT}{kV^l p_1^\circ} - \frac{(B_{11} - v_1^\circ) p_1^\circ}{RT} + \frac{(2B_{13} - v_1^\infty) J_2^3 p_\circ}{RT} \qquad (۱-۲۴)$$

همچنین آنها، کمیت kV^l که حجم خالص باقیمانده‌ی حل‌شونده را نشان می‌داد با V_N نشان دادند و توانستند این کمیت را با استفاده از داده‌های حاصل از GLC به صورت زیر به دست آورند:

۱-Desty

$$V_N = (J_2^3)^{-1} u_0 (t_R - t_G) \qquad (۲۵-۱)$$

۱-۳- مروری بر پیش‌بینی ضریب فعالیت در رقت بی‌نهایت

۱-۳-۱- پیش‌بینی توسط QSPR

برای اولین بار در سال ۲۰۰۴ ادوارد ماجین[۱] و همکارانش یک مطالعه‌ی رابطه‌ی کمی ساختار خاصیت (QSPR) را برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت روی ۳۸ ترکیب آلی به عنوان حل‌شونده در سه مایع یونی مختلف در دمای ۲۹۸ K انجام دادند. آنها توانستند برای $\ln \gamma_{13}^{\infty}$ در هر مایع یونی یک رابطه‌ی همبستگی ارائه کنند که مقدار ضریب تعیین[۲]، R^2، گزارش شده توسط آنها برای سه مایع یونی [emim][NTF_2]، [emmim][NTF_2]و [bmpyr][BF_4] به ترتیب ۰/۹۷۰، ۰/۹۷۵، و ۰/۹۵۲ بود [۵].

در سال ۲۰۰۷ جیکین[۳] و همکارانش ضریب فعالیت در رقت بی‌نهایت حل‌شونده‌های هیدروکربنی در چهار مایع یونی مختلف بر پایه‌ی ایمیدازولیوم در دمای ۲۹۸ تا ۳۱۸ کلوین مورد بررسی قرار دادند. در این تحقیق از ۴ توصیف‌کننده که با استفاده از شیوه‌ی نیمه تجربی کوانتومی PM3[۴] محاسبه شده‌بود، استفاده گردید. سپس مدل‌سازی با استفاده از رگرسیون خطی برای هر مایع یونی به طور جداگانه انجام گرفت که ضریب تعیین، R^2، در محدوده ۰/۹۷تا ۰/۹۹ قرار گرفت. نتایج نشان دادند که مقادیر پیش‌بینی شده توافق خوبی با مقادیر تجربی دارند [۶].

در سال ۲۰۱۰، هیوجان سان[۵] و همکارانش ۳۹ ترکیب حل‌شونده از ترکیبات آلی در مایع یونی

۱- Maginn
۲- Determination coefficient
۳- Jiqin
۴- Parameterized Model number 3
۵- Hui zhang Sun

تری‌هگزیل تترا (دسیل) فسفونیوم بیس (تری فلوئورو متیل سولفونیل) ایمید را در سه دمای مختلف با استفاده از روش ارتباط کمی ساختار-ویژگی برای پیش‌بینی ضریب فعالیت در رقت بینهایت مورد بررسی قرار دادند. آنها از الگوریتم ژنتیک (GA)[1] برای انتخاب بهترین توصیف‌کننده‌ها و از روش خطی حداقل مربعات متداول (OLS)[2] برای ساخت مدل استفاده کردند که ضریب تعیین مدل ساخته شده ۰/۹۵۲ و ضریب تعیین برای رد مرحله‌ای تک‌تک داده‌ها ۰/۹۴۵ گزارش شد [۷].

دیهیمی[3] و نامی[4] در سال ۲۰۱۰ برای اولین بار با استفاده از روش شبکه‌ی عصبی مصنوعی (ANN) توانستند ضریب فعالیت در رقت بی‌نهایت را برای ۲۴ ترکیب آلی در ۱۶ مایع یونی متداول حاوی ایمیدازولیوم، در دماهای مختلف از ۲۹۸ تا ۳۶۳ کلوین پیش‌بینی کنند [۸]. شبکه‌ی مورد استفاده‌ی آنها یک شبکه‌ی پیش رونده چند لایه به همراه الگوریتم بهینه سازی لونبرگ-مارکوات و تابع انتقال سیگموئیدی بود. نتایج نشان داد توافق خوبی بین مقادیر پیش‌بینی شده با ANN و مقادیر تجربی وجود دارد. جذر میانگین مربع خطا (RMSE) و ضریب تعیین برای سری تست به ترتیب ۰/۱۲۸ و ۰/۹۹۴ گزارش شد.

۱-۴- هدف تحقیق

اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت ترکیبات آلی به طور معمول با استفاده از تکنیک کروماتوگرافی گاز-مایع انجام می‌شود، وجود ستون‌های کروماتوگرافی بسیاری از مسائل مربوط به جداسازی و شناسائی ترکیبات را حل نموده‌است. اما در کنار این مزایا، تکنیک‌های کروماتوگرافی، محدودیت‌هایی نیز دارند که از آن جمله می‌توان به عدم توانایی آن‌ها در اندازه‌گیری پاره‌ای از ترکیبات

۲- Genetic Algorithm
۳- Ordinary Least Squares
۴- Deyhimi
۵- Nami

که تهیه و یا آماده‌سازی آن‌ها بسیار مشکل است یا ترکیباتی که در ستون‌های کروماتوگرافی ناپایدارند اشاره نمود. همچنین تکنیک‌های کروماتوگرافی نیاز به آماده‌سازی‌ها، انجام عملیات ویژه بر روی نمونه و ستون‌های کروماتوگرافی و در مواردی بهینه‌سازی پارامترها دارند که ممکن است با صرف زمان و هزینه‌های بسیار زیاد همراه باشند. بنابراین این محدودیت‌ها موجب می‌شوند که در کنار روش‌های آزمایشگاهی تجربی، روش‌های تئوری نیز برای استفاده از نتایج تجربی توسعه یابند. از جمله‌ی این روش‌ها مطالعات کمومتریکس به ویژه در زمینه‌ی ارتباط کمی ساختار- خاصیت[۱] (QSPR) می‌باشد. در این مطالعات می‌توان از تکنیک‌های مختلفی برای ایجاد روابط خطی و غیرخطی بین خصوصیات ساختاری و ضریب فعالیت در رقت بی‌نهایت تجربی ترکیبات استفاده نمود. هدف اصلی QSPR به دست آوردن رابطه‌ی کمی بهینه‌ای از روی ساختار مولکولی برای پیش‌بینی خواص آن‌ها می‌باشد و هنگامی که رابطه‌ی معتبری به دست آمد، امکان پیش‌بینی ضریب فعالیت در رقت بی‌نهایت برای ساختارهای مشابه با ترکیبات اندازه‌گیری شده یا ساختارهای دیگر که هنوز اندازه‌گیری نشده وجود دارد

در این پایان نامه هدف، پیش‌بینی γ_{13}^{∞} برای ۵۸ ترکیب آلی و آب در محیط مایع یونی، ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید [BMPYR][TCM]، در بازه‌ی دمایی ۳۱۸/۱۵ K تا ۳۶۸/۱۵ با استفاده از روش‌های غیر خطی است. مایع یونی به کار برده شده در این پروژه دارای قدرت انتخاب بسیار بالایی برای جدا کردن هیدروکربن‌های آلیفاتیک از آروماتیک مخصوصاً هپتان از تیوفن در فرآیند گوگردزدائی از سوخت‌ها می‌باشد [۱].

۱- Quntitative Structure- property Relationship

فصل دوم

کمومتریکس

۲-۱- کمومتریکس[۱]

اصطلاح کمومتریکس اولین بار توسط اسوانت ولد[۲] که در زمینه‌ی شیمی فیزیک آلی فعالیت داشت، مطرح گردید. همکاری ولد با بروس آر. کووالسکی[۳]، منجر به تأسیس انجمن بین المللی کمومتریکس[۴] (ICS) در سال ۱۹۷۴ گردید [۹]. بنا به تعریف ICS، کمومتریکس عبارت است از کاربرد روش‌های ریاضی و آماری برای برقراری ارتباط بین سنجش‌های انجام شده روی یک سیستم یا فرایند شیمیایی با ساختار شیمیایی به منظور درک بهتر اطلاعات شیمیایی [۱۰].

با توجه به رشد سریع تجهیزات مورد استفاده در شیمی و حجم بسیار زیاد داده‌ها، هنگام پردازش، تفسیر اطلاعات و استخراج نتایج مفید از آنها، نیاز به کامپیوتر امری اجتناب ناپذیر است. از طرف دیگر شیمیدانان، گاهی اوقات با موادی سر و کار دارند که بسیار گران، سمی و خطرناک بوده یا در مواردی به راحتی قابل دسترس نیستند. در این موارد می‌توان از روش‌های ریاضی و آمار جهت توصیف و توجیه نتایج آزمایش‌های مختلف استفاده نمود [۱۱]. می‌توان گفت که موارد فوق مهمترین دلایل افزایش کاربرد روش‌های کمومتریکس نزد شیمیدانان است.

۲-۲- ارتباط کمی ساختار-خاصیت یا ساختار-فعالیت[۵]

از نظر شیمیدانان فعالیت و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده‌ی مولکول انجام گیرد، به آن ارتباط

۱- Chemometrics

۲- Svante Wold

۳- Bruce R.Kowaski

۴- International Chemometrics Society

۵- Quntitative Structure-Activity Relationship

کمی ساختار- خاصیت (QSPR) می‌گویند مانند یافتن رابطه‌ای بین ساختار مولکول‌ها با خواصی نظیر نقطه‌ی جوش و ذوب، فشار بخار، حلالیت، اندیس بازداری و غیره ولی مطالعات ارتباط کمی ساختار-فعالیت (QSAR) مطالعاتی است که فعالیت بیولوژیکی ترکیبات را به ویژگی‌های ساختاری آنها ارتباط می‌دهد. در این مطالعات سعی بر این است تا رابطه‌ای هماهنگ میان فعالیت‌های شیمیایی و فیزیکی با ویژگی‌های مولکولی پیدا شود، به گونه‌ای که بتوان این قواعد را برای ارزیابی فعالیت ترکیبات جدید به کار برد [۱۲].

از روش‌هایی که به منظور مطالعه‌ی ارتباط خطی ساختار خاصیت مورد استفاده قرار می‌گیرند می‌توان به روش‌های خطی مثل حداقل مربعات متداول[۱] (OLS)، تحلیل مؤلفه‌های اساسی[۲] (PCA)، رگرسیون خطی چندگانه[۳](MLR)، رگرسیون اجزای اصلی[۴] (PCR) و حداقل مربعات جزئی[۵] (PLS) اشاره نمود. روش‌های دیگری مانند شبکه‌ی عصبی[۶](ANN)، ماشین‌های بردار پشتیبان[۷](SVM) و الگوریتم ژنتیک[۸] (GA)، ارتباط غیر خطی میان ساختار و خواص ترکیبات را مورد مطالعه قرار می‌دهند. هر مطالعه‌ی QSPR یا QSAR شامل مراحل زیر می‌باشد: [۱۳]

۱- جمع آوری سری داده‌ها

۲- رسم و بهینه‌سازی ساختار هندسی ترکیبات

۳- محاسبه‌ی توصیف‌کننده‌ها

۴- حذف توصیف‌کننده‌های نامناسب

۵- دسته‌بندی داده‌ها

۱-Ordinary Least Squares
۲- Principle Component Analysis
۳-Multiple Linear Regression
٤-Principal Component Regression
٥-Partial Least Squares
٦-Artificial Neural Networks
۷-Support Vector Machines
۸-Genetic Algorithm

۶- انتخاب بهترین توصیف‌کننده‌ها

۷- ساختن مدل

۸- ارزیابی قدرت مدل

۹- پیش‌بینی خاصیت مورد نظر

۲-۲-۱- جمع آوری سری داده‌ها

اولین مرحله در توسعه‌ی هر مدل QSAR یا QSPR گرد آوری مجموعه‌ای از ترکیبات است که خاصیت مورد مطالعه برای آنها در شرایط عملی یکسانی به دست آمده باشد. هر چه داده‌ها معتبرتر و سری داده‌ها وسیع‌تر باشد، مدل به دست آمده کارایی بیشتری خواهد داشت.

۲-۲-۲- رسم و بهینه‌سازی ساختار هندسی ترکیبات

ایجاد توصیف‌کننده‌های هندسی و هیبریدی بر مبنای ساختار و هندسه‌ی دقیق مولکولی استوار است و اگر ساختارها به صورت صورتبندی[۱] با حداقل انرژی نباشند، مقادیر غیر صحیحی برای این توصیف‌کننده‌ها ایجاد می‌شود [۱۴]. بنابراین، ساختار هندسی ترکیبات مورد مطالعه با استفاده از روشهای شیمی محاسباتی در نرم افزارهای مختلفی مثل Gaussian و Hyperchem ، بهینه[۲] می‌گردد. شیمی محاسباتی ساختارهای مولکولی را به صورت پارامترهای عددی معرفی و رفتارهای آنها را با معادلات کوانتومی و فیزیک کلاسیک شبیه سازی می‌نماید.

دو روش مهم در شیمی محاسباتی که برای بهینه‌سازی ساختار مولکول‌ها مورد استفاده قرارمی‌گیرد روش مکانیک مولکولی و روش مکانیک کوانتومی است [۱۵]. در روش مکانیک مولکولی که بر مبنای

۱- Conformation

۲- Optimize

روابط کلاسیک بنا شده است ساختمان مولکولی به صورت گوی و فنر در نظر گرفته می‌شود. در این روش انرژی مولکول به صورت مجموعه‌ای از انرژی‌های کششی، خمشی، الکتروستاتیکی و....بیان می‌شود. سپس می‌توان طول پیوندها، زوایای پیوندی و صورت‌بندی را به گونه‌ای تغییر داد تا ساختاری که عبارت انرژی مکانیک مولکولی را مینیمم کند، پیدا شود. این روش بهینه‌سازی، بیشتر برای ماکرومولکول‌ها به کار می‌رود [۱۶].

در بهینه‌سازی ساختار مولکولی به روش مکانیک کوانتومی از معادله‌ی شرودینگر[۱] برای محاسبه‌ی انرژی مولکول استفاده می‌شود. این روش نسبت به روش مکانیک مولکولی دارای صحت بیشتری است زیرا برخلاف مکانیک مولکولی که الکترون‌ها را در نظر نمی‌گرفت اثرات الکترونی اعمال شده روی مولکول نیز در محاسبات وارد می‌شود. اما در نظر گرفتن اثرات الکترونی در روش مکانیک کوانتومی سبب می‌شود که محاسبات پیچیده‌تر شده و وقت بیشتری را طلب کند.

روش‌های محاسباتی بر پایه‌ی مکانیک کوانتومی شامل روش‌های آغازین[۲] و روش‌های نیمه تجربی[۳] است [۱۵]. در روش‌های آغازین از داده‌های تجربی استفاده نمی‌شود بلکه ثابت‌هایی مانند جرم الکترون، ثابت پلانک، ثابت سرعت نور و بار الکترون به کار گرفته می‌شوند و محاسبه‌ی انرژی مولکول‌ها از حل معادله‌ی شرودینگر با استفاده از یک سری تقریب به دست می‌آید. در این روش چون کل الکترون‌ها را در محاسبات وارد می‌کنند، محاسبات پیچیده‌تر شده و به وقت بیشتری نیاز است [۱۷ و ۱۵].

روش‌های نیمه تجربی، روش‌هایی هستند که فقط الکترون‌های لایه ظرفیت را در محاسبات وارد می‌کنند. در نتیجه زمان محاسبات در این روش‌ها، کوتاهتر از روش‌های آغازین است و می‌تواند برای مولکول‌های بزرگتر به کار رود [۱۸]. روش‌های نیمه تجربی مختلفی وجود دارد که در این تحقیق،

۱- Schroodinger Equation

۲- Ab Initio

۳- Semi empirical

ساختار مولکول‌ها پس از ترسیم در نرم‌افزار Hyperchem ، با استفاده از روش نیمه تجربی AM1[1] دراین نرم‌افزار بهینه گردید [۱۵].

۲-۲-۳- محاسبه ی توصیف‌کننده‌ها

توصیف‌کننده‌های مولکولی نتیجه‌ی نهایی یک فرایند منطقی و ریاضی هستند که اطلاعات شیمیایی مربوط به ساختار مولکول را به اعداد تبدیل می‌کنند. این اعداد می‌توانند برای تفسیر خواص مولکولی استفاده شوند و یا برای پیش‌بینی تعدادی از ویژگی‌های مولکولی در یک مدل شرکت کنند.

برخی از ویژگی‌های یک توصیف‌کننده‌ی مناسب عبارت است از [۱۹]:

- ساده بودن
- توانایی تفسیر ساختار مولکول
- عدم همبستگی با سایر توصیف‌کننده‌ها
- قابلیت تمایز بین ایزومرهای مختلف یک مولکول
- قابلیت کاربرد برای دامنه‌ی وسیعی از ساختارهای مولکولی

توصیف‌کننده‌های مولکولی به دو دسته اصلی تقسیم‌بندی می‌شوند: توصیف‌کننده‌های حاصل از اندازه‌گیری‌های تجربی (مانند قطبش پذیری و ممان دوقطبی) و توصیف‌کننده‌های تئوری که با استفاده از ساختار مولکول و بدون نیاز به داده‌های تجربی محاسبه می‌شوند که باعث صرفه جویی در وقت، هزینه، مواد و تجهیزات خواهند شد و برای هر مولکول واقعی یا فرضی که هنوز سنتز نشده‌اند، در دسترس می‌باشند [۲۰ و ۱۴]. در این تحقیق ۱۴۸۱ توصیف‌کننده‌ی تئوری توسط نرم افزار دراگون[2] محاسبه شد که این توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به ۱۸ دسته‌ی متفاوت تقسیم

۱- Austian Method 1
۲- Dragon

می‌شوند[۲۱].

۲-۲-۴- حذف توصیف‌کننده‌های نامناسب

در این مرحله باید توصیف‌کننده‌هایی که اطلاعات کمتری در مورد تغییرات خواص شیمیایی و یا بیولوژیکی مواد مورد نظر دارند، حذف شوند. برای این منظور لازم است که از فرآیند کاهش متغیر استفاده شود. زیاد بودن تعداد توصیف‌کننده‌ها و همچنین همبستگی بالای برخی از آنها به یکدیگر، باعث طولانی و دشوار شدن فرایند مدل‌سازی می‌شود. در فرآیند کاهش متغیر توصیف‌کننده‌هایی حذف می‌شوند که معمولاً یک یا چند ویژگی زیر را داشته باشند:

۱- توصیف‌کننده‌هایی که دارای بیش از ۹۰٪ مقادیر یکسان باشند.

۲- توصیف‌کننده‌هایی که با توصیف‌کننده‌هایی دیگر همبستگی بیش از ۰/۹ دارند. برای این منظور می‌توان از نرم افزار SPSS استفاده کرد. اما انجام این کار با استفاده از این نرم افزار کاری بسیار وقت‌گیر است. اخیراً با استفاده از برنامه‌ی نوشته شده در محیط نرم افزار MATLAB، همبستگی توصیف‌کننده‌ها بررسی شده و توصیف‌کننده‌های حذفی مشخص می‌گردند.

۳- توصیف‌کننده‌هایی که با متغیر وابسته، همبستگی کمی داشته و یا ارتباط آنها با خاصیت مورد نظر مشکل است.

پس از حذف توصیف‌کننده‌های نامناسب، تعداد توصیف‌کننده‌ها و به دنبال آن پیچیدگی محاسبات به میزان زیادی کاهش می‌یابد.

۲-۲-۵- دسته بندی داده‌ها

قبل از ساختن مدل، سری داده‌ها به سه زیر مجموعه تقسیم می‌شوند. اولین زیر مجموعه، سری آموزش نام دارد که ازآن برای ساخت مدل‌های خطی و غیر خطی استفاده می‌شود. دومین زیر مجموعه سری ارزیابی است که مدل‌های به دست آمده از سری آموزش توسط آن ارزیابی شده و به این ترتیب بهترین مدل انتخاب می‌شود. زیر مجموعه‌ی سوم سری تست یا آزمایش نام دارد که در طول فرآیند مدل‌سازی دخالتی نداشته و از آن برای مقایسه‌ی مدل‌های متفاوت استفاده می‌شود. تقسیم بندی این سه زیر مجموعه معمولاً به صورت تصادفی صورت گرفته می‌گیرد.

۲-۲-۶- انتخاب بهترین توصیف‌کننده‌ها

با وجودی که با به کار بردن روش‌های کاهش متغیر، تعداد توصیف‌کننده‌ها به میزان زیادی کاهش می‌یابد. اما هنوز تعداد آنها برای ساخت یک مدل مناسب زیاد است. بنابراین باید به ترتیبی، بهترین توصیف‌کننده‌ها را از میان توصیف‌کننده‌های باقیمانده، انتخاب نمود. از جمله این روش‌ها می‌توان به روش رگرسیون گام به گام یا مرحله‌ای[۱] و الگوریتم ژنتیک اشاره کرد. در ادامه به اختصار راجع به هر کدام از این روش‌ها توضیحاتی ارائه می‌گردد.

۲-۲-۶-۱- انتخاب بهترین توصیف‌کننده‌ها توسط رگرسیون مرحله‌ای(SR)

در این روش خصوصیت شیمیایی و یا بیولوژیکی مورد نظر به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیرهای مستقل در نظر گرفته می‌شوند. در روش مرحله‌ای، ابتدا با توجه به ضریب تعیین، متغیر اولیه انتخاب می‌شود. سپس مدل دو متغیری ایجاد شده و پارامترهای آماری مدل حاصل برای

۱- Stepwise Regression

سری ارزیابی مورد بررسی قرار می‌گیرند. در صورتی که مدل از نظر آماری بی معنا باشد، متغیر به کار گرفته شده در آن مدل حذف می‌گردد. سپس مدل سه متغیری ساخته می‌شود و دوباره پارامترهای آماری آن شامل ضریب تعیین، مقدار میانگین مربعات خطا و آماره‌ی F برای سری ارزیابی مورد بررسی قرار می‌گیرد. مدلی که ضریب تعیین و آماره‌ی F آن بیشتر و مقدار میانگین مربعات خطای آن کمتر باشد به عنوان بهترین مدل انتخاب شده و توصیف‌کننده‌های موجود در مدل به عنوان توصیف‌کننده‌های بهینه به روش رگرسیون مرحله‌ای در نظر گرفته می‌شود.

۲-۲-۶-۲- انتخاب بهترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک (GA)

۲-۲-۶-۲-الف- مقدمه

امروزه با پیشرفت سریع دانش، بهینه‌سازی از اهمیت بالایی در علوم مختلف مهندسی برخوردار شده است. بهینه‌سازی یکی از مفاهیمی است که به خوبی از پل ارتباطی بین تئوری و عمل عبور کرده و دارای کاربرد گسترده‌ای می‌باشد. یکی از ابزارهای مناسب برای بهینه‌سازی مسائل چند هدفه، الگوریتم ژنتیک است که شاخه‌ای از محاسبات تکاملی محسوب می‌گردد. مبانی این روش از طبیعت الهام گرفته شده است.

در سال ۱۹۷۵ جان هولند[۱] نظریه‌ی بنیادی ژنتیک را مطرح کرد. ایده‌ی اساسی این الگوریتم، انتقال خصوصیات موروثی توسط ژن‌هاست. فرض کنید مجموعه‌ی خصوصیات انسان توسط کروموزوم‌های او به نسل بعدی منتقل می‌شوند. هر ژن در این کروموزوم‌ها نماینده‌ی یک خصوصیت است. حال اگر این کروموزوم کاملاً، به نسل بعد انتقال یابد، تمامی خصوصیات نسل بعدی شبیه به خصوصیات نسل قبل خواهد بود. بدیهی است که در عمل چنین رخ نمی‌دهد زیرا کروموزوم‌ها دستخوش تغییراتی از قبیل

۱- John Holand

جهش و تقاطع می‌شوند[۲۲].

۲-۲-۶-۲-ب- الگوریتم ژنتیک و ساختار آن

الگوریتم‌های ژنتیک (GA) یک تکنیک جستجو در علم رایانه برای یافتن راه حل بهینه و مسائل جستجو هستند. الگوریتم‌های ژنتیک یکی از انواع الگوریتم‌های تکاملی هستند که از علم زیست شناسی مانند وراثت، جهش، انتخاب ناگهانی، انتخاب طبیعی و ترکیب الهام گرفته‌اند.

ساز وکار الگوریتم ژنتیک بر مبنای قوانین موجود در علم ژنتیک است. در این بخش با برخی مفاهیم و تعاریف ژنتیک آشنا خواهیم شد:

الف)- افراد[۱] یا کروموزوم‌ها[۲] : هر کدام از افراد جمعیت، که تقریب‌هایی از جواب نهایی‌اند، به صورت رشته‌هایی از حروف یا ارقام، کدگذاری می شوند. این رشته ها را کروموزوم می نامند. متداولترین حالت، نمایش با ارقام صفر و یک است [۲۴ و ۲۳].

ب)- جمعیت: به تعداد مشخصی از کروموزوم‌ها که در کنار هم قرار می‌گیرند جمعیت گفته می‌شود و با نماد N نشان داده می‌شود. هر جمعیت توسط عملگرهای وراثتی مثل انتخاب، تلفیق و جهش جمعیت بعدی را می‌سازد.

پ)- نمایش و کد گذاری کروموزوم‌ها: تاکنون در حل مسائل بهینه‌سازی روش‌های مختلفی برای نمایش پارامترها و اطلاعات مسأله (کروموزوم‌ها) به کار برده شده است که انتخاب هر کدام از این روش‌ها باید با توجه به نوع مسأله و فضای جستجوی مورد نیاز برای حل و بهینه‌سازی آن صورت پذیرد. از پرکاربردترین روش‌ها، کدگذاری رشته‌ای است.

ت)- تعیین جمعیت اولیه: بعد از تصمیم‌گیری در مورد شیوه‌ی کدگذاری کروموزوم‌ها، جمعیت

۱- Individuals
۲- Chromosome

اولیه باید اتخاذ گردد. این مرحله معمولاً با انتخاب تصادفی مقادیر در محدوده مجاز صورت می‌گیرد.

ث)- تابع شایستگی[۱] : برتری کروموزوم‌ها (افراد) نسبت به یکدیگر با توجه به یک معیار سنجیده می‌شود. این معیار تابع شایستگی است

ج)- روش‌های انتخاب کروموزوم‌ها: روش‌های مختلفی برای انتخاب کروموزوم‌ها وجود دارند که در زیر به معمولترین آن‌ها اشاره می‌شود.

- انتخاب چرخ رولت[۲]: در این روش هر عنصری که عدد برازش (تناسب) بیشتری داشته باشد، شانس انتخاب بیشتری دارد.
- انتخاب نخبه‌گرا یا ممتازگرا[۳]: در این روش مناسب‌ترین عضو هر اجتماع انتخاب می‌شود.
- انتخاب مقیاسی[۴]: در این روش به موازات افزایش متوسط عدد برازش جامعه، سنگینی انتخاب هم بیشتر می‌شود. این روش وقتی به کار می‌رود که مجموعه دارای عناصری با عدد برازش بزرگ باشد و فقط تفاوت‌های کوچکی آنها را از هم تفکیک کند
- انتخاب مسابقه‌ای[۵]: روشی است که در آن یک زیرمجموعه از صفات یک جامعه، انتخاب می‌شوند و اعضای آن مجموعه با هم رقابت می‌کنند تا سرانجام فقط یک صفت از هر زیر گروه برای ادغام، انتخاب شوند[۲۵].

چ)- روش‌های ادغام کروموزوم‌ها: وقتی با روش‌های انتخاب، کروموزوم‌ها انتخاب شدند، باید به طور تصادفی برای افزایش تناسبشان اصلاح شوند. دو راه حل اساسی برای این کار وجود داردکه اولین و ساده‌ترین آنها جهش[۶] نام دارد. جهش به معنی تغییر از یک ژن به دیگری می‌باشد. در

۱- Fitness function
۲- Roulette wheel
۳- Elitist
٤- Scalin
۵- Tournament
۶- Mutation

الگوریتم‌های ژنتیک، جهش تغییر کوچکی در یک نقطه از کد خصوصیاتی ایجاد می‌کند [۲۵]. در نمایش دودویی[۱] رشته‌ها، جهش به معنای تغییر مقدار یکی از خانه‌های رشته، از یک به صفر و یا از صفر به یک می‌باشد. احتمال جهش در کروموزوم‌ها معمولاً در حدود ۰/۰۱ تا ۰/۰۰۱ در نظر گرفته می‌شود. به کمک این عملگر می‌توان امید داشت که کروموزوم‌های خوبی که در مراحل انتخاب و یا تکثیر حذف شده‌اند، دوباره احیا شوند. این عملگر همچنین تضمین می‌کند که بدون توجه به پراکندگی جمعیت اولیه، احتمال جستجوی هر نقطه از فضای مسأله هیچ گاه صفر نشود [۲۷ و ۲۶]. دومین روش، تقاطع[۲] نام دارد که به معنای چسبیدن دو کروموزوم از طول به یکدیگر و تبادل برخی قطعات بین کروموزوم هاست. این روش اغلب شامل تقاطع تک نقطه ای بوده و نقطه‌ی تعویض در محلی تصادفی بین کروموزوم‌ها قرار دارد [۲۵].

۲-۲-۶-۲-پ- شرایط توقف الگوریتم ژنتیک

از شرایط زیر می‌توان برای توقف روند تکرار الگوریتم ژنتیک استفاده نمود:

- تعداد نسل‌ها: الگوریتم زمانی متوقف می‌شود که تعداد نسل‌ها به مقدار معینی برسد.
- محدودیت زمانی: در جعبه ابزار GA می‌توان این محدودیت را برحسب واحد ثانیه تعیین نمود.
- محدودیت شایستگی: الگوریتم زمانی متوقف می‌شود که بهترین مقدار شایستگی در جمعیت حاضر کوچکتر یا مساوی یک مقدار معین گردد.
- رکود نسلی: اگر در نسل‌ها به تعداد معینی هیچ گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.

۱- Binary
۲- Cross over

- رکود زمانی: در صورتی که در تابع هدف برای زمان معینی هیچ گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.

الگوریتم در صورتی که هر یک از شرایط فوق برآورده گردد متوقف می‌شود.

۲-۲-۷- ساختن مدل

هدف اصلی مدل‌های QSPR این است که بین توصیف‌کننده‌های مولکولی یک ترکیب و خواص فیزیکی یا شیمیایی آن، یک رابطه‌ی کمی برقرار کند.

همان طور که قبلاً در بخش (۲-۲) در مطالعات ارتباط کمی ساختار-خاصیت، مدل‌سازی به دو صورت خطی و یا غیر خطی انجام می‌شود، که به دلیل استفاده از دو روش غیر خطی (ANN) و (SVM) در این پایان نامه، در ادامه فقط راجع به این دو روش توضیح داده خواهد شد.

۲-۲-۷-۱- شبکه‌ی عصبی مصنوعی

شبکه‌های عصبی زیر مجموعه‌ای از هوش مصنوعی هستند. این شبکه‌ها در واقع تقلیدی از سیستم‌های عصبی زیستی هستند و مانند آنها از عناصر عملیاتی ساده‌ای به نام نرون[۱] که به صورت موازی عمل می‌کنند، ساخته شده‌اند. پردازش موازی به معنای اجرای برنامه‌هایی است که در هر بار بیش از یک عمل را انجام می‌دهند. این نوع پردازش، به دو دلیل مورد توجه هوش مصنوعی است. نخست این که به وسیله‌ی پردازش‌های موازی، زمان طولانی مدت استفاده از کامپیوتر برای گشودن تعدادی از مسئله‌های هوشمند ریاضی کاهش می‌یابد. دوم اینکه برخی از مسائل با شیوه‌های موازی بهتر حل می‌شوند.

۱- Neuron

شبکه‌های عصبی مصنوعی که با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای آنها را کشف می‌کنند. به همین دلیل به این سیستم‌ها هوشمند گفته می‌شود زیرا براساس محاسبات روی داده‌های عددی یا مثال‌ها، قوانین کلی را فرا می‌گیرند [۲۸]. برای آشنایی با شبکه‌های عصبی مصنوعی، ابتدا باید سیستم عصبی زیستی را به طور مختصر مورد بررسی قرار داد.

۲-۲-۷-۱-الف: سیستم عصبی زیستی

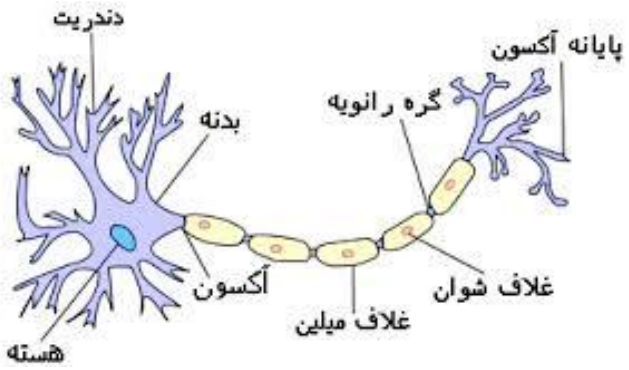
مغز انسان یک سیستم پردازش اطلاعات با ساختار موازی و کاملاً پیچیده است که دو درصد وزن بدن را تشکیل می‌دهد اما بیش از بیست درصد کل اکسیژن بدن را مصرف می‌کند مرکز بسیاری از رفتارهای خودآگاه و ناخودآگاه انسان است. مغز از حدود یک صد میلیارد سلول عصبی به نام نرون تشکیل شده که هر نرون به ده هزار نرون دیگر مرتبط می‌باشند. بیشتر نرون‌ها از سه قسمت اساسی دندریت‌ها[۱]، بدنه‌ی سلول و آکسون[۲] تشکیل شده‌اند [۲۹]:

- دندریت‌ها، به عنوان مناطق دریافت سیگنال الکتریکی، شبکه‌های تشکیل یافته از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بیشماری می‌باشند. به همین علت آنها را شبکه‌های دریافت کننده‌ی درخت مانند نیز می‌نامند. دندریت‌ها سیگنال‌های الکتریکی را به هسته منتقل می‌کنند.
- بدنه سلول، که شامل هسته و قسمت‌های حفاظتی دیگر می‌باشد انجام عملیات لازم بر روی اطلاعات ورودی را بر عهده دارد.
- آکسون‌ها، بر عکس دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتری برخوردار هستند.آنها طول بیشتری دارند و سیگنال الکتروشیمیایی دریافتی از هسته سلول را به

۱- Dendrites
۲- Axon

نرون‌های دیگر منتقل می‌کنند. محل اتصال آکسون یک سلول به دندریت‌های سلول‌های دیگر را سیناپس[1] می‌گویند. پیام‌های عصبی تنها به صورت یک طرفه از دندریت‌ها به بدنه سلول و از آنجا به آکسون حرکت می‌کنند. وقتی پیام به انتهای آکسون می‌رسد، موجب آزاد شدن یک ماده‌ی شیمیایی به نام انتقال دهنده نرونی از انتهای آکسون می‌شود و پس از نفوذ در سیناپس‌ها، گیرنده‌های سلول‌های مجاور را فعال می‌کند [۳۰].

شکل (۲-۱) نشان دهنده‌ی ساختار کلی یک نرون می‌باشد.



شکل(۲-۱)- ساختار یک نرون زیستی [۳۰]

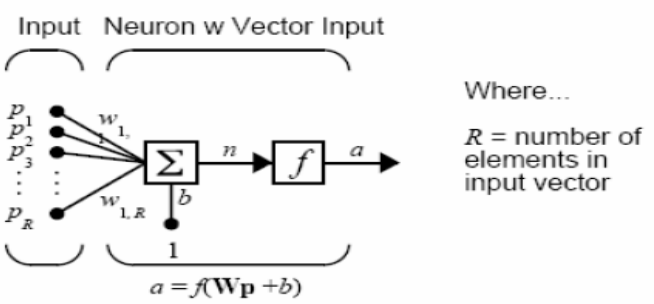
۲-۲-۷-۱-ب-مدل ریاضی نرون

با بحث مختصری که در مورد نرون‌های طبیعی انجام شد، می‌توان فهمید که برای ایجاد هر شبکه‌ی عصبی مصنوعی باید سیستمی مشابه مغز طراحی شود که از تعداد زیادی نورون تشکیل شده باشد نرون در شبکه عصبی مصنوعی یک واحد محاسباتی-پردازشی است که مدل ساده شده‌ای از نرون طبیعی است. هر نرون تعدادی ورودی دارد که پس از وزن‌دهی تک‌تک آنها و سپس محاسبه‌ی مجموع آنها، سیگنال تحریک را ایجاد می‌کند. سپس در مرحله‌ی بعد، تابع انتقال[2] روی سیگنال تحریک اثر کرده و پیغام خروجی نرون به نرون‌های دیگر را به وجود می‌آورد.

۱- Synapse

۲- Transfer function

شکل(۲-۲) الگویی از یک نرون محاسباتی را با توجه به نحوه‌ی عملکرد نرون طبیعی ارائه می‌دهد. با نگاهی به ساختار این نرون محاسباتی، می‌توان ورودی‌ها را به دندریت‌ها، مجموعه‌ی جمع کننده و تابع محرک را به بدنه‌ی سلول، وزن‌ها[۱] را به شدت سیناپس‌ها و خروجی را به سیگنال گذرنده از آکسون تشبیه نمود. همچنین ارتباط بین خروجی و سیگنال تحریک در قسمتی به نام تابع انتقال وجود دارد [۳۲ و ۳۱].



شکل (۲-۲)- ساختار یک نرون محاسباتی با چند ورودی [۱۳]

در شکل (۲-۲) که یک مدل نرون محاسباتی با R ورودی را ارائه می‌دهد. بردار ورودی با **p** نمایش داده می‌شود که اسکالرهای p_i (i=1,2,…R) عناصر بردار **p** هستند و مجموعه‌ی سیناپس‌های $w_{1,j}$، عناصر ماتریس وزن **W** را تشکیل می‌دهند. در این حالت **W** یک بردار سطری با عناصر j=1,…,R و $w_{1,j}$ است. هر عنصر از بردار ورودی **p** درعنصر متناظر بردار **W** ضرب می‌شود. نرون یک جمله بایاس، b، داردکه با حاصلضرب ماتریس وزن **W** در بردار ورودی **p** جمع می‌شود تا ورودی خالص، n، یا همان سیگنال تحریک را ایجاد کند [۳۱].

$$n = \mathbf{Wp} + b = w_{1,1}p_1 + w_{1,2}p_2 + \ldots + w_{1,R}p_R + b \qquad (۱-۲)$$

سپس تابع انتقال روی این ورودی خالص n اثر کرده و خروجی نرون، a، را به وجود می‌آورد.

۱- Weight

$$a=f(n)=f(\ \mathbf{W}\mathbf{p}+b)$$ (۲-۲)

۲-۲-۷-۱-پ- توابع انتقال

سه تابع انتقال رایج در فرایند بهینه‌سازی شبکه های عصبی، تابع انتقال خطی[۱] ، تابع انتقال لگـاریتم سیگموییدِ[۲] و تابع انتقال تانژانت هایپربولیک سیگموییدِ[۳] نام دارند[۳۳].

۱- تابع انتقال خطی: تابع که به اختصار به صورت purline نمایش داده می‌شود یک تابع پیوسـته است که معمولاً برای تقریب خطی در فیلترهای خطی به کار می‌روند. این تابع همـان مقـدار ورودی را بـه عنوان خروجی برمی‌گرداند.

۲- تابع انتقال لگاریتم سیگموییدِ: این تابع که در برنامه‌ی نرم افزاری بـا LOGSIG نمـایش داده می‌شود بیشتر در شبکه‌های پس انتشار استفاده می‌شود. این تابع انتقال مقادیر ورودی را در محـدوده ∞- تا ∞+ دریافت کرده و خروجی بین صفر و یک تولید می‌نماید.

۳- تابع انتقال تانژانت سیگموییدِ: که با TANSIG نمایش داده می‌شود ورودی را در محدوده ∞- تا ∞+ دریافت کرده و خروجی بین ۱ و ۱- تولید می‌نماید.

۲-۲-۷-۱-ت- انواع شبکه‌های عصبی از نظر ارتباط بین نرونی

شبکه‌های عصبی براساس شیوه‌ی پردازش اطلاعات و ارتباطات بین نرونی به دو دسته‌ی کلی تقسـیم می‌شوند: شبکه‌های پیشخور[۴] و شبکه‌های پس خور[۵]. شبکه‌های پیشخور، شبکه‌هـایی هسـتند کـه مسـیر

۱- Linear Transfer Function
۲- Log-Sigmoid Transfer Function
۳- Hyperbolic Tangent Sigmoid Transfer Function
٤- Feed Forward
٥- Feed Back

پاسخ در آنها همواره رو به جلو پردازش شده و به نرون‌های همان لایه یا لایه قبل باز نمی‌گردد. در این نوع شبکه‌ها به سیگنال اجازه داده می‌شود که از مسیر یکطرفه، یعنی از ورودی تا خروجی، عبور کند و در نتیجه بازخوردی[۱] وجود نداشته و خروجی هر لایه تأثیری بر همان لایه و همچنین لایه‌های قبلی ندارد.

در شبکه‌های پس‌خور، حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های همان لایه و یا نرون‌های لایه‌ی قبلی وجود دارد. بنابراین آنها می‌توانند با استفاده از حلقه‌های برگشتی، سیگنال‌هایی داشته باشند که در هر دو مسیر از ورودی به خروجی و بالعکس حرکت کنند [۳۴].

۲-۲-۷-۱-ث- روش‌های آموزش

برای آموزش شبکه‌های عصبی دو روش آموزش با ناظر[۲] و آموزش بدون ناظر[۳] مورد استفاده قرار می‌گیرند. در یادگیری با ناظر، در هر مرحله از تکرار الگوریتم یادگیری، جواب واقعی سیستم یادگیرنده وجود دارد. لذا الگوریتم یادگیری به خطای یادگیری؛ یعنی تفاوت بین مقدار واقعی و مقدار پیش‌بینی شده دسترسی دارد. در اکثر شبکه‌های عصبی از این نوع یادگیری استفاده می‌شود [۳۵ و ۳۴].

نوع دیگری از یادگیری به نام یادگیری بدون ناظر یا خودسامان‌ده می‌باشد که در آن جواب واقعی برای سیستم یادگیرنده موجود نیست و شبکه می‌آموزد که الگوهای ورودی را به تعداد مساوی از گروه‌ها تقسیم‌بندی کرده، ارتباطات موجود بین آن‌ها را پیدا و در خروجی شبکه کد نماید. باید توجه داشت که در این حالت، فرد کاربر است که هدف نهایی را مشخص می‌کند [۳۶].

رفتار سیستم‌های آموزش‌پذیر توسط الگوریتم‌های برگشتی بیان می‌شود که به این الگوریتم‌ها قوانین یادگیری می‌گویند. انواع مختلفی از قوانین یادگیری برای شبکه‌های عصبی وجود دارند که یادگیری

۱- Recurrent
۲- Supervised training
۳- Unsupervised training

عملکردی[1] یکی از آن‌ها است. در این نوع یادگیری، پارامترهای شبکه (وزن‌ها و پیش‌قدرها) به نحوی تنظیم می‌شوند که عملکرد شبکه بهینه شود. منظور از بهینه کردن عملکرد شبکه، حداقل شدن خطایی است که بین مقادیر تجربی و پاسخ شبکه وجود دارد [۲۱].

برای بهینه‌سازی عملکرد شبکه، ابتدا باید یک شاخص عملکرد پیدا کرد. شاخص عملکرد، معیاری برای بیان عملکرد شبکه است و با عملکرد شبکه، رابطه‌ی عکس دارد، یعنی هر چه عملکرد شبکه بهتر باشد، مقدار شاخص عملکرد کوچکتر خواهد بود و بالعکس. شاخص عملکرد در بیشتر الگوریتم‌ها متوسط مربعات خطا (MSE)[2] است. شاخص عملکرد را اصطلاحاً تابع هدف یا تابع خطا نیز می‌گویند، یعنی تابعی که کمینه کردن آن مورد نظر است. پس از تعیین شاخص عملکرد، باید پارامترهای شبکه برای کاهش مقدار این شاخص عملکرد تنظیم گردد [۳۵].

برای بهینه‌سازی عملکرد شبکه، روش‌های مختلفی وجود دارد که از آن جمله می‌توان به الگوریتم‌های آموزشی لونبرگ-مارکوارت و تنظیم بایزین اشاره کرد. در این پایان‌نامه از این دو الگوریتم به منظور بهینه‌سازی عملکرد شبکه استفاده شده است[۳۷].

۲-۲-۷-۲- ماشین برداری پشتیبان (SVM)

ماشین‌های برداری پشتیبان (SVM)[3] یکی از روش‌های یادگیری با ناظر هستند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر مثل شبکه‌های عصبی برای طبقه‌بندی، نشان داده است.

از الگوریتم‌های آموزش دیده توسط ماشین‌های برداری پشتیبان می‌توان برای موارد زیر استفاده

۱- Performance Learning
۲- Mean Square Error
۳- Support Vector Machine

نمود:

۱- طبقه‌بندی برای سیستم‌های خطی و غیرخطی، که این سیستم‌ها ممکن است به صورت دو کلاسه یا چند کلاسه باشند.

۲- آنالیز رگرسیون برای سیستم‌های خطی و غیر خطی

ماشین‌های برداری پشتیبان در مقایسه با شبکه‌های عصبی دارای دو مزیت عمده هستند که همین امر موجب توجه خاص محققین به این ماشین‌های فراگیر شده است:

۱- قابلیت مدل‌سازی توسط SVM با تعداد داده‌های کم نسبت به شبکه‌های عصبی بیشتر است.

۲- تئوری قوی ماشین بردار پشتیبان قابلیت تعمیم بالایی به آن داده و از قرار گرفتن آن در بهینه‌ی محلی جلوگیری می‌کنند. در آموزش شبکه‌های عصبی از حداقل‌سازی خطای تجربی برای یافتن بهترین مدل استفاده می‌کنند؛ یعنی، شبکه خطای داده‌های آموزشی را حداقل می‌کند اما در SVM از حداقل‌سازی خطای ساختاری برای یافتن بهترین مدل استفاده می‌شود که در آن خطای تعمیم حداقل می‌شود.

به منظور سادگی در فهم، برای بیان تئوری ماشین برداری پشتیبان از ساده‌ترین حالت ممکن، یعنی دسته بندی دو کلاسه در حالت جدایی‌پذیر به‌صورت خطی شروع می‌کنیم.

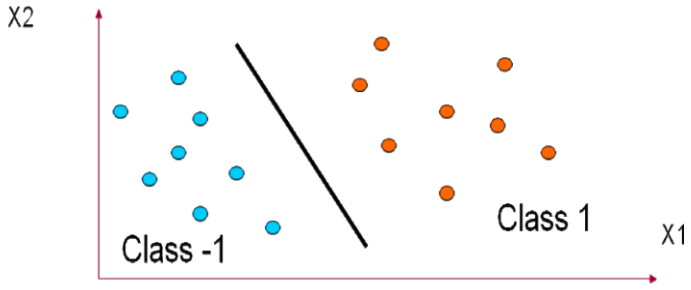
۲-۲-۷-۲-الف- طبقه‌بندی خطی دو کلاسه با ماشین‌های برداری پشتیبان

فرض کنید دو دسته داده‌ی آموزشی که به صورت خطی از هم جداپذیر باشند در اختیار داریم. برای این داده‌ها تعداد زیادی خط جدا کننده وجود دارد که قادر هستند داده‌ها را به صورت خطی جدا کنند اما همان‌طور که در شکل (۲-۳) نشان داده شده است، تنها یکی از آن‌ها دارای ماکزیمم فاصله بین خط جداکننده و نزدیک‌ترین نقاط آموزشی در هر طرف خط هستند که به آن خط جداساز بهینه می‌گویند. به

طور کلی این داده‌ها به صورت زیر تعریف می‌شود:

$$D = \{(X_i, y_i) x_i \in R^n, y_i \in [-1.1]\}_{i=1}^{n} \qquad (۳-۲)$$

که در آن X_i یک بردار حقیقی n بعدی است و y_i برابر ۱ و ۱- است و نشان می‌دهد که هر یک از نقـاط x_i به کدام طبقه تعلق دارند. هدف پیدا کردن ابر صفحه‌ی جداکننده[۱] با بیشترین فاصـله از نقـاط حاشـیه ای است که نقاط با $y_i = +1$ را از نقاط با $y_i = -1$ جدا کند.

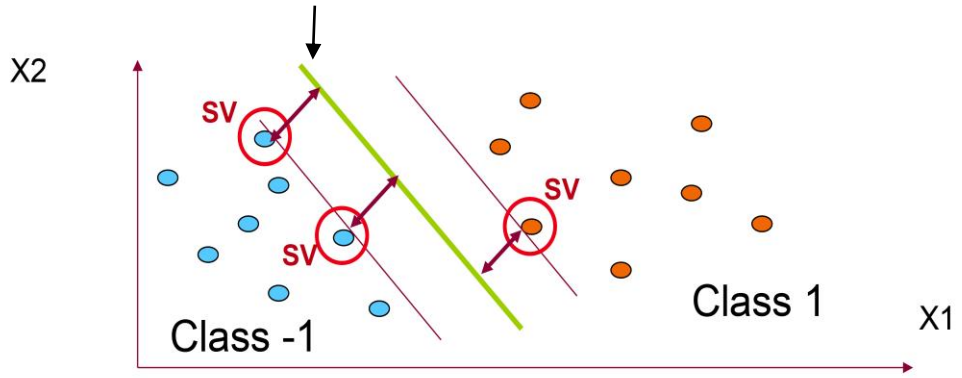


شکل(۲-۳): خط جداساز بهینه با حداکثر مقدار حاشیه [۳۸]

به نزدیک‌ترین داده‌های آموزشی به ابرصفحه‌ی جداساز بهینه که در واقع روی ابرصـفحه‌هـای مـوازی مرزی قرار گرفته‌اند، بردارهای پشتیبان[۲] می‌گوینـد. ایـن داده‌هـا کـه در شـکل (۲-۴) بـا SV نشـان داده شده‌اند، معیاری هستند که ماشین برداری پشتیبان از آن‌هـا بـرای طبقـه‌بنـدی صـحیح داده‌هـا اسـتفاده می‌کند [۳۸] در حالت دو بعدی معادله‌ی این خط به صورت زیر نوشته می‌شود:

$$W_1 X_1 + W_2 X_2 + b = 0 \qquad (۴-۲)$$

۱- Separating hyperplane
۲- Support vectors (SV)



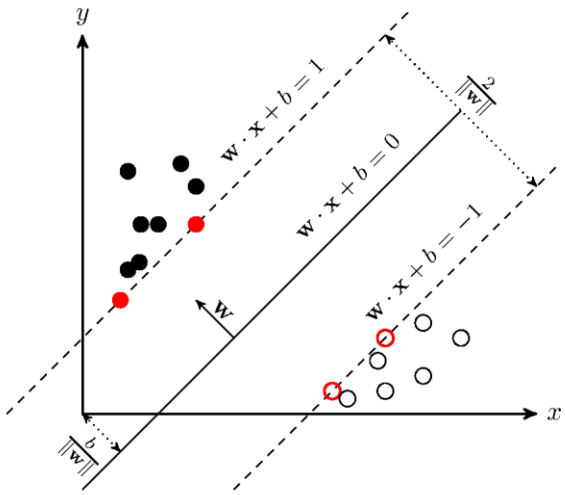
شکل (۲-۴) : نمایش بردارهای پشتیبان روی ابرصفحه‌های موازی مرزی [۳۸]

اکنون موضوع دسته بندی خطی داده‌ها را به فضای n بعدی گسترش می‌دهیم. در این فضا هدف، جداسازی نقاط دو دسته توسط یک ابرصفحه‌ی جداساز $(n-1)$ بعدی است. در این حالت معادله‌ی ابرصفحه‌ی جداکننده با رابطه‌ی زیر بیان می‌شود :

$$W^T.X+b=0 \qquad (۵-۲)$$

که در آن W بردار نرمال ابر صفحه‌ی جداساز بهینه و b عرض از مبدأ آن است. علامت"." بین بردارهای X و W به معنی ضرب داخلی این دو بردار بوده و علامت T بالای بردار W ، ترانهاده‌ی این بردار می‌باشد [۳۸]. باید W و b به گونه‌ای پیدا شوند که اولاً، نمونه‌های آموزشی بدون اشتباه در کلاس خود دسته‌بندی شوند ثانیاً، فاصله‌ی بین نزدیک‌ترین نقاط هر کلاس داده تا ابرصفحه‌ی جداساز، ماکزیمم باشد؛ یعنی حداکثر حاشیه‌ی ممکن بین ابرصفحه‌های مرزی موازی که داده‌ها را از هم جدا می‌کنند ایجاد شود. این موضوع به‌صورت ریاضی چنین بیان می‌شود:

$$W^T.X+b=\pm 1 \qquad 1\le i\le n \qquad (۶-۲)$$



شکل(۲-۵)- صفحه‌ی جداساز و حاشیه‌ها [۳۸]

در شکل (۲-۵) معادلات در نظر گرفته شده برای ابر صفحه‌های مرزی و صفحه‌ی جداساز بهینه نشان داده شده‌است.

فاصله بین دو ابرصفحه‌ی مرزی را می‌توان به کمک هندسه به صورت زیر به دست آورد:

$$d = \frac{\left|\left(W^T.X + b - 1\right) - \left(W^T.X + b + 1\right)\right|}{\|W\|} = \frac{2}{\|W\|} \qquad (۲-۷)$$

براساس این رابطه اگر تابع $\frac{2}{\|W\|}$ را با درنظر گرفتن قید (۲-۸) ماکزیمم کنیم حاشیه‌ی مورد نظر ماکزیمم خواهد شد.

$$y_i\left(W^T.X + b\right) \geq 1 \qquad 1 \leq i \leq n \qquad (۲-۸)$$

اما حل این مسئله‌ی بهینه‌سازی به دلیل وابستگی به $\|W\|$ سخت است، لذا برای سادگی کار می‌توان بدون تغییر در مسئله به جای تابع $\frac{2}{\|W\|}$، تابع $\frac{1}{2}\|W\|$ را با درنظر گرفتن قید (۲-۸) مینیمم کرد که این تابع به صورت $\frac{1}{2}W^T.W$ نوشته می‌شود

برای حل این مسئله از روش ضرایب لاگرانژ استفاده می‌شود. تابع لاگرانژ برای این مسئله به صورت

زیر نوشته می‌شود :

$$L_p(W,b,\alpha)=\frac{1}{2}W^T.W-\sum_{i=1}^{n}\alpha_i\left[y_i\left(W^T.X_i+b\right)-1\right] \qquad (۹-۲)$$

که در آن α_iها ضرایب نامعین لاگرانژ هستند. اکنون هدف مینیمم‌سازی تابع لاگرانژ با درنظر گرفتن $\alpha_i > 0$ است. اگر از رابطه‌ی (۲-۹) نسبت به W و b و مشتق بگیریم و مساوی صفر قرار دهیم مقدار بهینه‌ی W به دست می‌آید.

$$\frac{\partial L}{\partial W}=0\Rightarrow W=\sum_{i=1}^{N}\alpha_i X_i y_i \qquad (۱۰-۲)$$

$$\frac{\partial L}{\partial b}=0\Rightarrow \sum_{i=1}^{N}\alpha_i y_i=0 \qquad (۱۱-۲)$$

حال اگر مقدار W به دست آمده از مشتقات جزیی رابطه‌ی (۲-۱۰) را در خود رابطه‌ی (۲-۹) قرار دهیم معادله‌ی اساسی ماشین‌های برداری به صورت زیر به دست خواهد آمد:

$$L_d(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\ y_i y_j\alpha_i\alpha_j X_i^T X_j \qquad (۱۲-۲)$$

بنابراین، هدف در ماشین‌های برداری حل معادله‌ی (۲-۱۲) با توجه به دو محدودیت (۲-۱۳) و (۲-۱۴) می‌باشد:

$$\alpha_i\geq 0 \qquad (۱۳-۲)$$

$$\sum_{i=1}^{N}\alpha_i y_i=0 \qquad (۱۴-۲)$$

مقدار بهینه‌ی b نیز به کمک قید (۲-۸) به دست می‌آید:

$$W^T.X_i+b=\frac{1}{y_i}=y_i\Rightarrow b=y_i-W^T.X_i \qquad (۱۵-۲)$$

البته بهتر است که با متوسط‌گیری روی همه‌ی بردارهای پشتیبان، مقدار بهینه‌ی b را محاسبه کرد تا در عمل الگوریتم مقاوم‌تری به دست آید:

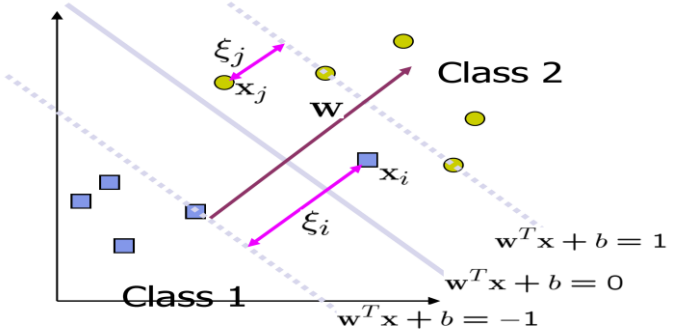
$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} \left(y_i - W^T . X_i \right) \qquad (۱۶-۲)$$

کسر N_{SV} تعداد ابر صفحه‌های موازی است.

با حل مسئله‌ی بهینه‌سازی (۲-۱۲) و استفاده از رابطه‌ی (۲-۱۶) می‌توان به بهینه‌ترین ابرصفحه‌ی جداساز دست یافت و سپس از این ابرصفحه جداساز برای طبقه‌بندی نمونه‌های جدید استفاده نمود [۳۸-۴۱].

۲-۲-۷-۲-ب- طبقه‌بندی خطی سیستم‌های دو کلاسه با ایده‌ی حاشیه‌ی نرم

گاهی اوقات در سیستم‌های خطی همان‌طور که در شکل(۲-۶) نشان داده شده است، داده‌هایی حضور دارند که در کلاس اشتباه طبقه‌بندی شده‌اند. در این حالت با استثناء در نظر گرفتن این داده‌های اشتباه، می‌توان داده‌های آموزشی را به صورت خطی از هم جدا کرد. برای چنین شرایطی ایده‌ی حاشیه‌ی نرم توسط وپنایک[۱] و کورتینا[۲] در سال ۱۹۹۵ مطرح شد [۳۸]. در این روش یک متغیر ξ_i معرفی می‌شود که میزان خطای طبقه‌بندی اشتباه برای هر داده‌ی X_i را نشان می‌دهد. در این حالت تابع هدف که با توجه به دو قید (۲-۱۸) و (۲-۱۹) بهینه شده، به‌صورت زیر تعریف می‌شود:



شکل (۲-۶) : سیستم‌های خطی جدا ناپذیر با میزان خطای ξ_i [۴۲ و ۳۸]

۱- Vapnik
۲- Cortinna

$$Min_{W,b} = \frac{1}{2} W^T.W + C\sum_{i=1}^{N} \xi_i \qquad (۱۷-۲)$$

$$y_i\left(W^T.X_i + b\right) \geq 1 - \xi_i \qquad (۱۸-۲)$$

$$F\left(\xi\right) = \sum_{i=1}^{N} \xi_i \qquad (۱۹-۲)$$

که در رابطه‌ی فوق C ضریب تنظیم جهت ماکزیمم کردن حاشیه‌ها و مینیمم کردن خطاست. با درنظر گرفتن ضرایب لاگرانژ α و β که بزرگتر از صفر هستند می‌توان تابع لاگرانژ را چنین نوشت:

$$L_p\left(W,b,\xi,\alpha,\beta\right) = \frac{1}{2} W^T.W + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \left\{y_i\left(W^T.X_i + b\right) - 1 + \xi_i\right\} - \sum_{i=1}^{N} \beta_i \xi_i \qquad (۲۰-۲)$$

اگر از رابطه‌ی (۲-۲۰) نسبت به W، b و ξ_i مشتق گرفته و مساوی صفر قرار داده شود، مقادیر زیر به دست می‌آیند:

$$\frac{\partial L}{\partial W} = 0 \Rightarrow W = \sum_{i=1}^{N} \alpha_i y_i X_i \qquad (۲۱-۲)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (۲۲-۲)$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow \alpha_i + \beta_i = C \qquad (۲۳-۲)$$

با قرار دادن این روابط در رابطه‌ی (۲-۲۰)، معادله‌ی اساسی ماشین‌های برداری در حالت خطی جداناپذیر به دست می‌آید که به صورت معادله‌ی زیر خواهد بود:

$$L_d\left(\alpha\right) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i^T x_j \qquad (۲۴-۲)$$

در این حالت هدف حل معادله‌ی (۲-۲۴) با توجه به دو محدودیت (۲-۲۵) و (۲-۲۶) می‌باشد :

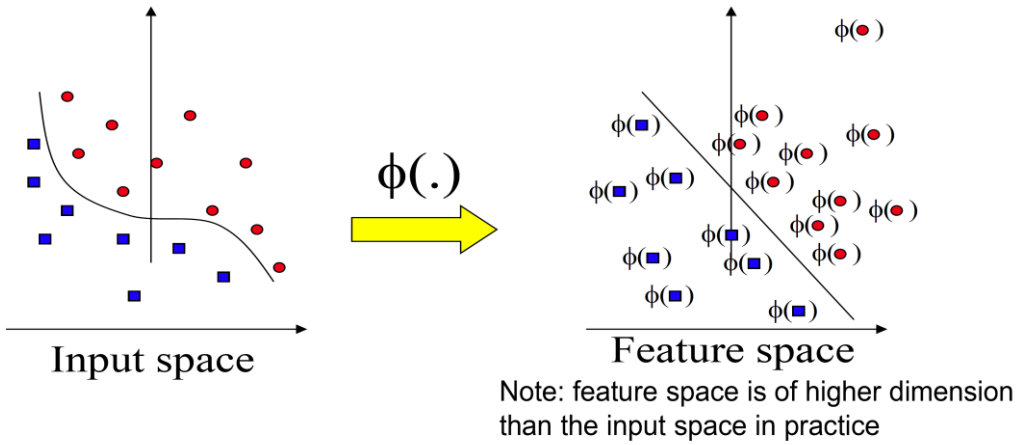
$$0 \leq \alpha_i \leq C \qquad (۲۵-۲)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (۲۶-۲)$$

آنگونه که مشاهده می‌شود، تابع هدف سیستم‌های جداناپذیر خطی مشابه با سیستم‌های جداپذیر خطی است با این تفاوت که ضرایب لاگرانژ α_i دارای کران بالای C هستند. در این حالت پارامتر C نیز که قابلیت کنترل ظرفیت اضافی در طبقه‌بندی کننده را فراهم می‌آورد، باید تعیین شود [۴۰ و ۳۸].

۲-۲-۷-۲-ج- طبقه بندی غیرخطی[۱] با ماشین‌های برداری پشتیبان

در این حالت که نزدیک‌ترین حالت به موارد واقعی می‌باشد، داده‌های آموزشی به صورت غیرخطی از هم جدا می‌شوند. برای این منظور، ابتدا بردارهای ورودی به فضایی با ابعاد بالاتر که فضای ویژگی[۲] نام دارد نگاشته می‌شودکه نگاشت بردار X_i در این فضای جدید، به صورت $\varphi(X_i)$ نمایش داده می‌شود. سپس در این فضای بالاتر ماشین بردار پشتیبان می‌تواند به جداسازی داده‌ها به صورت خطی بپردازد در حالی که فضای ورودی در همان حالت غیرخطی باقی مانده است. شماتیک ساده‌ای از این فرایند در شکل (۲-۷) نشان داده شده است [۴۰].



شکل (۲-۷) : داده‌های ورودی ارجاع داده شده به فضای بالاتر [۴۲ و ۴۰].

۱ -Nonlinear Classification
۲ -Feature Space

بنابراین روابط موجود در سیستم‌های خطی را باید در فضای بالاتر به‌صورت زیر بازنویسی کرد:

$$W = \sum_{i=1}^{N} \alpha_i y_i \varphi(X_i) \qquad (۲۷-۲)$$

$$b = y_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i y_i \varphi(X_i)^T . \varphi(X_j) \qquad (۲۸-۲)$$

به منظور کم کردن حجم محاسبات به دلیل زیاد شدن ابعاد مسئله، تئوری مرکر[۱] در سال ۱۹۹۹ ضرب داخلی $\varphi(X_i).\varphi(X_j)$ را به صورت تابع کرنل $K(X_i, X_j)$ در فضای ویژگی معرفی کرد که به آن کرنل مرکر می‌گویند [۴۰ ، ۳۸]. در نتیجه رابطه‌ی (۲-۲۸) به صورت زیر تغییر می‌یابد :

$$b = y_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i y_i K(X_i, X_j) \qquad (۲۹-۲)$$

با استفاده از ترفند کرنل می‌توان معادله‌ی ابرصفحه‌ی جداساز بهینه در سیستم‌های غیرخطی را به صورت معادله‌ی زیر نوشت:

(۳۰-۲)

$$d(x) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i \alpha_i K(X_i, X_j) + b$$

همان‌گونه که مشاهده می‌شود، معادله‌ی صفحه‌ی جداساز را می توان بدون محاسبه $\varphi(X_i)$ در سیستم‌های غیرخطی به دست آورد و تنها کافیست که از کرنل مناسب برای حل معادله‌ی صفحه استفاده شود. کرنل‌های مختلفی در ریاضیات برای استفاده در فضای ویژگی معرفی شده‌اند که بسته به شرایط، مورد استفاده قرار می‌گیرند. بعضی از این توابع کرنل که در واقع ارتباط دهنده‌ی بین پارامترهای مدل و هدف هستند در جدول (۲-۱) گردآوری شده‌اند [۴۱].

۱- Mercer

جدول(۲-۱) : توابع کرنل در فضای ویژگی

نوع طبقه‌بندی	تابع کرنل
خطی	$K(X_i, X_j) = (X_i^T X_j)^{\rho}$
چند جمله‌ای از درجه ρ	$K(X_i, X_j) = (X_i^T X_j + 1)^{\rho}$
گوسین یا نمایی	$K(X_i, X_j) = e^{\frac{\|X_i - X_j\|^2}{2\sigma^2}}$
پرسپترون چند لایه	$K(X_i, X_j) = \tanh(\gamma X_i^T X_j + \mu)$
برای کلیه مسائل شرایط مرزی	$K(X_i, X_j) = \frac{\sin((n+1/2)(X_i - X_j))}{2\sin((X_i - X_j)/2)}$

۲-۲-۷-۲-د- بردار پشتیبانی رگرسیونگر برای سیستم‌های خطی

یک روش از SVM برای آنالیز رگرسیون در سال ۱۹۹۶ توسط وپنایک و همکارانش پیشنهاد شدکه این روش بردار پشتیبان رگرسیونگر (SVR) نامیده می‌شود. هدف SVR پیدا نمودن تابعی است که بر داده با کمترین انحراف از کمیتی مانند ε برای هر جفت y_i , $\boldsymbol{X}_i$ برازش یابد. به عبارت دیگر، تابع رگرسیون در این حالت یک ابر صفحه جداساز است که بر روی داده‌ها با در نظر گرفتن کمترین مربع خطا بین ابرصفحه و داده‌ها، برازش می یابد. بنابراین تابع رگرسیون به‌صورت زیر بیان می‌شود [۴۲].

$$f(x) = W^T.X + b \qquad (۳۱-۲)$$

در این‌جا نیز تلاش می‌شود که کمترین مقدار ممکن برای $\|W\|$ با توجه به دو قید (۲-۳۲) و (۲-۳۳) به

دست آید.

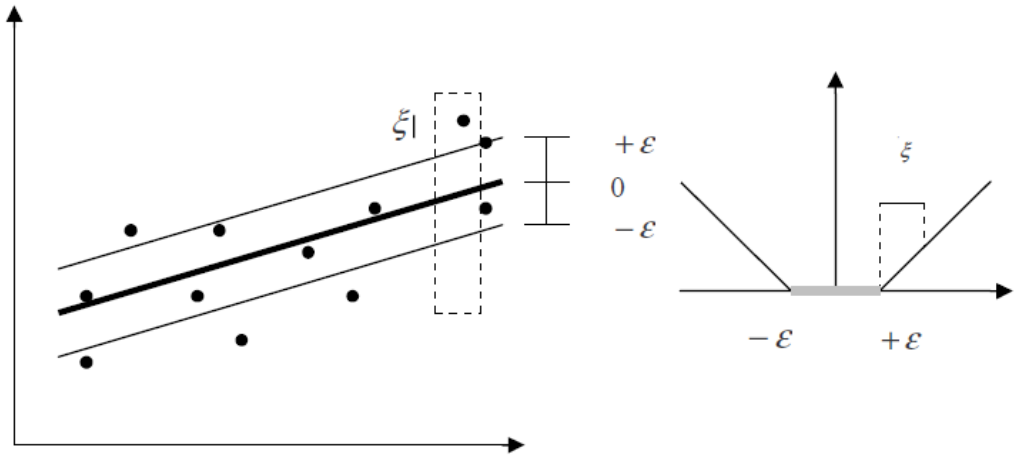
$$y_i - W^T.X - b \le \varepsilon \qquad (۳۲-۲)$$

$$y_i + W^T.X + b \le \varepsilon \qquad (۳۳-۲)$$

برخی مواقع، برخی خطاها ممکن است از مقدار ε بیشتر باشد، بنابراین برای این حالت تابع حساسیت ε[۱] یا متغیر ξ_i معرفی می‌شود که به صورت زیر بیان می‌شود.

$$|\xi| = \begin{cases} 0 & if\ |\xi| < \varepsilon \\ |\xi| - \varepsilon & otherwise \end{cases} \qquad (۳۴-۲)$$

شکل (۲-۸) به‌صورت شماتیکی این حالت را بیان می‌کند.



شکل (۲-۸) : تابع حساسیت به مقدار ε.

بنابراین رگرسیون خطی برای ماشین‌های برداری می‌تواند به صورت تابع اولیه (۲-۳۵) نوشته شود.

$$L_p = \frac{1}{2}\|W\| + C\sum_{i=1}^{N}(\xi_i + \xi_i') \qquad (۳۵-۲)$$

که در آن هدف، مینیمم‌سازی خطای کلی و $\|W\|$ با توجه به قیدهای (۲-۳۶)، (۲-۳۷) و (۲-۳۸) است

۱- ε-Sensitivity

$$y_i - W^T.X - b \le \xi_i + \varepsilon \qquad (۳۶-۲)$$

$$y_i + W^T.X + b \le \xi_i + \varepsilon \qquad (۳۷-۲)$$

$$\xi_i, X_i \ge 0 \qquad (۳۸-۲)$$

در این نوع از مدل‌های رگرسیون، تابع هدف مجموع مقادیر خطا را مینیمم خواهد ساخت و تنها نمونه‌هایی مورد استفاده قرار خواهند گرفت که خطای آن‌ها بیش از ε باشد، بنابراین راه‌حل، تابعی از این نمونه‌ها خواهد بود.

برای حل مسئله‌ی بهینه‌سازی دارای محدویت (۲-۳۵)، باید از بهینه‌سازی لاگرانژ برای تبدیل این معادله به یک معادله‌ی بدون محدودیت استفاده نمود. با در نظر گرفتن تابع لاگرانژ و مشتق‌گیری از تابع هدف بدون محدودیت نسبت به دو پارامتر W و b، دو معادله به صورت (۲-۳۹) و (۲-۴۰) به دست می‌آیند.

$$W = \sum_{i=1}^{N} (\alpha_i - \alpha_i') X_i \qquad (۳۹-۲)$$

$$\sum_{i=1}^{N} (\alpha_i - \alpha_i') = 0 \qquad (۴۰-۲)$$

با قرار دادن رابطه‌ی به دست آمده از معادله‌ی (۲-۳۹) برای مقدار W، در معادله‌ی به دست آمده از تابع لاگرانژ، معادله‌ی اساسی ماشین‌های برداری رگرسیونگر به صورت معادله‌ی زیر نوشته خواهد شد:

$$L_d = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i') {X_i}^T . X_j (\alpha_i - \alpha_i') + \sum_{i=1}^{N} \left((\alpha_i - \alpha_i') y_i - (\alpha_i + \alpha_i') \varepsilon \right) \qquad (۴۱-۲)$$

با توجه به این که :

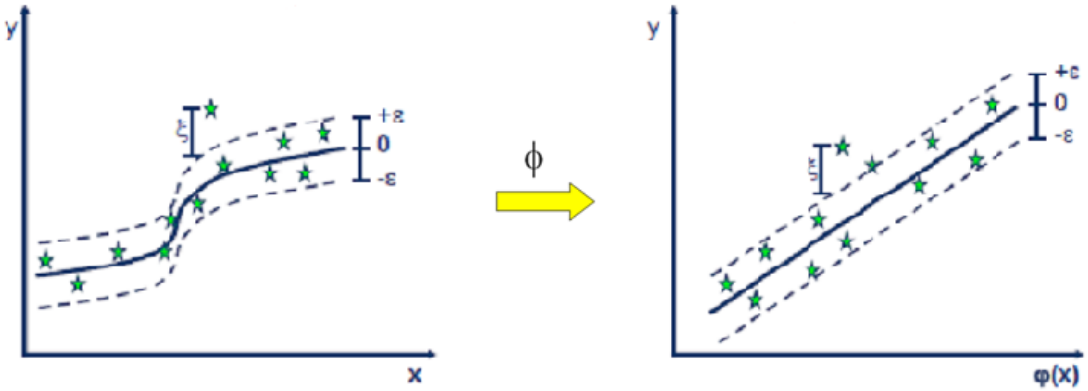
$$0 \le (\alpha_i - \alpha_i') \le C \qquad (۴۲-۲)$$

۲-۲-۷-۲-ه- بردار پشتیبان رگرسیونگر برای سیستم‌های غیرخطی

همان‌گونه که در بحث برداری پشتیبان رگرسیونگر در سیستم‌های خطی بیان شد، مقدار بهینه W از رابطه‌ی $W=\sum_{i=1}^{N}(\alpha_i-\alpha_i')X_i$ به دست می‌آید که این رابطه در سیستم‌های غیرخطی تبدیل به رابطه‌ی $W=\sum_{i=1}^{N}(\alpha_i-\alpha_i')\varphi(X_i)$ می‌شود. مسئله‌ای که در این حالت ایجاد می‌شود، همان مشکلی است که در سیستم‌های غیرخطی وجود دارد و آن بردن به فضای بالاتر و به دست آوردن مقدار $\varphi(X_i)$ است که مقدار آن ناشناخته است. در این شرایط مانند حالت طبقه‌بندی داده‌ها از حقه‌ی کرنل استفاده می‌شود تا بتوان بدون محاسبه‌ی مقدار $\varphi(X_i)$ و تنها با استفاده از کرنل‌های موجود، بهینه‌ترین مدل ریاضی را بر داده‌ها برازش نمود. براساس مطالب بیان شده، رابطه‌ی صفحه‌ی جداساز در سیستم‌های غیرخطی به صورت $y_i=\sum_{i=1}^{N}W^T\varphi(X_i)+b$ بیان می‌گردد. با قراردادن رابطه‌ی به دست آمده برای W در رابطه‌ی صفحه جداساز، می‌توان معادله‌ی اساسی (۲-۴۳) را برای ماشین‌های برداری رگرسیونگر در سیستم‌های غیرخطی به صورت زیر معرفی نمود.

$$y_i=\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i-\alpha_i')\varphi(X_i)^T\varphi(X_j)+b=\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha-\alpha_i')K(X_i,X_j)+b \qquad (۴۳-۲)$$

با در نظر گرفتن رابطه‌ی (۲-۴۳) نیازی به محاسبه مقدار $\varphi(X_i)$ نیست و مقدار b از این رابطه محاسبه می‌شود [۳۸ ، ۴۰]. عملکرد ماشین‌های برداری پشتیبان در قیاس با مدل‌های دیگری که در سیستم‌های غیرخطی وجود دارد، حاکی از قدرت و عملکرد بالای این ماشین‌ها در اکثر سیستم‌های به کارگرفته شده، خصوصاً در سیستم‌های غیرخطی است. مدل ریاضی ارائه شده توسط ماشین برداری برای یک رگرسیون غیرخطی در شکل (۲-۹) نشان داده شده است.



شکل (۲-۹) : مدل ارائه شده توسط ماشین برداری برای سیستم غیرخطی [۴۰]

۲-۲-۸- ارزیابی قدرت پیش‌بینی مدل

۲-۲-۸-۱- با استفاده از پارامتر آماری

برای اطمینان از این که مدل به دست آمده توانایی پیش‌بینی نمونه‌های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی با محاسبه‌ی پارامترهای آماری صورت می‌گیرد. رابطه‌ی ریاضی پارامترهای آماری استفاده شده در این پایان‌نامه، در ادامه توضیح داده خواهد شد.

ضریب تعیین: به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته توضیح داده شده توسط متغیر مستقل است. به عنوان مثال R^2 برابر با ۰/۹۲۴۳ نشان می‌دهد که ۹۲/۴۳درصد تغییرات در متغیر وابسته می‌تواند توسط متغیر مستقل توضیح داده شود. رابطه‌ی ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (۴۴-۲)$$

که SSR[۱] طبق رابطه‌ی (۲-۴۵)، بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است:

۱- Sum Square Regression

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad (۴۵-۲)$$

SST[۱] طبق رابطه‌ی (۲-۴۶) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقـادیر آن است

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (۴۶-۲)$$

که در این روابط، $\hat{y}_i$ مقدار پیش‌بینی شده‌ی متغیر وابسته، y_i مقدار واقعی متغیر وابسته و $\bar{y}$ در هر رابطـه، میانگین مقادیر متغیر وابسته است. همچنین SSE[۲] نیز مبین مجموع مربعات انحراف مقادیر واقعی متغیـر وابسته از مقادیر پیش‌بینی شده برای آن است

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (۴۷-۲)$$

بنابراین با توجه به روابط فوق می‌توان نوشت

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (۴۸-۲)$$

طبق رابطه‌ی (۲-۴۸)، اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشـند، یعنـی بـه ازای تمام نقاط $y_i = \hat{y}_i$ باشد، مقدار R^2 برابر یک شده و هر گونه انحرافی از این حالت باعث می‌شـود کـه مقـدار R^2 از یک کوچکتر شود.

آماره‌ی F : آزمون F یا آزمون فیشر در واقع آزمون معنی‌دار بـودن آمـاری در تحلیـل رگرسـیون سـاده و چند متغیره است و برابر با نسبت میانگین مربعات رگرسیون[۳] (MSR) به میـانگین مربعـات باقیمانـده‌هـا[۴]

۱- Sum Square Total
۲- Sum Square Error
۳- Mean Square Regression
٤- Mean Square Error

(MSE) است. بیان ریاضی آن به صورت زیر می‌باشد:

$$F=\frac{\mathrm{MSR}}{\mathrm{MSE}}=\frac{\mathrm{SSR}/\mathrm{df_m}}{\mathrm{SSE}/\mathrm{df_{res}}} \qquad (۴۹-۲)$$

که در آن df_{m} درجه آزادی مدل بوده و برابر،p تعداد متغیرهای مستقل مدل، است و df_{res} درجه آزادی باقیمانده‌هاست و برابر$n-p-1$ می‌باشد که n تعداد کل ترکیبات مربوط به مدل است.

میانگین مربع خطاها (MSE)[۱]: آماره‌ی MSE از رابطه‌ی (۲-۵۰) به دست می‌آید:

$$\mathrm{MSE}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_{\mathrm{i}}-\hat{y}_{\mathrm{i}}\right)^2}{n} \qquad (۵۰-۲)$$

مجموع مربع باقیمانده‌ها[۲] (PRESS):

$$\mathrm{PRESS}=\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_i-\hat{y}_i\right)^2 \qquad (۵۱-۲)$$

خطای مطلق میانگین[۳] (MAE):

$$\mathrm{MAE}=\frac{\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|}{n} \qquad (۵۲-۲)$$

متوسط درصد انحراف مطلق[۴] (AAD): [۴۳]

$$\mathrm{AAD}=\frac{\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|}{n}\times 100 \qquad (۵۳-۲)$$

۱- Mean Square Error
۲- Predictive Residual Sum of Squares
۳- Mean Absolute Error
۴- Absolute Average Percent Deviation

خطای استاندارد پیش‌بینی[۱] (SEP)

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}} \qquad (۵۴-۲)$$

میانگین درصد انحراف نسبی[۲] (Bias) [۴۳]

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)/y_i}{n} \times 100 \qquad (۵۵-۲)$$

ماکزیمم درصد انحراف مطلق[۳] (D_{max})

$$D_{\max} = \max\left(100\left|\frac{y_i - \hat{y_i}}{y_i}\right|\right) \qquad (۵۶-۲)$$

۲-۲-۸-۲- با استفاده از نمودار برگشتی: در این نمودار مقادیر پیش‌بینی شده‌ی کمیت مورد نظر بر حسب مقادیر تجربی رسم می‌شود و به کمک ضریب تعیین (R^2) به دست آمده از نمودار، پراکندگی نقاط اطراف خط برگشت تعیین می‌شود. محدوده‌ی تغییرات بین صفر و یک است. اگر R^2 =1 باشد، همبستگی کاملی بین داده‌ها وجود دارد و نتایج پیش‌بینی شده به واقعیت نزدیک ترند. اما اگر R^2 =0 باشد، بین داده‌ها هیچ گونه همبستگی وجود ندارد.

۲-۲-۸-۳- با استفاده از نمودار خطای باقیمانده: منظور از عبارت خطای باقیمانده، اختلاف بین مقادیر پیش‌بینی شده و مقادیر تجربی است. اگر پراکندگی مقادیر در دو طرف نمودار صفر باشد، این

۱- Standard Error of Prediction
۲- Relative Average Percent Deviation
۳- Absolute Maximum Percent Deviation

امر نشان دهنده‌ی تصادفی بودن خطاهاست. ولی اگر عمده‌ی نقاط، در این نمودار، در یک طرف صفر باشد، این بدان معناست که خطای جهت داری رخ داده است.

۲-۲-۸-۴- با استفاده از آزمون Y–تصادفی: این تکنیک برای مطالعه همبستگی‌های تصادفی مدل غیر خطی طراحی شده است. در این آزمون، مقادیر تجربی (که در اینجا بردار Y نامیده می‌شود) به صورت تصادفی در محدوده همان مقادیر، تغییر داده شده و سپس همبستگی متغیرهای مستقل با متغیرهای وابسته با استفاده از یکی از شاخص‌های آماری که معمولاً R^2 است، مورد بررسی قرار می‌گیرد. اختلاف زیاد بین شاخص آماری به دست آمده از این روش با شاخص آماری به دست آمده از مدل اصلی، نشان دهنده‌ی عدم وجود همبستگی تصادفی می‌باشد. به طور معمول این فرایند چندین بار انجام می‌شود.

۲-۲-۸-۵- با استفاده از ارزیابی متقاطع یا اعتبار سنجی تقاطعی: رایج ترین تکنیک اعتبار سنجی است که در آن در هر بار یکی یا یک گروه کوچک از داده‌ها کنار گذاشته شده و سپس برای داده‌های باقی مانده، مدلی به دست می‌آید. بعد از آن پاسخ برای داده‌های کنار گذاشته شده از روی این مدل پیش‌بینی می‌شود.این روش ها به ترتیب به نام‌های رد تک تک داده‌ها[۱] و رد گروهی از داده‌ها[۲] نامیده می‌شوند [۴۴].

۱- Leave one out
۲- Leave group out

فصل سوم

پیش‌بینی ضریب فعالیت در رقت بی‌نهایت، ترکیبات آلی و آب در محیط مایع یونی [BMPYR][TCM] با استفاده از روش‌های غیر خطی

در این تحقیق، از تکنیک‌های رگرسیون مرحله‌ای و الگوریتم ژنتیک به عنوان روش‌های انتخاب توصیف‌کننده و از روش‌های شبکه‌ی عصبی مصنوعی و ماشین بردار پشتیبان جهت مدل‌سازی غیرخطی برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞})، ترکیبات آلی و آب در محیط مایع یونی، ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید [BMPYR][TCM]، در ۶ دمای مختلف استفاده گردید. همچنین توانایی مدل‌های به دست آمده با روش‌های مختلف مورد ارزیابی قرار گرفت. این فصل شامل معرفی سری داده‌ها، بهینه‌سازی ساختار مولکول‌ها و محاسبه توصیف‌کننده‌های مولکولی، انتخاب توصیف‌کننده‌های مناسب، مدل‌سازی و ارزیابی مدل برتر است.

۳-۱- مراحل مدل‌سازی

۳-۱-۱- سری داده‌ها

ضریب فعالیت در رقت بی‌نهایت ۵۹ ترکیب حل شونده شامل: آلکان‌ها، سیکلو آلکان‌ها، آلکن‌ها، آلکین‌ها، هیدروکربن‌های آروماتیک، الکل‌ها، تیوفن، اتر، کتون، استونیتریل و پیریدین، در شش دمای مختلف به عنوان سری داده‌ها جمع آوری گردید که این مجموعه داده‌ها شامل ۳۵۴ نقطه داده‌ی ضریب فعالیت در رقت بی‌نهایت می‌باشد. نام این ترکیبات، به همراه محدوده‌ی دمایی و مقدار ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞})، در جدول(۳-۱) نشان داده شده است [۱].

جدول (۳-۱)- نام ترکیبات، محدوده‌ی دمایی و مقدار ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞})

Solute	*T*, K					
	۳۱۸/۱۵	۳۲۸/۱۵	۳۳۸/۱۵	۳۴۸/۱۵	۳۵۸/۱۵	۳۶۸/۱۵
Pentane	۱۶/۷	۱۶/۰	۱۵/۳	۱۴/۵	۱۳/۹	۱۳/۴
Hexane	۲۳/۴	۲۲/۳	۲۱/۱	۲۰/۱	۱۹/۲	۱۸/۳
3- Methylopentane	۲۱/۰	۲۰/۱	۱۹/۱	۱۸/۱	۱۷/۳	۱۶/۶
2,2- Dimethylbutane	۲۱/۱	۲۰/۳	۱۹/۲	۱۸/۴	۱۷/۸	۱۷/۰
Heptane	۳۳/۷	۳۱/۸	۲۹/۹	۲۷/۹	۲۶/۳	۲۴/۹
Octane	۴۸/۱	۴۵/۳	۴۱/۶	۳۹/۱	۳۶/۹	۳۴/۵
2,2-4-trimethylpentane	۳۹/۱	۳۶/۹	۳۵/۰	۳۲/۸	۳۱/۱	۲۹/۴
Nonane	۶۸/۸	۶۳/۹	۵۸/۲	۵۴/۳	۵۰/۸	۴۷/۱
Decane	۹۹/۲	۹۱/۳	۸۲/۴	۷۶/۴	۷۰/۹	۶۵/۲
Cyclopentane	۷/۵۳	۷/۲۹	۷/۰۰	۶/۷۱	۶/۴۸	۶/۲۷
Cyclohexane	۱۱/۱	۱۰/۶	۱۰/۰	۹/۴۹	۹/۱۰	۸/۷۴
Methylcyclohexane	۱۵/۹	۱۵/۳	۱۴/۴	۱۳/۶	۱۳/۰	۱۲/۴
Cycloheptane	۱۳/۵	۱۲/۹	۱۲/۲	۱۱/۵	۱۰/۹	۱۰/۴
Cyclooctane	۱۷/۲	۱۶/۳	۱۵/۳	۱۴/۳	۱۳/۶	۱۲/۹
Pente-1-ne	۷/۵۱	۷/۴۰	۷/۲۹	۷/۱۴	۷/۰۰	۶/۹۲
Hex-1-ene	۱۰/۶	۱۰/۳	۱۰/۰	۹/۶۸	۹/۴۲	۹/۱۵
Cyclohexene	۵/۱۶	۵/۱۰	۴/۹۹	۴/۸۶	۴/۷۷	۴/۶۸
Hept-1-ene	۱۵/۰	۱۴/۷	۱۴/۲	۱۳/۷	۱۳/۳	۱۲/۹
Oct-1-ene	۲۱/۸	۲۱/۰	۲۰/۱	۱۹/۲	۱۸/۵	۱۷/۸
Dec-1-ene	۴۳/۲	۴۱/۴	۳۹/۴	۳۷/۴	۳۵/۶	۳۳/۸
Pen-1-yne	۲/۱۰	۲/۱۷	۲/۲۳	۲/۲۷	۲/۳۰	۲/۳۵
Hex-1-yne	۲/۸۷	۲/۹۷	۳/۰۳	۳/۰۷	۳/۱۱	۳/۱۶
Hept-1-yne	۴/۰۳	۴/۱۵	۴/۲۱	۴/۲۴	۴/۲۸	۴/۳۲
Oct-1-yne	۵/۷۳	۵/۸۵	۵/۸۹	۵/۹۰	۵/۹۱	۵/۹۳
Benzene	۱/۰۲	۱/۰۴	۱/۰۷	۱/۰۹	۱/۱۱	۱/۱۳
Toluene	۱/۴۳	۱/۴۷	۱/۵۱	۱/۵۵	۱/۵۷	۱/۶۰
Ethylbenzene	۲/۱۱	۲/۱۵	۲/۱۹	۲/۲۳	۲/۲۵	۲/۲۹
o-Xylene	۱/۷۶	۱/۸۱	۱/۸۵	۱/۹۰	۱/۹۳	۱/۹۷
m-Xylene	۲/۰۶	۲/۱۲	۲/۱۷	۲/۲۱	۲/۲۵	۲/۳۰
p-Xylene	۲/۰۱	۲/۰۶	۲/۱۲	۲/۱۶	۲/۲۱	۲/۲۵
Styrene	۱/۱۲	۱/۱۶	۱/۲۰	۱/۲۴	۱/۲۷	۳/۳۱
Methanol	۰/۶۹۹	۰/۶۷۰	۰/۶۴۶	۰/۶۲۵	۰/۶۰۷	۰/۵۹۰

ادامه جدول (۳-۱)

Ethanol	۰/۹۶۷	۰/۹۱۸	۰/۸۷۶	۰/۸۳۶	۰/۸۰۱	۰/۷۷۰
Propan-1-ol	۱/۱۴	۱/۰۸	۱/۰۳	۰/۹۸۳	۰/۹۴۵	۰/۹۱۳
Propan-2-ol	۱/۲۰	۱/۱۳	۱/۰۸	۱/۰۲	۰/۹۸۴	۰/۹۵۰
Butan-1-ol	۱/۴۳	۱/۳۴	۱/۲۷	۱/۲۲	۱/۱۶	۱/۱۱
Butan-2-ol	۱/۳۴	۱/۲۷	۱/۲۱	۱/۱۶	۱/۱۲	۱/۰۹
2-Methyl-propan-1-ol	۱/۴۲	۱/۳۲	۱/۲۴	۱/۱۷	۱/۱۲	۱/۰۷
Tert-Butanol	۱/۳۰	۱/۲۳	۱/۱۸	۱/۱۴	۱/۱۱	۱/۰۹
Water	۰/۹۷۳	۰/۹۲۲	۰/۸۸۴	۰/۸۳۳	۰/۸۰۱	۰/۷۷۰
Thiophene	۰/۷۰۹	۰/۷۳۰	۰/۷۵۲	۰/۷۷۳	۰/۷۹۴	۰/۸۲۱
Tetrahydrofuran	۰/۸۴۸	۰/۸۶۸	۰/۸۹۰	۰/۹۱۱	۰/۹۳۴	۰/۹۵۲
1,4- Dioxane	۰/۶۴۲	۰/۶۶۹	۰/۶۹۷	۰/۷۲۱	۰/۷۴۵	۰/۷۶۸
Tert-Buthyl methyl ether	۳/۴۵	۳/۵۲	۳/۵۶	۳/۵۸	۳/۶۲	۳/۶۹
Ethyl tert-butyl ether	۸/۴۴	۸/۳۹	۸/۳۱	۸/۱۶	۸/۰۹	۸/۰۲
Diethyl ether	۳/۳۴	۳/۳۷	۳/۳۸	۳/۳۷	۳/۳۹	۳/۴۰
Di-n-propyl ether	۸/۴۳	۸/۳۳	۸/۱۶	۷/۹۷	۷/۸۲	۷/۷۳
Di-iso-propyl ether	۸/۹۹	۹/۰۲	۸/۹۱	۸/۷۹	۸/۶۳	۸/۵۷
Di-n-butyl ether	۱۷/۳	۱۶/۷	۱۶/۱	۱۵/۵	۱۴/۹	۱۴/۴
Acetone	۰/۶۲۴	۰/۶۳۹	۰/۶۵۰	۰/۶۶۲	۰/۶۷۵	۰/۶۸۷
Pentan-2-one	۱/۰۳	۱/۰۵	۱/۰۸	۱/۱۰	۱/۱۲	۱/۱۴
Pentan-3-one	۱/۰۰	۱/۰۳	۱/۰۶	۱/۰۸	۱/۱۱	۱/۱۳
Methyl acetate	۱/۰۰	۱/۰۲	۱/۰۴	۱/۰۶	۱/۰۸	۱/۰۹
Ethyl acetate	۱/۴۱	۱/۴۳	۱/۴۶	۱/۴۷	۱/۵۰	۱/۵۳
Methyl propanoate	۱/۲۴	۱/۲۷	۱/۳۰	۱/۳۲	۱/۳۵	۱/۳۸
Methyl butanoate	۱/۶۹	۱/۷۲	۱/۷۶	۱/۷۹	۱/۸۲	۱/۸۵
Butanal	۰/۹۰۷	۰/۹۳۰	۰/۹۴۹	۰/۹۶۸	۰/۹۸۶	۱/۰۱
Acetonitrile	۰/۵۵۸	۰/۵۶۳	۰/۵۶۶	۰/۵۶۹	۰/۹۷۲	۰/۵۷۶
Pyridine	۰/۵۶۷	۰/۵۸۰	۰/۵۹۲	۰/۶۰۳	۰/۶۱۷	۰/۶۲۷

۳-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها و محاسبه توصیف‌کننده‌ها

ابتدا ساختار مولکول‌ها با استفاده از نرم افزار HyperChem رسم و سپس با احتساب اتم‌های هیدروژن و استفاده از روش نیمه تجربی AM1، ساختار سه بعدی ترکیبات بهینه گردید. این بهینه‌سازی تا حد گرادیانی ۰/۰۰۱ کیلوکالری بر مول انجام شد.با انجام محاسبات این مرحله، ۲۱ توصیف‌کننده بدست آمد که توصیف‌کننده به دلیل خطای دستگاهی حذف گردید و باقیمانده آنها جهت ادامه محاسبات انتخاب گردید. سپس ساختار بهینه شده‌ی مولکول‌ها به عنوان ورودی به نرم افزار Dragon وارد گردید و برای هر مولکول ۱۴۸۱ توصیف‌کننده در ۱۸ طبقه‌ی مختلف به دست آمد. با توجه به این که ضریب فعالیت به دما وابسته است، لذا یک متغیر تجربی دما هم به آن اضافه گردید. به این ترتیب برای هر ترکیب ۱۵۰۲ توصیف‌کننده به دست آمد.

۳-۱-۳- حذف توصیف‌کننده‌های نامناسب

از آنجا که تعداد ۱۵۰۲ توصیف‌کننده برای مدل‌سازی بسیار زیاد است، باید به دنبال روشی جهت کاهش تعداد توصیف‌کننده‌ها بود تا از انجام محاسبات وقت‌گیر و بیهوده جلوگیری شده و از پیچیدگی کار پرهیز شود. روش مورد استفاده باید قادر باشد تا توصیف‌کننده‌هایی که بیشترین ارتباط را با متغیر وابسته γ_{13}^{∞}، دارند انتخاب کرده و سایر توصیف‌کننده‌ها را حذف نماید. به همین منظور ابتدا توصیف‌کننده‌هایی که برای تمام مولکول‌ها مقادیر ثابت یا تقریباً ثابت داشتند، حذف شدند. سپس به وسیله‌ی برنامه‌ی نوشته شده در محیط نرم افزارMATLAB، توصیف‌کننده‌هایی که همبستگی بزرگتر از ۰/۹داشتند، نیز حذف شدند. در پایان این مرحله ۲۶۶ توصیف‌کننده از مجموع ۱۵۰۲ توصیف‌کننده باقی ماند.

۳-۱-۴- دسته بندی داده‌ها

قبل از شروع مرحله‌ی انتخاب بهترین توصیف‌کننده‌ها باید داده‌ها به سه سری آموزش، ارزیابی و تست تقسیم شوند. بنابراین از بین ۳۵۴ نقطه داده‌ی ضریب فعالیت در رقت بی‌نهایت، به طورتصادفی ۲۵۱ نقطه به عنوان سری آموزش، ۵۳ نقطه به عنوان سری ارزیابی و ۵۰ نقطه به عنوان سری تست انتخاب شدند. داده‌های سری آموزش به منظور ساخت مدل‌های مختلف، داده‌های سری ارزیابی برای بهینه‌سازی پارامترهای موثر بر قدرت پیش‌بینی مدل‌ها و داده‌های سری تست برای ارزیابی و مقایسه‌ی توانایی مدل‌های مختلف در پیش‌بینی ضریب فعالیت در رقت بی‌نهایت ترکیبات به کاربرده شدند. این تقسیم‌بندی در جدول (پ-۱) در بخش پیوست آمده است.

۳-۱-۵- انتخاب بهترین توصیف‌کننده‌ها

پس از دسته بندی داده‌ها باید از بین ۲۶۶ توصیف‌کننده‌ی باقی مانده، بهترین توصیف‌کننده‌ها برای ساختن مدل انتخاب شوند. در این تحقیق از روش خطی رگرسیون مرحله‌ای و روش غیرخطی الگوریتم ژنتیک برای انتخاب توصیف‌کننده‌هایی مناسب استفاده شد که در ادامه به توضیح هر یک از این روش‌ها خواهیم پرداخت.

۳-۱-۵-۱- انتخاب بهترین توصیف‌کننده‌ها به روش رگرسیون مرحله‌ای (SR)

در این مرحله از داده‌های سری آموزش توسط نرم افزار SPSS رگرسیون مرحله‌ای گرفته شد. به این ترتیب که با انتخاب سربرگ Analyze در این نرم افزار، گزینه‌ی Regression و سپس گزینه‌ی linear انتخاب و ۱۵ مدل مختلف به دست آمد. سپس این ۱۵ مدل توسط سری ارزیابی مورد بررسی قرار گرفتند. میانگین مربعات خطا (MSE) و ضریب تعیین (R^2) و آماره‌ی F محاسبه شده برای سری ارزیابی

مدل‌های ۱ تا ۱۵، در جدول (۳-۲) آمده است. مدلی که تعداد توصیف‌کننده‌های کمتر، مقدار ضریب رگرسیون و آماره‌یF بیشتر و مقدار میانگین مربعات خطای کمتری برای سری ارزیابی داشته باشد، به عنوان بهترین مدل انتخاب می‌شود. با توجه به پارامترهای آماری به دست آمده در این مرحله، مدل ۱۲ به عنوان مدل بهینه انتخاب گردید.

برای ارزیابی اهمیت و سهم هر یک از توصیف‌کننده‌ها در مدل، اثر متوسط[۱] (ME) برای هریک از توصیف‌کننده‌ها از رابطه‌ی زیر محاسبه می‌شود [۴۴].

$$ME_j = \frac{\beta_j \sum_{i=1}^{n} d_{ij}}{\sum_{j}^{m} \beta_j \sum_{i=1}^{n} d_{ij}} \qquad (۱-۳)$$

که در این رابطه ME_j، اثر متوسط توصیف‌کننده‌ی j، β_j، ضریب توصیف‌کننده‌ی j، d_{ij} مقدار توصیف‌کننده برای هر مولکول و m تعداد توصیف‌کننده‌ها در مدل می‌باشد.

نام کامل توصیف‌کننده‌های وارد شده در مدل بهینه به همراه نام گروه و اثر متوسط آنها در جدول (۳-۳) آمده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۴) ارائه شده که عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

۱-Mean Effect

جدول(۳-۲)- مقایسه‌ی آماره‌های سری ارزیابی مدل‌های به دست آمده از روش رگرسیون مرحله‌ای (SR)

مدل	توصیف‌کننده	R^2	MSE	F
۱	RDF065u	۰/۵۲۵	۶۷/۶۶	۵۷۵/۸۱۲
۲	RDF065u, RDF030P	۰/۷۹۶۱	۲۸/۱۷۹	۸۴۴/۴۹۸
۳	RDF065u, RDF030P, I_{TH}	۰/۷۹۰	۲۷/۲۴۸	۵۲۸/۵۰
۴	RDF065u, RDF030P, I_{TH}, RDF055u	۰/۸۵۳	۱۸/۸۸۶	۵۶۹/۶۸
۵	RDF065u, RDF030p, RDF055u, I_{TH}, H-046	۰/۸۹۲	۱۴/۰۰۵	۶۳۵/۳۸
۶	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002	۰/۹۴۴	۷/۵۷۶	۱۰۳۵/۱۸۸
۷	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e	۰/۹۳۸	۸/۲۸۶	۷۹۶/۴۹
۸	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e, *T*	۰/۹۳۷	۸/۴۲۲	۶۸۷/۵۹
۹	RDF065u, RDF030p, I_{TH}, RDF055u, H-046 Mor26v C002, MATS6e, *T*,	۰/۹۴۳۷	۷/۸۶۳	۷۹۶/۴۹
۱۰	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, Mor26v, RARS *T*, MATS6e،C002	۰/۹۴۳۲	۸/۱۰۷	۵۸۸/۷۸
۱۱	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, Mor26v, RARS, HOMO C002, MATS6e, *T*,	۰/۹۵۱۷	۷/۰۶۷	۶۱۵/۳۴
***۱۲**	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e, T, Mor26v, RARS, HOMO, BELm7	**۰/۹۶۰۸**	**۵/۴۷۰**	**۷۰۳/۳۱**
۱۳	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e, *T*, Mor26v, RARS, HOMO, BELm7, GGI2	۰/۹۵۳۸	۷/۴۱۰۹	۵۱۶/۳۸۷
۱۴	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e, *T*, Mor26v, RARS, HOMO, BELm7, GGI2, Mor06v,	۰/۹۳۷۹	۹/۲۰۵	۳۷۳/۸۸۶
۱۵	RDF065u, RDF030p, I_{TH}, RDF055u, H-046, C002, MATS6e, *T*, Mor26v, RARS, HOMO, BELm7, GGI2, Mor06v, ATS4e	۰/۹۴۵۸	۸/۵۰۴	۳۸۷/۶۸۹

جدول(۳-۳)- نشان، گروه، نام کامل و اثر متوسط توصیف‌کننده‌های انتخاب شده توسط روش رگرسیون مرحله‌ای (SR)

شماره	نشان	گروه	نام کامل	اثر متوسط
۱	RDF065u	RDF	Radial Distribution Function-6.5/unweighted	۲۸/۳۸
۲	RDF030p	RDF	Radial Distribution Function-3.0/weighted by atomic polarizabilities	۱۰/۹۱۱
۳	I_{TH}	GET AWAY descriptors	Total information content on the leverage equality	-۷/۷۸۹
۴	RDF055u	RDF	Radial Distribution Function-5.5/unweighted	۴/۷۸۴
۵	H-046	Atom-centred fragment	H attached to C0 (sp3) no X attached to next C	۸/۵۰۱
۶	C002	Atom-centred fragment	CH2R2	-۵/۷۸۲
۷	MATS6e	2D auto correlations	Moran auto correlation of lag7 weighted by sanderson electronegativity	۳/۸۳۵
۸	Temperature	Experimental	Temperature	-۳/۸۴۲
۹	Mor26v	3D-MoRSE descriptors	Signal 26/wighted by van der waals volume	-۳/۸۱۰
۱۰	RARS	GET AWAY descriptors	R matrix average row sum	-۴/۲۳۷
۱۱	HOMO	Electron descriptors	The highest occupied molecular orbital	-۳/۰۳۸
۱۲	BELm7	BCUT descriptors	Lowest eigenvalue n. 7 of Burden matrix/weighted by atomic masses	-۳/۱۰۳

جدول(۳-۴)- ماتریس همبستگی بین توصیف‌کننده های انتخاب شده توسط رگرسیون مرحله‌ای

	Temperature	RDF065u	BELm7	C002	RDF030p	I_{TH}	RARS	HOMO	RDF055u	MATS6e	Mor26v	H-046
Temperature	۱											
RDF065u	۰/۰۰۶	۱										
BELm7	-۰/۰۱۵	۰/۸۲۱	۱									
C002	-۰.۰۳۳	۰/۶۷۶	۰/۷۸۱	۱								
RDF030p	-۰/۰۵۳	۰/۳۰۶	۰/۵۳۹	۰/۶۱۱	۱							
I_{TH}	-۰/۰۰۱	۰/۴۸۱	۰/۵۹۱	۰/۳۴۹	۰/۳۹۰	۱						
RARS	-۰/۰۳۶	-۰/۴۸۶	-۰/۴۶۷	-۰/۲۵۱	-۰/۰۹۳	-۰/۶۲۱	۱					
HOMO	۰/۰۲۵	-۰/۱۸۶	-۰/۰۶۷	-۰/۱۹۲	-۰/۰۵۷	۰/۳۸۹	-۰/۵۲۲	۱				
RDF055u	-۰/۰۱۶	-۰/۰۷۴	۰/۱۲۶	۰/۰۷۹	۰/۳۰۶	۰/۴۱۶	-۰/۱۸۲	۰/۲۶۵	۱			
MATS6e	۰/۰۶۷	-۰/۱۷۳	-۰/۰۲۸	-۰/۰۷۳	۰/۱۱۹	۰/۱۹۴	-۰/۰۳۰	۰/۱۲۸	۰/۳۵۶	۱		
Mor26v	-۰/۰۳۵	-۰/۱۳۷	-۰/۰۴۵	۰/۱۵۵	۰/۰۵۲	-۰/۴۶۶	۰/۵۲۵	-۰/۴۵۵	-۰/۱۷۷	-۰/۰۳۷	۱	
H-046	-۰/۰۲۷	۰/۷۰۱	۰/۷۶۶	۰/۸۷۶	۰/۶۷۴	۰/۴۸۴	-۰/۳۴۳	-۰/۱۳۸	۰/۱۰۱	۰/۰۱۸	-۰/۰۴۷	۱

۳-۱-۵-۲- انتخاب توصیف‌کننده‌های معتبر به روش الگوریتم ژنتیک (GA)

در این تحقیق از دو نوع روش انتخاب متغیر استفاده شده که روش دوم انتخاب متغیر روش الگوریتم ژنتیک (GA) است. به منظور اجرای برنامه‌ی الگوریتم ژنتیک، تمام ۲۶۶ توصیف‌کننده باقی مانده به عنوان ورودی به این برنامه داده شدند. برنامه M-file استفاده شده در این تحقیق در بسته‌ی نرم افزاری MATLAB دارای مقادیر پیش فرض احتمال ترکیب ۰/۵، احتمال جهش ۰/۰۱، اندازه‌ی جمعیتی برابر ۳۰ و با تکرار ۵۰۰ می‌باشد. این برنامه ۱۶ بار تکرار گردید تا احتمال انتخاب توصیف‌کننده‌های تصادفی به حداقل برسد. نشان توصیف‌کننده‌هایی که بیش از ۵۰٪ در این الگوریتم تکرار شده‌اند به همراه طبقه و نام کامل آنها در جدول (۳-۵) آورده شده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول(۳-۶) ارائه شده که این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۵)- توصیف‌کننده های انتخاب شده توسط الگوریتم ژنتیکGA

شماره	نشان	گروه	نام کامل
۱	AMW	Average molecular weight	Constitutional indices
۲	PSA	Properties	Fragment-based polar surface area properties
۳	MPC07	Walk and path counts	Molecular path count of order 7
۴	BIC0	Information indices	Bond information content index (neighborhood symmetry of 0-order)
۵	LUMO	Electron descriptors	The lowest unoccupied molecular orbital
۶	GATS2e	2D autocorrelations	Geary autocorrelation of lag2 weighted by Sanderson electronegativity
۷	H-047	Atom-centred fragments	H attached to C1(sp3)/C0(sp2)
۸	O058	Atom-centred fragments	O=
۹	Dipole x	Electron descriptor	Dipole x
۱۰	Temperature	Experimental	Temperature

جدول(۳-۶)-ماتریس همبستگی توصیف‌کننده های انتخاب شده توسط الگوریتم ژنتیک (GA)

	Temperature	AMW	PSA	MPC07	BIC0	LUMO	GATS2e	H-047	O058	Dipole x
Temperature	۱									
AMW	۰/۰۰۶	۱								
PSA	۰/۰۰۴	-۰/۴۲۹	۱							
MPC07	۰/۰۰۲	-۰/۱۴۶	-۰/۲۷۰	۱						
BIC0	۰/۰۰۳	۰/۴۸۰	۰/۶۹۴	-۰/۲۳۵	۱					
LUMO	-۰/۰۰۵	-۰/۶۵۱	-۰/۱۰۴	۰/۰۵۲	-۰/۳۳	۱				
GATS2e	-۰/۰۰۳	-۰/۳۳۴	-۰/۳۹۴	۰/۲۱۴	-۰/۲۸۸	۰/۲۴۶	۱			
H-047	۰/۰۰۵	۰/۴۱۷	۰/۰۶۸	۰/۰۰۱	۰/۱۱۳	-۰/۴۶۹	۰/۲۳۹	۱		
O058	۰/۰۰۲	۰/۲۷۹	۰/۵۴۴	-۰/۱۴۱	۰/۴۰۶	-۰/۴۰۱	-۰/۵۰۳	-۰/۱۰۵	۱	
Dipolex	۰/۰۰۱	۰/۰۷۴	۰/۱۹۲	-۰/۰۳۴	-۰/۰۳۵	-۰/۰۰۲	-۰/۲۶۴	-۰/۱۰۸	۰/۳۴۶	۱

۳-۱-۶- مدل‌سازی غیر خطی با استفاده از شبکه عصبی مصنوعی

یکی از راه‌های یافتن رابطه‌ی غیر خطی بین متغیرهای مستقل و متغیر وابسته، استفاده از شبکه‌ی عصبی مصنوعی برای مدل‌سازی می‌باشد. شبکه‌ی مورد استفاده در این تحقیق، یک شبکه‌ی پیشخور با تکنیک پس‌انتشار است که برنامه نویسی و اجرای آن، در محیط MATLAB انجام شده است. مقدار عددی توصیف‌کننده‌های انتخاب شده توسط دو روش SR و GA به عنوان ورودی و مقادیرتجربی ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞}) نیز به عنوان هدف به شبکه‌ی عصبی داده شدند تا پاسخ شبکه با آنها سنجیده شود. در تمامی مطالعات مربوط به شبکه‌ی عصبی، از دو تابع آموزشی لونبرگ-مارکوات (trainlm) و بایزین (trainbr) و دو تابع انتقال لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig) استفاده گردید. به این ترتیب با ترکیب دو تابع آموزشی با دو تابع انتقال گفته شده، چهار شبکه‌ی عصبی مصنوعی طراحی گردید که هر کدام باید جداگانه بهینه شود. در هریک از شبکه‌های طراحی شده، توصیف‌کننده‌هایی که با روش‌های مختلف انتخاب شده بودند به عنوان ورودی به شبکه داده شدند. بنابراین در هر یک از شبکه‌ها، به ازای هر توصیف‌کننده‌ی انتخابی یک نرون وجود دارد. لایه‌ی خروجی نیز شامل یک نرون می‌باشد که ضریب فعالیت در رقت بی‌نهایت متناظر با ورودی‌ها را نشان می‌دهد. در هر کدام از شبکه‌های طراحی شده، پارامترهای مؤثر بر قدرت پیش‌بینی شبکه که شامل تعداد توصیف‌کننده‌های ورودی (K)، تعداد گره در لایه پنهان (n)، ،تعداد دور آموزش (Epoch) و پارامتر ممنتوم (μ) برای دستیابی به بهترین نتیجه، باید بهینه شوند. تابع کارایی شبکه نیز میانگین مربع خطا (MSE) می‌باشد.

۳-۱-۶-۱- بهینه‌سازی پارامترهای مؤثر بر شبکه با استفاده از توصیف‌کننده‌های حاصل از رگرسیون مرحله‌ای (SR-ANN)

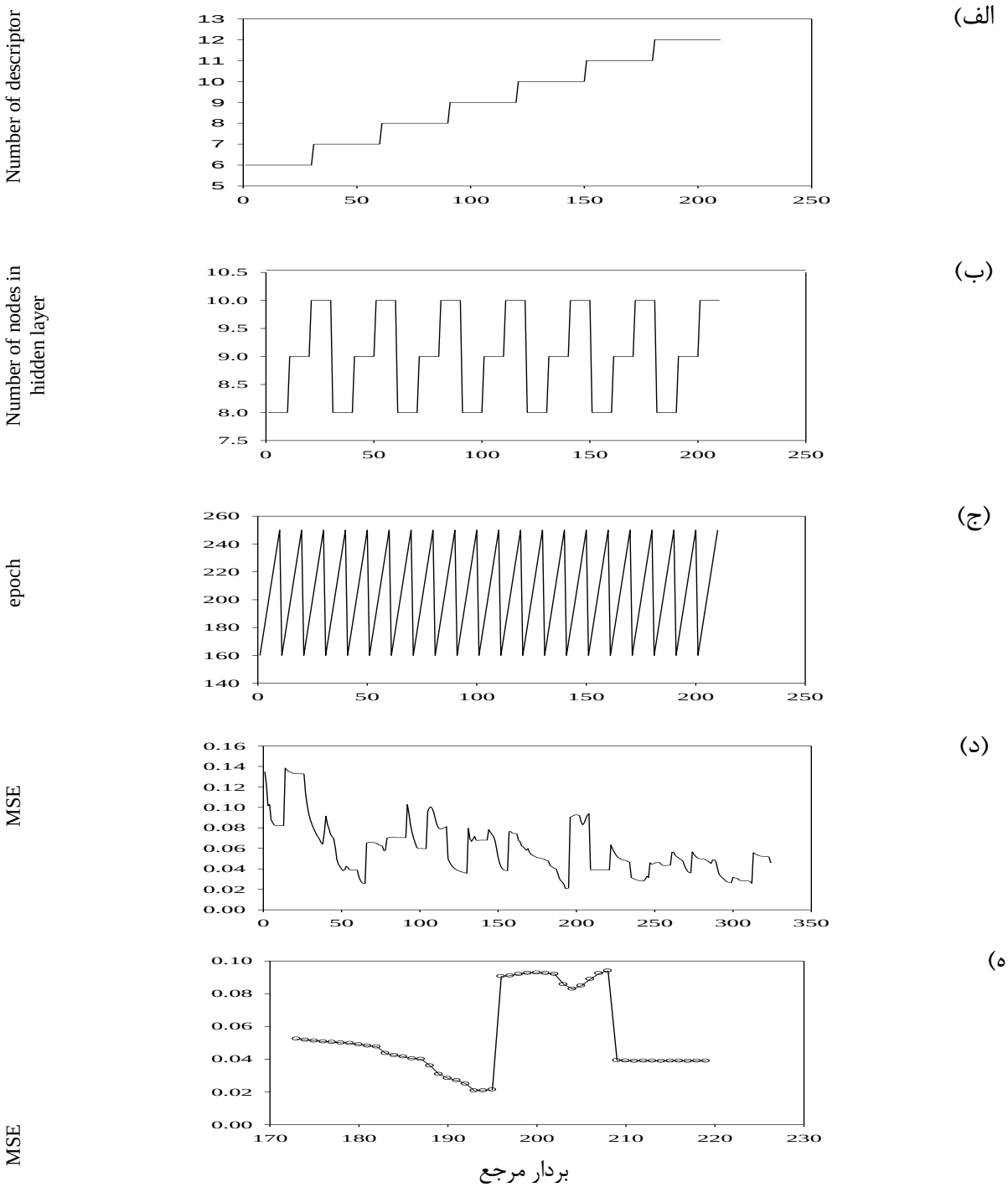
هر شبکه دارای یک لایه‌ی ورودی، یک لایه‌ی خروجی و تعدادی لایه‌ی پنهان می‌باشد. تعداد گره‌های لایه‌ی ورودی برابر توصیف‌کننده‌های بهینه‌ی حاصل از روش رگرسیون مرحله‌ای یعنی تعداد ۱۲ توصیف‌کننده‌ی ذکر شده در بخش (۳-۱-۵-۱) می‌باشد و لایه‌ی خروجی نیز دارای یک نرون است که نشان دهنده‌ی ضریب فعالیت در رقت بی‌نهایت متناظر با هر ترکیب می‌باشد. با توجه به اینکه شبکه، توصیف‌کننده‌ها را به صورت مرحله‌ای به مدل وارد می‌کند لذا ترتیب چیدمان توصیف‌کننده‌ها در ماتریس ورودی شبکه از اهمیت ویژه‌ای برخوردار است. یکی از روش‌های مرتب‌سازی توصیف‌کننده‌ها، بر حسب میزان همبستگی با متغیر وابسته است که به آن رتبه‌بندی همبستگی[۱] (CR) می‌گویند. در جدول (۳-۷) مقادیر همبستگی توصیف‌کننده‌های انتخابی به روش SR با متغیر وابسته نشان داده شده است. توصیف‌کننده‌ای که بیشترین مقدار همبستگی با متغیر وابسته را دارد در ستون اول ماتریس ورودی شبکه قرار می‌گیرد و سایر توصیف‌کننده‌ها نیز به همین ترتیب در ماتریس ورودی شبکه مرتب می‌شوند.

جدول(۳-۷)-مقادیر همبستگی توصیف‌کننده‌های انتخابی با SR با ضریب فعالیت در رقت بی‌نهایت

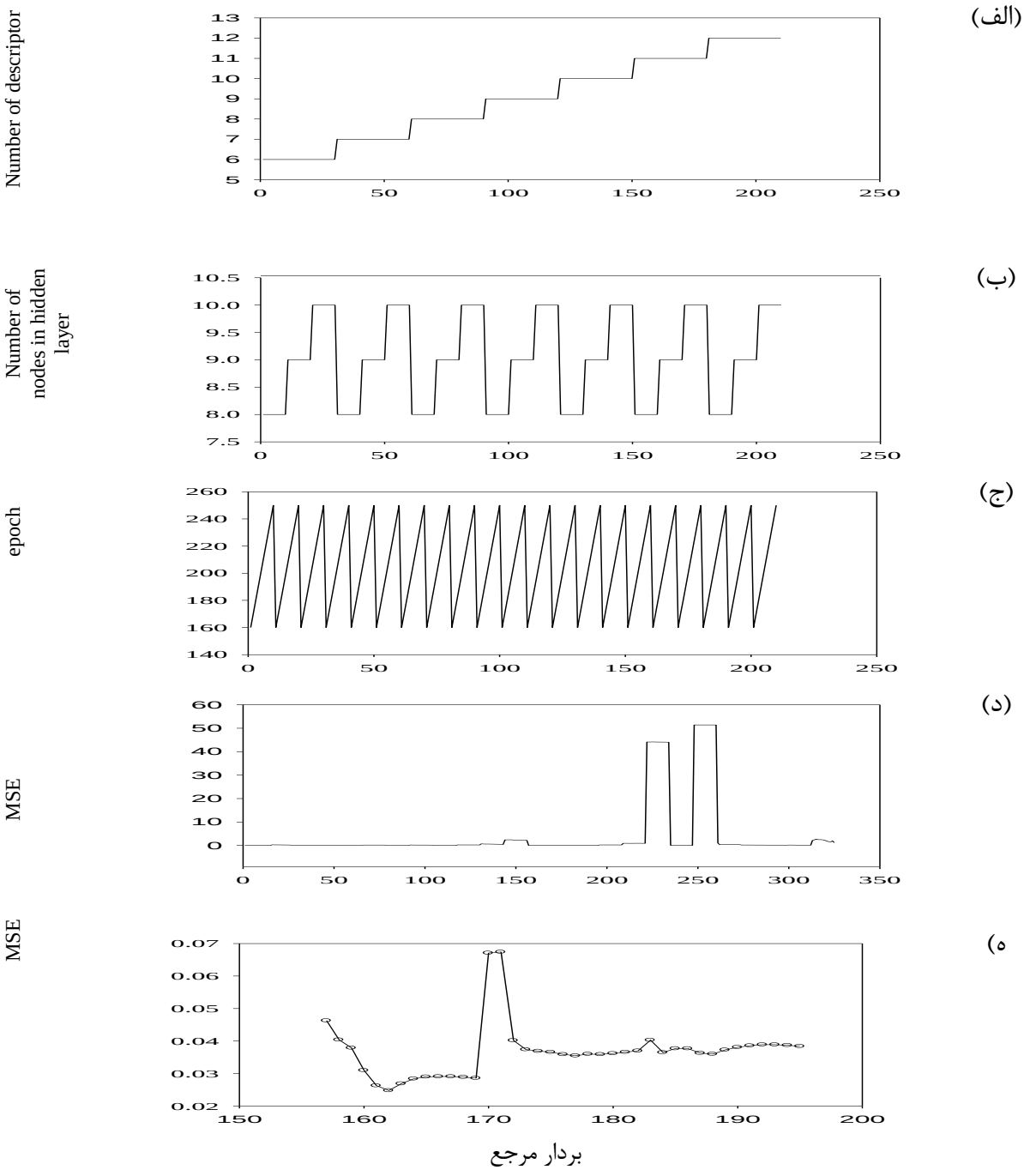
	RDF065u	H-046	BELm7	C002	RDF030P	I_{TH}	RARS	HOMO	RDF055u	MATS6e	T	Mor26v
مقدار همبستگی با $(\frac{1}{C})$	۰/۹۱۴	۰/۷۹۸	۰/۷۸۹	۰/۷۱۰	۰/۵۱۵	۰/۴۰۸	-۰/۳۹۷	-۰/۲۵۲	۰/۰۶۴	-۰/۰۶۱	-۰/۰۶	-۰/۰۳۷

۱- Correlation Ranking

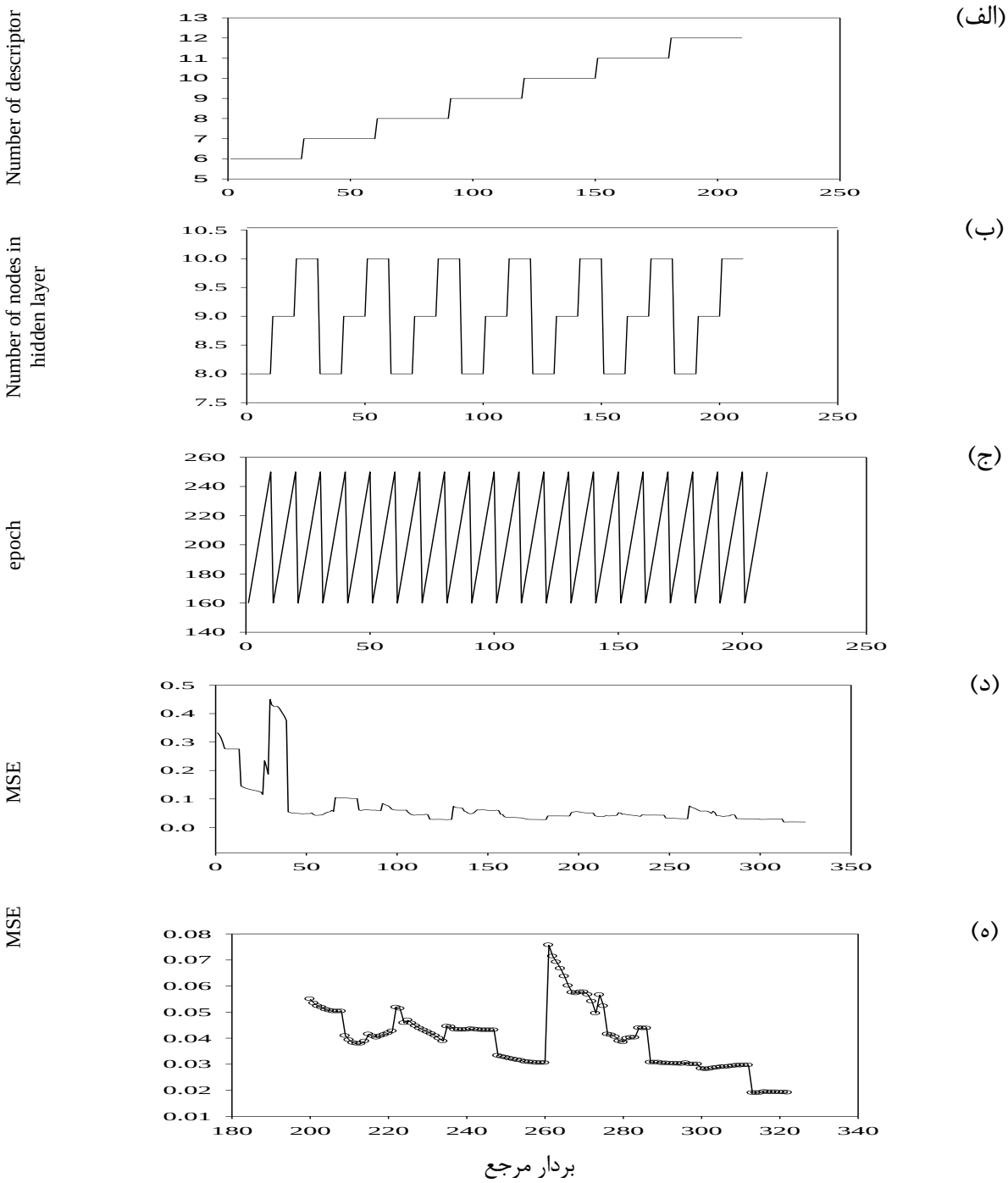
برای بهینه‌سازی تعداد ورودی‌های شبکه، تعداد گره‌های لایه‌ی پنهان و تعداد دور آموزشی از روش بهینه‌سازی همزمان استفاده گردید. برای این منظور، در هریک از شبکه‌های طراحی شده، تعداد ورودی‌های شبکه از ۱تا ۱۲ با گام ۱، تعداد گره‌ی لایه پنهان شبکه از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی شبکه از ۱۰ تا ۲۵۰ با گام ۱۰ تغییر داده شد و به ازای همه‌ی ترکیب‌های ممکن از این سه پارامتر، شبکه های طراحی شده به طور جداگانه آموزش داده شدند. ترکیبی از سه پارامتر که بتواند کمترین مقدار MSE را برای سری ارزیابی ایجاد کند به عنوان پارامترهای بهینه‌ی شبکه انتخاب می‌شوند. لازم به ذکر است که در لایه‌ی خروجی تمام شبکه‌های مورد مطالعه از تابع انتقال خطی (pureline) استفاده گردید. به دلیل این که رسم نمودارهای مربوط به کل محدوده‌ی تغییرات پارامترها بسیار شلوغ و غیر واضح خواهد شد، بنابراین قسمتی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه مقادیر MSE به دست آمده برای سری ارزیابی حول مقادیر بهینه بر حسب یک بردار مرجع در شکل‌های (۳-۱) تا (۳-۴) نشان داده شده است. بردار مرجع، تعداد ترکیب‌های ممکن سه پارامتر، تعداد توصیف‌کننده، تعداد لایه‌ی پنهان و تعداد دور آموزشی را نشان می‌دهد.



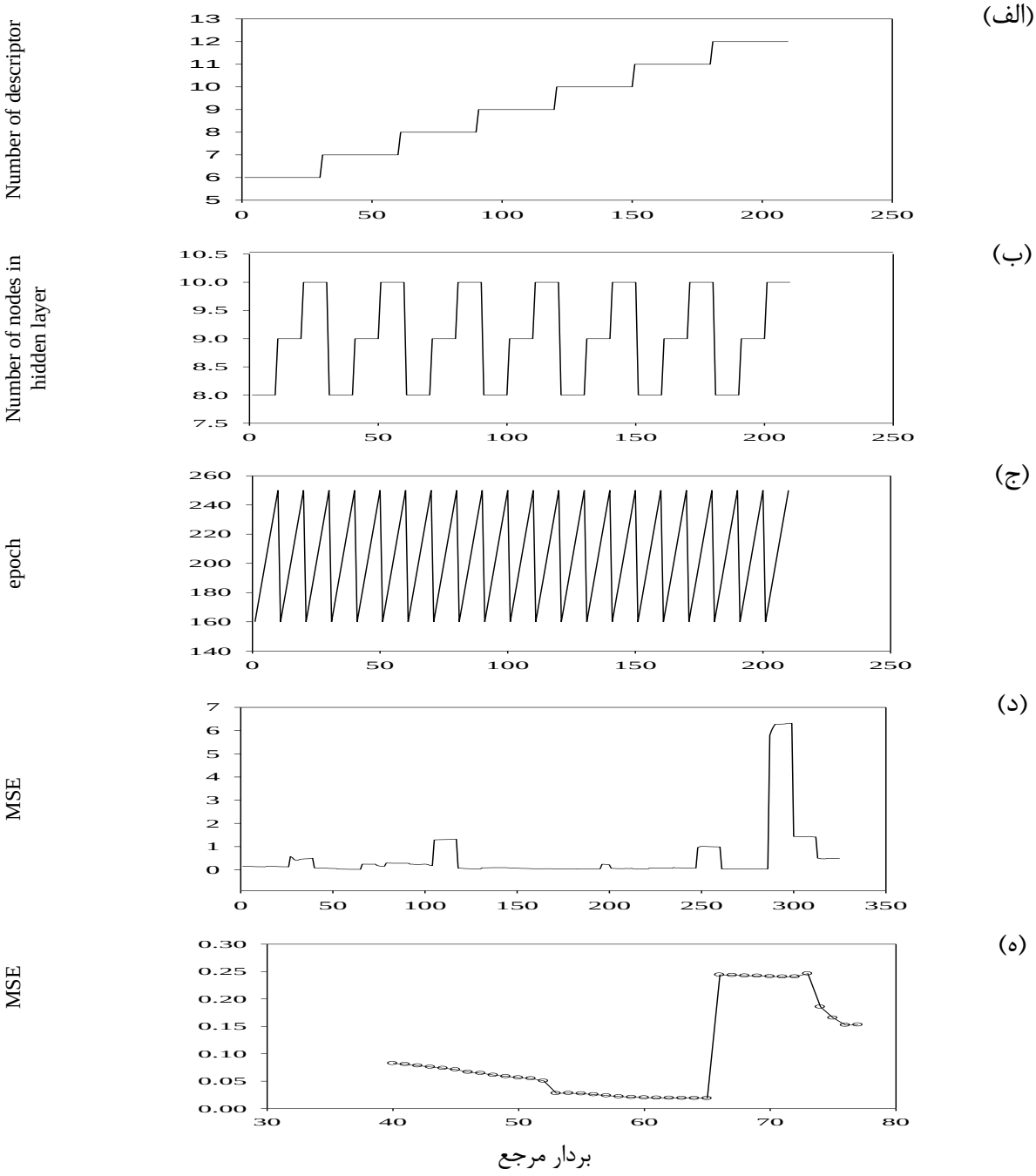
شکل (۳-۱)- نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از SR



شکل (۳-۲)- نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از SR



شکل (۳-۳)- نمودارهای (الف) تعداد توصیف‌کننده (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د)) ناحیه‌ای از نمودار MSE سری ارزیابی برحسب بردار مرجع که مینیمم آن را بهتر نشان می‌دهد، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از SR



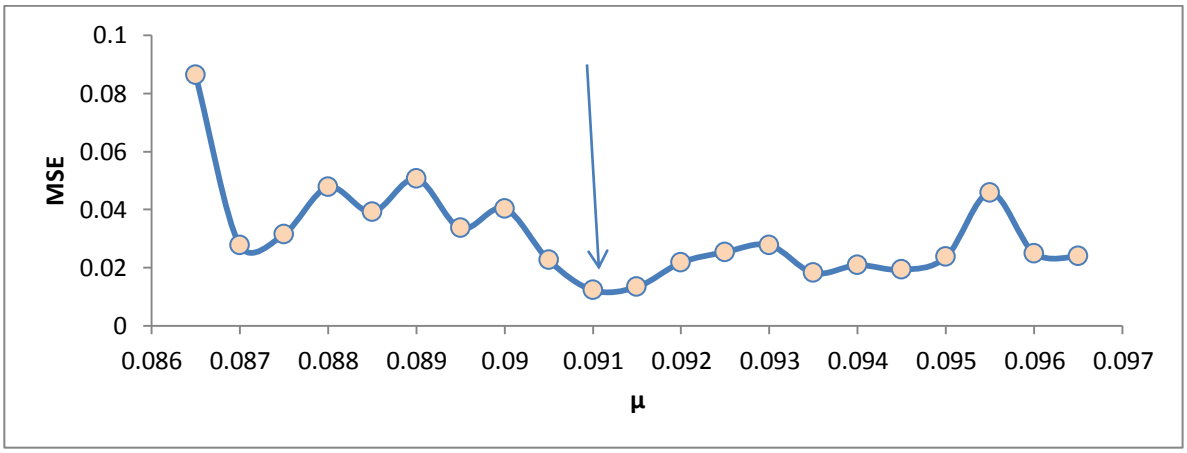
شکل (۳-۴)- نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از SR

مقادیر بهینه‌ی به دست آمده برای هر یک از شبکه‌های طراحی شده در جدول (۳-۸) گردآوری شده است. شبکه‌ای که تعداد توصیف‌کننده‌ی کمتر و میانگین مربع خطای کمتری دارد به عنوان شبکه بهینه انتخاب می‌شود. بنابراین طبق نتایج موجود در جدول (۳-۸) شبکه‌ای با تابع آموزشی تانژانت سیگموییدی و تابع انتقال لونبرگ-مارکوات با ۸ توصیف‌کننده در لایه‌ورودی و ۱۰ گره در لایه‌ی پنهان و دورآموزشی ۲۵۰، به عنوان شبکه بهینه انتخاب گردید.

جدول(۳-۸)- توابع و پارامترهای بهینه‌ی شبکه‌های بهینه (SR-ANN) به دست امده

تابع انتقال	تابع آموزش	تعداد توصیف‌کننده	تعداد نرون لایه‌پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموید	تنظیم بایزین	۱۰	۱۰	۲۳۰	۰/۰۲۰۷
لگاریتم سیگموید	لونبرگ-مارکوارت	۱۰	۸	۱۸۰	۰/۰۲۴۸
تانژانت سیگموید	تنظیم بایزین	۱۰	۱۲	۸۰	۰/۰۱۹۲
تانژانت سیگموید	**لونبرگ-مارکوارت**	**۸**	**۱۰**	**۲۵۰**	**۰/۰۱۸۷**

پس از مشخص شدن شبکه‌ی بهینه، باید پارامتر μ بهینه شود. برای این منظور در شبکه‌ی بهینه، تغییرات μ از ۰/۰۰۰۵ تا ۰/۱ با گام ۰/۰۰۰۵ اعمال شد و برای هر مقدار مومنتوم، مقدار میانگین مربعات خطا برای سری ارزیابی محاسبه گردید. نمودار تغییرات MSE بر حسب μ حول نقطه‌ی بهینه در شکل (۳-۵) نشان داده شده است. ممنتم ۰/۰۹۱ دارای کمترین مقدار میانگین مربعات خطا است و این مقدار به عنوان مقدار ممنتم بهینه انتخاب شد.



شکل(۳-۵) نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ

توابع و مقادیر بهینه شده پارامترهای شبکه‌ی عصبی برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت، با استفاده از روش رگرسیون مرحله‌ای در جدول (۳-۹) ارائه شده است.

جدول (۳-۹)- توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی (SR-ANN)

تابع آموزش	**trainlm**
تابع انتقال لایه‌ی پنهان	**tansig**
تابع انتقال لایه‌ی خروجی	**pureline**
تعداد نرون لایه‌ی پنهان	**۱۰**
تعداد متغیرهای ورودی	**۸**
تعداد دورهای آموزش	**۲۵۰**
پارامتر **μ**	**۰/۰۹۱**
مقدار **MSE** سری ارزیابی	**۰/۰۱۲۲**

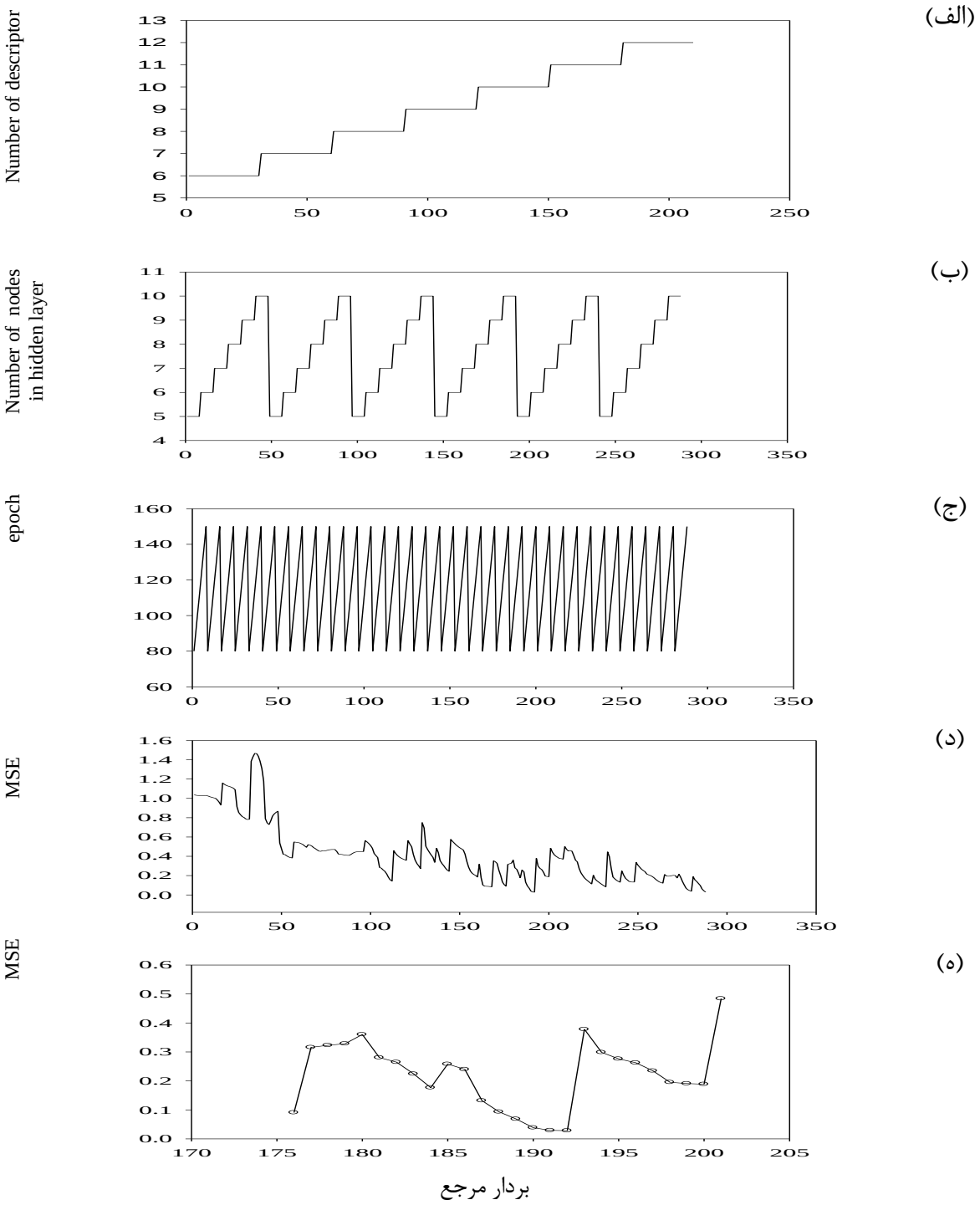
۳-۱-۶-۲- بهینه‌سازی شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک (GA)

شبکه‌ی عصبی مورد مطالعه در این بخش نیز، شبکه‌ای سه لایه می‌باشد. در این مرحله نیز شبکه‌هایی از تلفیق دو الگوریتم آموزشی متفاوت لونبرگ-مارکوات (trainlm) و بایزین (trainbr) و دو تابع انتقال لگاریتم سیگموئید (logsig) و تانژانت سیگموئید (tansig) طراحی شدند و سپس در هریک از شبکه‌های طراحی شده پارامترهای شبکه بهینه‌سازی گردیدند. برای این منظور شبکه‌هایی با مقدار μ ثابت و تعداد ورودی‌های از ۱ تا ۱۰ توصیف‌کننده با گام ۱، تعداد گره لایه‌ی پنهان از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی از ۱۰ تا ۱۵۰ با گام ۱۰ در نظر گرفته شدند و به ازای همه ترکیب‌های ممکن از این سه پارامتر، چهار شبکه عصبی مصنوعی طراحی شده، بهینه‌سازی گردیدند. لازم به ذکر است که در اینجا نیز توصیف‌کننده‌های انتخاب شده توسط روش GA که در جدول (۳-۵) نمایش داده شده‌اند، به روش رتبه‌بندی همبستگی، CR، به شبکه عصبی وارد شدند. در جدول (۳-۱۰) مقادیر همبستگی توصیف‌کننده‌های انتخاب شده توسط روش GA با متغیر وابسته گزارش شده است.

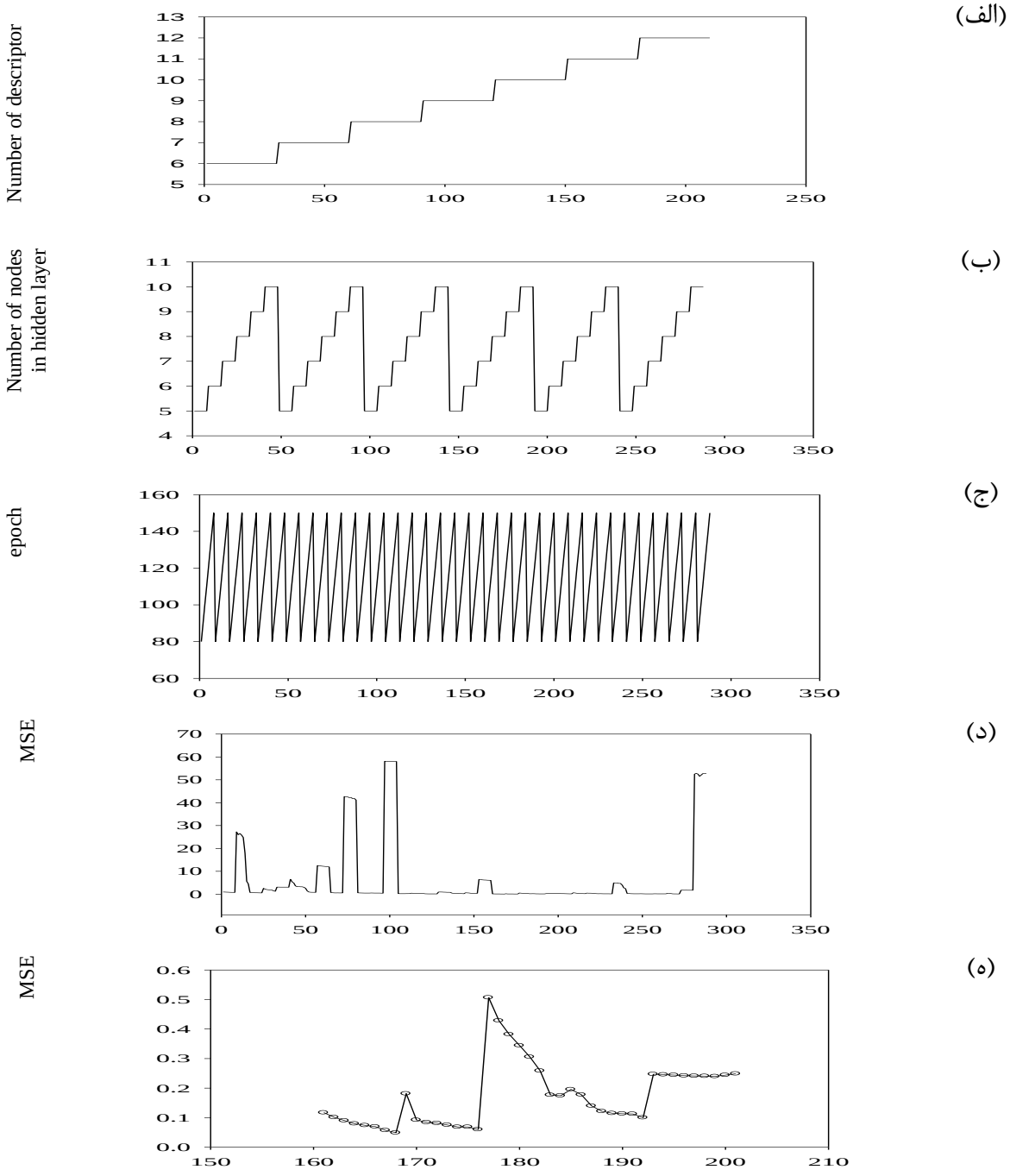
جدول(۳-۱۰)- مقادیر همبستگی توصیف‌کننده‌های انتخابی با روش GA با ضریب فعالیت در رقت بی‌نهایت

	AMW	PSA	MPC07	BIC0	LUMO	GATS2e	H-047	O058	Dipole x	T
همبستگی با مقدار (γ^{∞}_{13})	-۰/۴۸۲	-۰/۴۲۹	۰/۴۱۱	-۰/۳۸۴	۰/۳۵۴	۰/۳۵۰	-۰/۳۰۴	-۰/۲۲۲	-۰/۰۸۳	-۰/۰۶

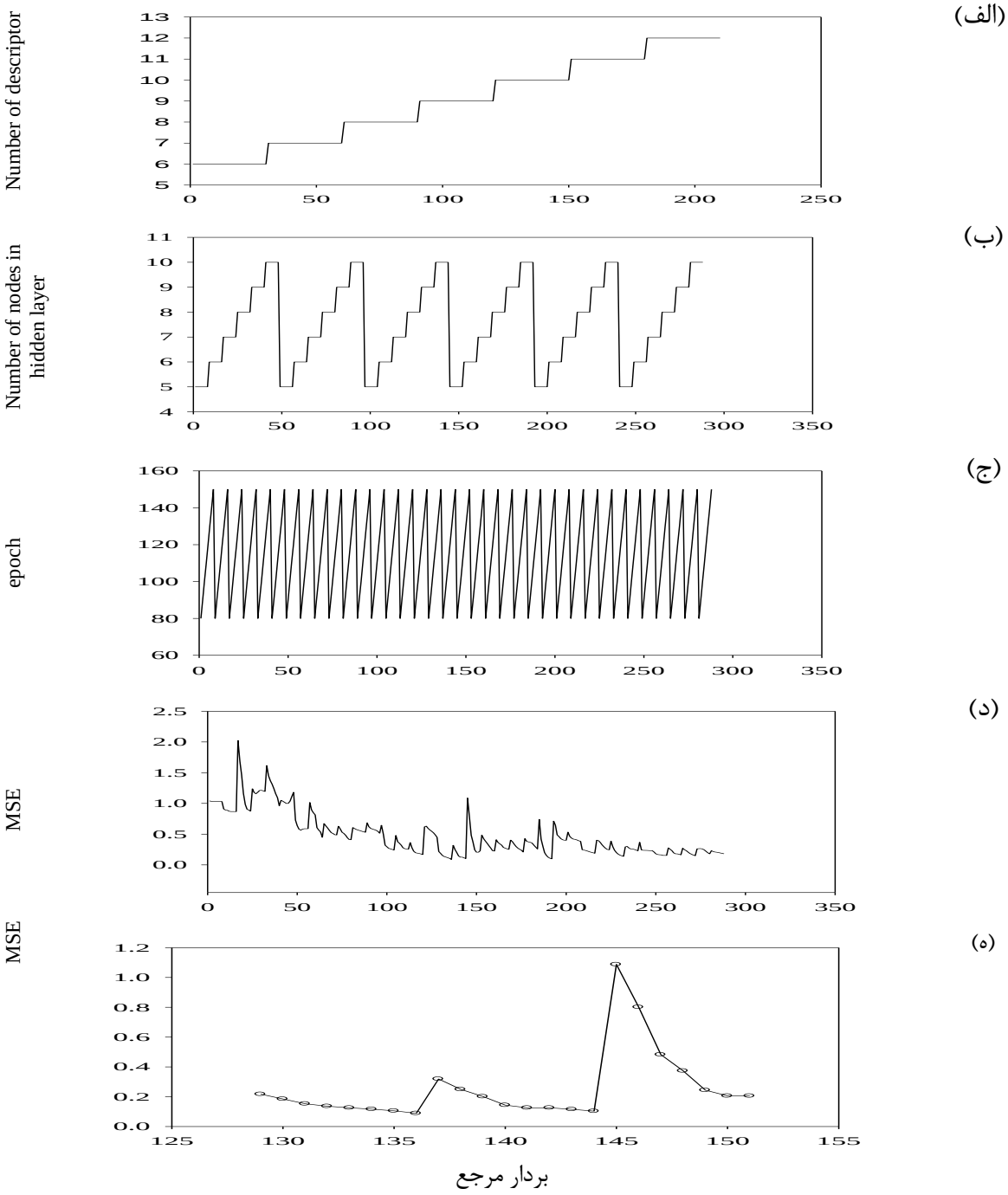
در اینجا مثل بخش (۳-۱-۶-۱)، به حداقل رسیدن میانگین مربعات خطا (MSE) برای سری ارزیابی به عنوان معیار انتخاب متغیرهای بهینه مورد استفاده قرار گرفت. قسمتی از روند تغییرات پارامترهای شبکه حین بهینه‌سازی، حول مقادیر بهینه برای هریک از سه کمیت تعداد توصیف‌کننده (K)، تعداد نرون لایه‌ی پنهان (n) و تعداد دور آموزش بر حسب یک بردار مرجع در شکل‌های (۳-۶) تا (۳-۹) نشان داده شده است.



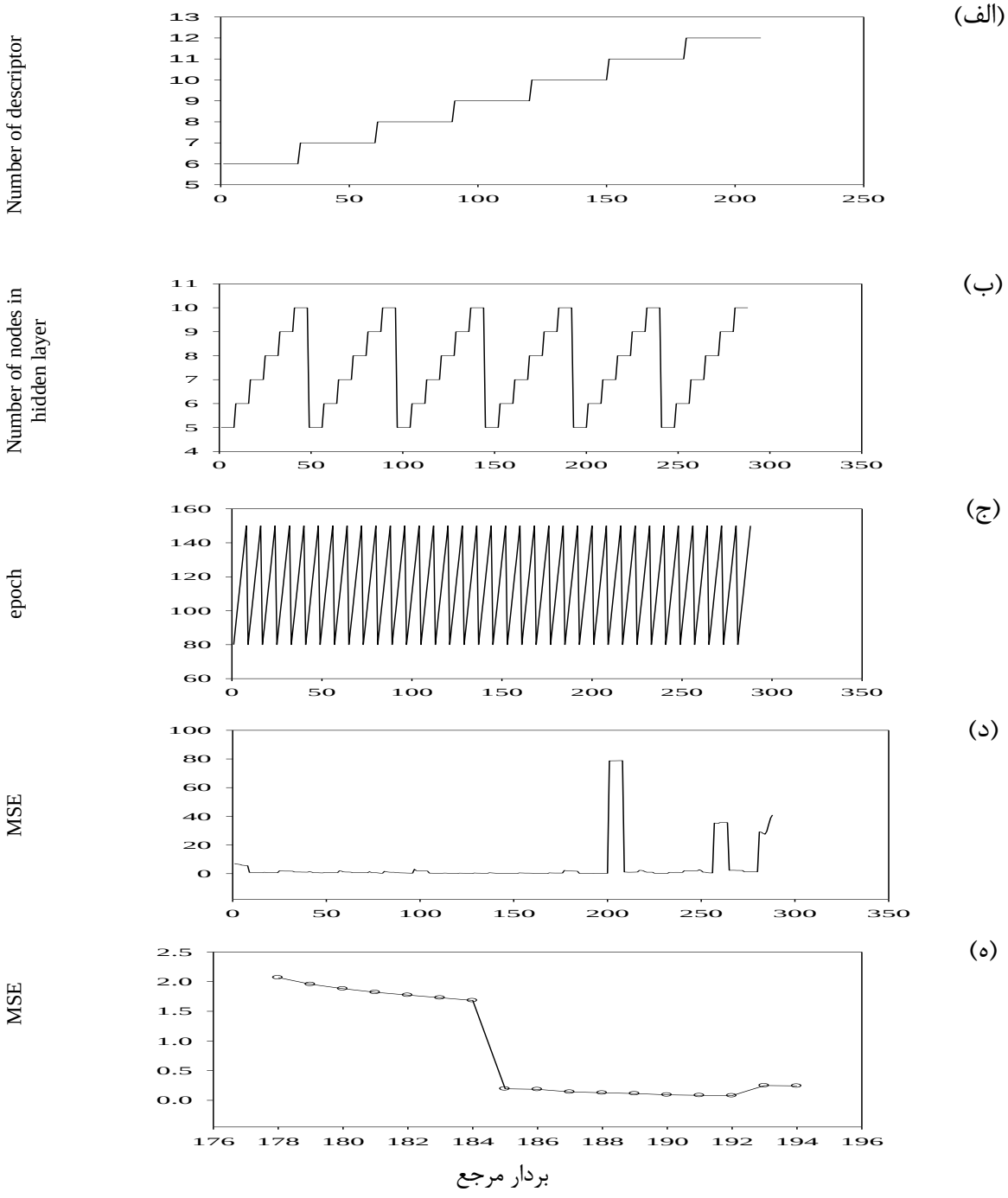
شکل (۳-۶)- نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از GA



شکل (۳-۷)- نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از GA



شکل (۳-۸)- نمودارهای (الف) تعداد توصیف‌کننده (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د)) ناحیه‌ای از نمودار MSE سری ارزیابی برحسب بردار مرجع که مینیمم آن را بهتر نشان می‌دهد، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از GA



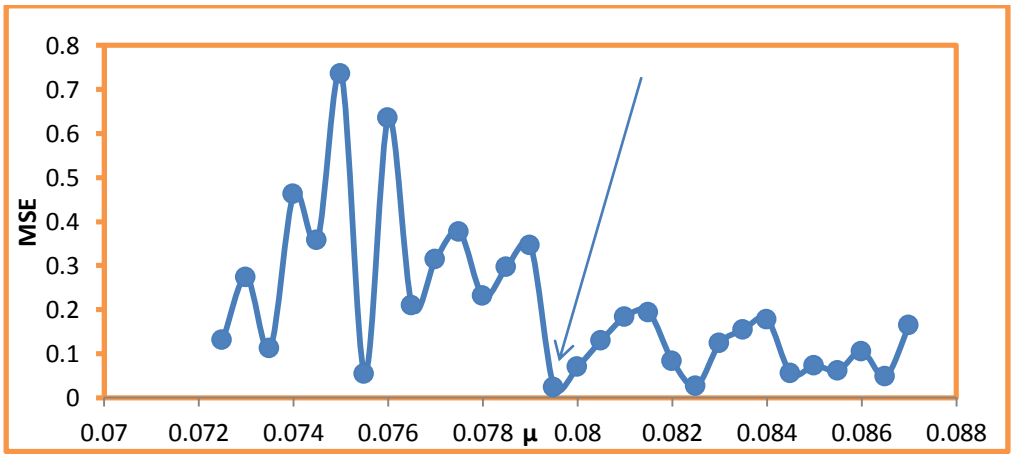
شکل (۳-۹)- نمودارهای (الف) تعداد توصیف‌کننده (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د)) ناحیه‌ای از نمودار MSE سری ارزیابی برحسب بردار مرجع که مینیمم آن را بهتر نشان می‌دهد، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با توصیف‌کننده‌های حاصل از GA

مقدار و پارامترهای بهینه شده‌ی هر یک از شبکه‌های GA-ANN در جدول (۳-۱۱) گزارش شده است. نتایج موجود در این جدول نشان می‌دهد از میان چهار شبکه بهینه شده، شبکه بهینه‌ای که تعداد توصیف‌کننده و MSE کمتر داشته باشد به عنوان بهترین شبکه بهینه انتخاب می‌شود. به همین دلیل شبکه‌ای با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با ۸ توصیف‌کننده در لایه ورودی، ۱۰ گره در لایه‌پنهان و دور آموزشی بهینه‌ی ۱۵۰ کمترین میانگین مربع خطا را داشته و به عنوان شبکه‌ی بهینه انتخاب می‌شود.

جدول(۳-۱۱)- توابع و پارامترهای بهینه‌ی شبکه (GA-ANN)

تابع انتقال	تابع آموزش	تعداد توصیف‌کننده	تعداد نرون لایه‌ی پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموید	تنظیم بایزین	**۸**	**۱۰**	**۱۵۰**	**۰/۰۲۸۵**
لگاریتم سیگموید	لونبرگ-مارکوارت	۸	۷	۱۵۰	۰/۰۴۹۰
تانژانت سیگموید	تنظیم بایزین	۷	۹	۱۵۰	۰/۰۸۸۹
تانژانت سیگموید	لونبرگ-مارکوارت	۸	۱۰	۱۵۰	۰/۰۷۹۲

در ادامه بهینه کردن شبکه مورد بررسی، مقدار پارامتر μ نیز بایستی بهینه شود که بدین منظور، در ساختار شبکه‌ی بهینه با ۸ متغیر ورودی ،۱۰ گره در لایه‌ی پنهان، الگوریتم آموزشی لونبرگ مارکوات، تابع انتقال لگاریتم سیگموییدی و دور آموزشی ۱۵۰، مقدار μ از ۰/۰۰۰۵ تا ۰/۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطا برای سری ارزیابی محاسبه و ثبت گردید که مقدار ۰/۰۷۹۵ به عنوان مقدار بهینه‌ی μ انتخاب شد. نمودار مقدار میانگین مربع خطا بر حسب μ حول نقطه‌ی بهینه در شکل (۳-۱۰) رسم شده است.



شکل(۳-۱۰)- نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ

با توجه به روند بهینه‌سازی ارائه شده، مقدار پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های منتخب الگوریتم ژنتیک (GA-ANN) در جدول (۳-۱۲) گزارش شده است.

جدول (۳-۱۲)- توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های حاصل از GA

تابع آموزش	**trainbr**
تابع انتقال لایه‌ی پنهان	**logsig**
تابع انتقال لایه‌ی خروجی	**pureline**
تعداد نرون لایه‌ی پنهان	**۱۰**
تعداد متغیرهای ورودی	**۸**
تعداد دورهای آموزش	**۱۵۰**
پارامتر μ	**۰/۰۷۹۵**
مقدار **MSE** سری ارزیابی	**۰/۰۲۳۷**

۳-۱-۶-۳- مدل‌سازی ماشین برداری پشتیبان

روش غیر خطی دیگری که برای مدل‌سازی ضریب فعالیت در رقت بی‌نهایت در این پایان نامه استفاده شد، روش ماشین برداری پشتیبان است. برای انجام این روش، داده‌ها به دو دسته آموزش و تست تقسیم شدند. داده های سری تست همان داده‌های سری تست در روش شبکه‌ی عصبی مصنوعی هستند. برای آموزش ماشین برداری پشتیبان باید پارامترهای آن شامل تابع حساسیت وپنایک، پارامتر موازنه ی C و پارامتر کرنل بهینه گردد. بنابراین تابع حساسیت وپنایک دربازه‌ی ۰/۰۰۰۱ تا۰/۱، پارامترموازنه‌ی C در بازه‌ی ۱۰۰ تا ۱۰۰۰۰۰ و پارامتر کرنل دربازه ۰/۰۰۰۱ تا ۰/۱ انتخاب گردید. لازم به ذکر است که در ماشین برداری پشتیبان مورد استفاده در این پایان نامه از کرنل گوسین استفاده گردید. برای بهینه کردن هریک از این سه پارامتر، دو پارامتر ثابت نگه داشته می‌شود و پارامتر سوم در بازه‌ی مورد نظر تغییر داده می‌شود تا جایی که تابع MSE کمترین مقدار شود. در این صورت مقدار آن پارامتر، به عنوان مقدار بهینه انتخاب می‌شود.

بهینه‌سازی سه پارامتر ماشین برداری پشتیبان یکبار با استفاده از توصیف‌کننده‌های انتخاب شده به روش SR انجام گرفت که نتایج آن در جدول (۳-۱۳) گزارش شده است و بار دوم همین کار با استفاده از توصیف‌کننده‌های انتخاب شده به روش GA انجام گرفت که نتایج آن در جدول (۳-۱۴) ارائه شده است.

نتایج جدول (۳-۱۳) نشان می‌دهد که ماشین برداری پشتیبان با مقادیر ۰/۱ برای تابع حساسیت وپنایک، ۳۹۰۰۰۰ برای پارامتر موازنه‌ی C و ۰/۰۰۰۷ برای پارامتر کرنل به عنوان ماشین بهینه برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت ترکیبات آلی و آب در مایع یونی [BMPYR][TCM]، با استفاده از توصیف‌کننده‌های منتخب SR می‌باشد

جدول(۳-۱۳)- مقادیر مختلف پارامترهای ماشین برداری پشتیبان با استفاده از توصیف‌کننده‌های انتخاب شده توسط روش SR و MSE مربوط به آنها

پارامتر کرنل	پارامتر موازنه C	تابع حساسیت وپنایک	MSE
۰/۱	۱۰۰	۰/۱	۵۲/۳۹
۰/۰۱	۱۰۰	۰/۱	۷/۵۵
۰/۰۰۱	۱۰۰	۰/۱	۶/۷۱۶
۰/۰۰۰۸	۱۰۰	۰/۱	۶/۸۹۷
۰/۰۰۰۷	۱۰۰	۰/۱	۶/۵۶۵
۰/۰۰۰۶	۱۰۰	۰/۱	۶/۶۶۱
۰/۰۰۰۱	۱۰۰	۰/۱	۱۲/۵۳۸
۰/۰۰۰۷	۱۰۰۰	۰/۱	۴/۴۴
۰/۰۰۰۷	۱۰۰۰۰	۰/۱	۱/۳۸۸۹
۰/۰۰۰۷	۵۰۰۰۰	۰/۱	۰/۳۷۷۱
۰/۰۰۰۷	۱۰۰۰۰۰	۰/۱	۰/۲۱۹۲
۰/۰۰۰۷	۲۰۰۰۰۰	۰/۱	۰/۱۵۱۹
۰/۰۰۰۷	۳۰۰۰۰۰	۰/۱	۰/۰۹۱۵
۰/۰۰۰۷	۳۶۰۰۰۰	۰/۱	۰/۰۸۱۴
۰/۰۰۰۷	۳۸۰۰۰۰	۰/۱	۰/۰۷۵۹۹
۰/۰۰۰۷	۳۹۰۰۰۰	۰/۱	۰/۰۷۰۶۷
۰/۰۰۰۷	۴۰۰۰۰۰	۰/۱	۰/۰۷۵۴۰
۰/۰۰۰۷	۵۰۰۰۰۰	۰/۱	۰/۰۸۲۵۸
۰/۰۰۰۷	۸۰۰۰۰۰	۰/۱	۰/۰۹۷۳
۰/۰۰۰۷	۱۰۰۰۰۰	۰/۱	۰/۱۱۷۸
۰/۰۰۰۷	۳۹۰۰۰۰	۰/۰۱	۰/۰۷۷۷۷
۰/۰۰۰۷	۳۹۰۰۰۰	۰/۰۰۱	۰/۰۸۶۵۲۲
۰/۰۰۰۷	۳۹۰۰۰۰	۰/۰۰۰۱	۰/۰۸۶۲۲

جدول(۳-۱۴)- مقادیر مختلف پارامترهای ماشین برداری پشتیبان با استفاده از توصیف‌کننده‌های انتخاب شده توسط روش GA و MSE مربوط به آنها

پارامتر کرنل	پارامتر موازنه C	تابع حساسیت وپنایک	MSE
۰/۰۰۱	۱۰۰	۰/۱	۳۲/۹۶۱۱
۰/۰۰۱	۱۰۰۰	۰/۱	۲۴/۰۸۵
۰/۰۰۱	۱۰۰۰۰	۰/۱	۲۴/۸۰۵۴
۰/۰۰۱	۱۰۰۰۰۰	۰/۱	۱۳/۷۸۸
۰/۰۰۱	۲۰۰۰۰۰	۰/۱	۹/۸۳
۰/۰۰۱	۵۰۰۰۰۰	۰/۱	۷/۳۶
۰/۰۰۱	۱۰۰۰۰۰۰	۰/۱	۵/۳۵
۰/۰۰۱	۱۳۰۰۰۰۰	۰/۱	۴/۹۹۵
۰/۰۰۱	**۱۴۰۰۰۰۰**	۰/۱	**۴/۸۶**
۰/۰۰۱	۱۵۰۰۰۰۰	۰/۱	۴/۸۷۵
۰/۰۰۱	۲۰۰۰۰۰۰	۰/۱	۵/۱۷۶
۰/۰۱	۱۴۰۰۰۰۰	۰/۱	۸۵۷/۷۴۳
۰/۰۰۰۸	۱۴۰۰۰۰۰	۰/۱	۵/۵۴
۰/۰۰۰۷	۱۴۰۰۰۰۰	۰/۱	**۴/۱۷**
۰/۰۰۰۶	۱۴۰۰۰۰۰	۰/۱	۵/۹۱
۰/۰۰۰۱	۱۴۰۰۰۰۰	۰/۱	۱۴/۰۵۲۲
۰/۰۰۰۷	۱۴۰۰۰۰۰	۰/۰۳	۴/۵۲
۰/۰۰۰۷	**۱۴۰۰۰۰۰**	**۰/۰۲**	**۴/۰۳**
۰/۰۰۰۷	۱۴۰۰۰۰۰	۰/۰۱	۴/۳۰
۰/۰۰۰۷	۱۴۰۰۰۰۰	۰/۰۰۱	۴/۸۶

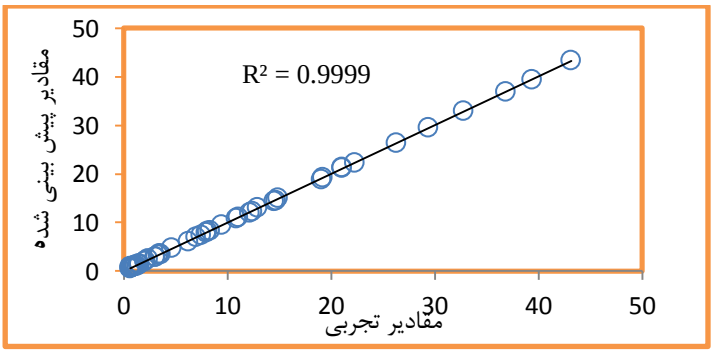
به طور مشابه نتایج موجود در جدول (۳-۱۴) نشان می‌دهد که ماشین بـرداری پشـتیبان بـا مقـادیر ۰/۰۲، ۱۴۰۰۰۰۰ و ۰/۰۰۰۷ به ترتیب برای تابع حساسیت وپنایک، پـارامتر موازنـه‌ی C و پـارامتر کرنـل می‌تواند به عنوان ماشین بهینه برای پیش‌بینی γ_{13}^{∞} ترکیبات مورد نظر در مایع یونی [BMPYR][TCM]، با استفاده از توصیف‌کننده‌های حاصل از GA به کار رود.

۳-۲-۷- ارزیابی قدرت پیش‌بینی مدل‌های غیر خطی

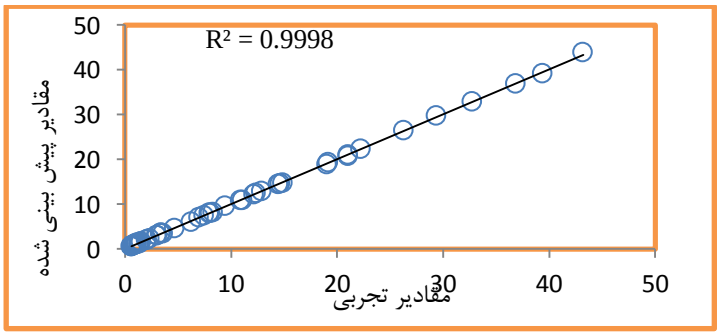
۳-۲-۷-۱- با استفاده از نمودار برگشتی

در نمودار برگشتی مقادیر پیش‌بینی شده بر حسب مقادیر تجربی رسم می‌گردد و با توجه به مقدار ضریب تعیین (R^2) به دست آمده از نمودار، پراکندگی نقاط در اطراف خط برگشت تعیین می‌شود. هر چه مقدار ضریب تعیین به یک نزدیک‌تر باشد، مدل ساخته شده، مدل بهتری است. نتایج حاصل از نمودارهای برگشتی مربوط به چهار روش انجام شده در این کار برای سری ارزیابی و تست در شکل‌های (۳-۱۱) و (۳-۱۲) و (۳-۱۳) آورده شده است. لازم به ذکر است که مقادیر پیش‌بینی شده سری تست γ_{13}^{∞} توسط چهار روش SR-ANN، GA-ANN، SR-SVM و GA-SVM به همراه مقادیر واقعی مربوطه در جدول (پ-۲) در بخش پیوست گزارش شده است.

الف)

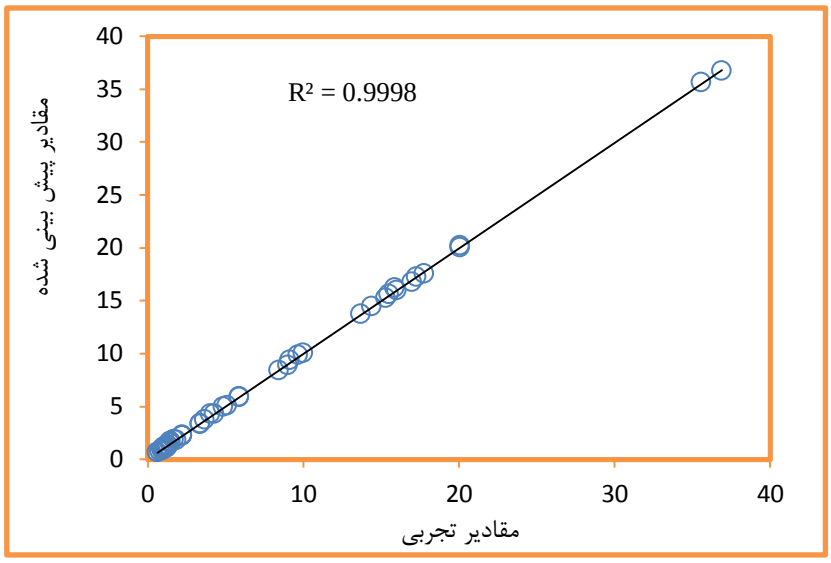


ب)

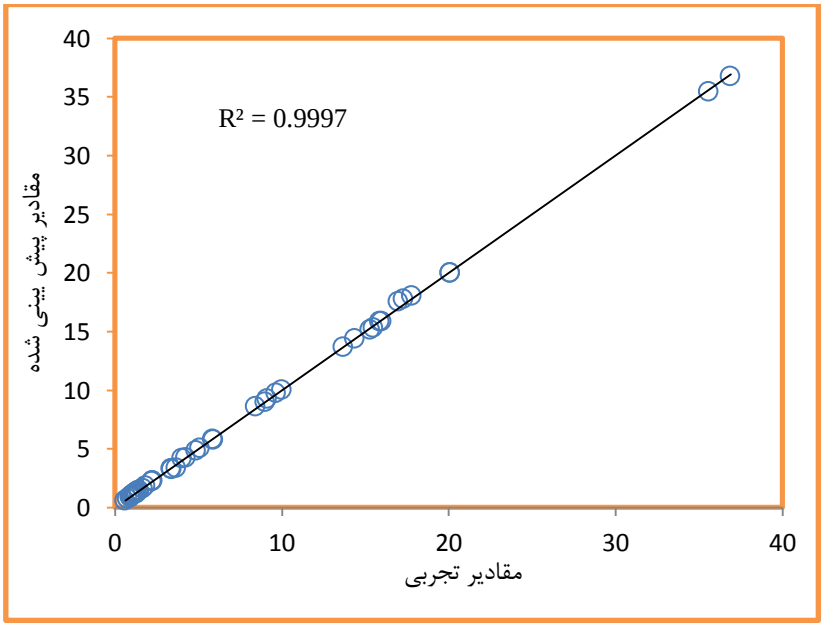


شکل (۳-۱۱)- نمودار مقادیر پیش‌بینی شده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN و (ب) مدل GA-ANN

الف)

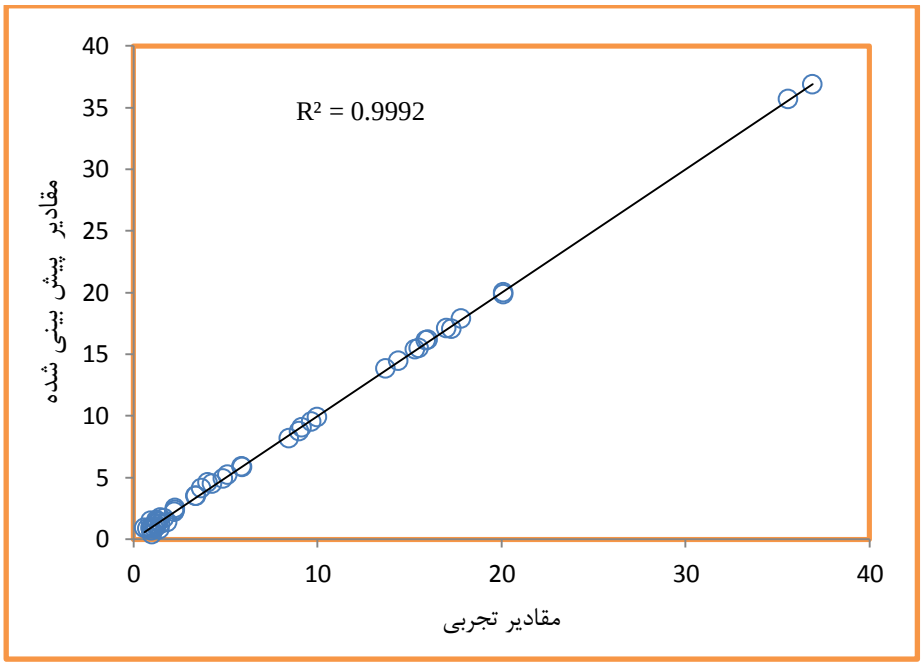


ب)

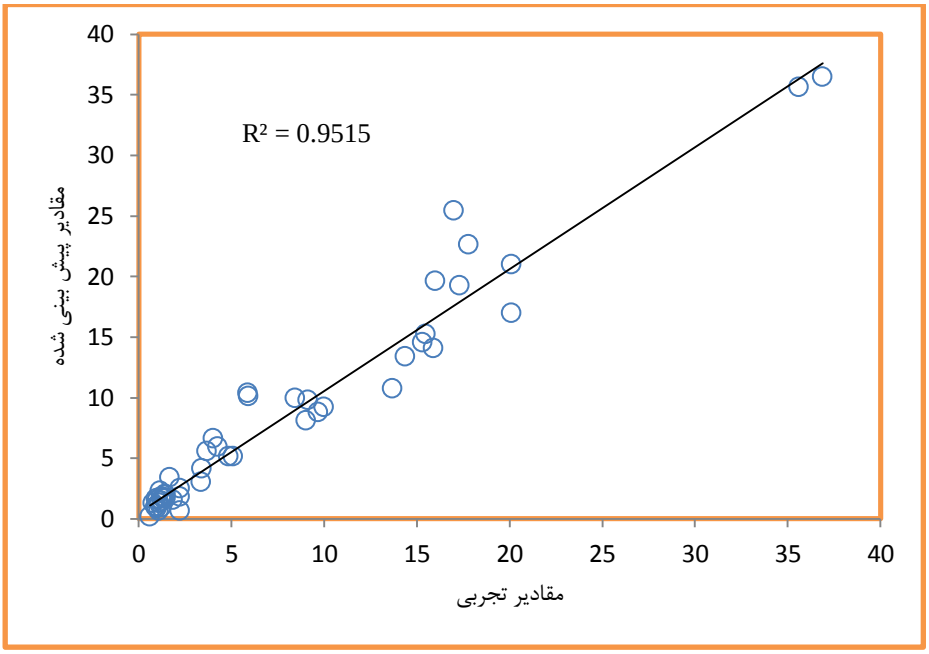


شکل (۳-۱۲)- نمودار مقادیر پیش‌بینی شده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری تست
(الف) مدل SR-ANN (ب) مدل GA-ANN

الف)



ب)



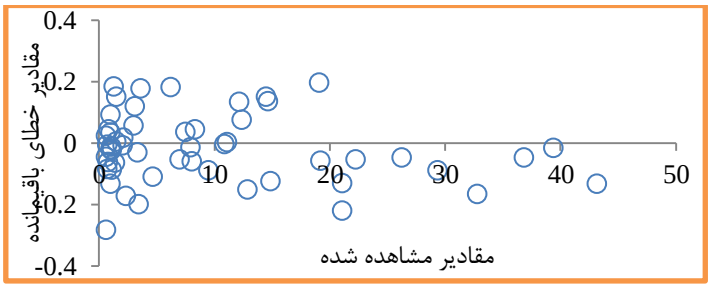
شکل (۳-۱۳)- نمودار مقادیر پیش‌بینی شده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری تست (الف) مدل SR-SVM (ب) مدل GA-SVM

همان طور که در این شکل‌ها مشاهده می‌شود، ضریب تعیین به دست آمده برای روش‌های SR-ANN، GA- ANN و SR-SVM بسیار به یک نزدیک است که این موضوع بر توافق مقادیر γ_{13}^{∞} پیش‌بینی شده توسط این سه روش نسبت به مقادیر تجربی دلالت دارد.

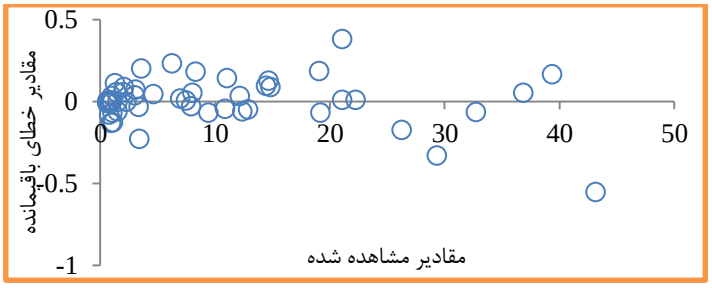
۳-۲-۷-۲- با استفاده از نمودار خطای باقیمانده

اختلاف مقادیر پیش‌بینی شده و مقادیر تجربی، خطای باقیمانده نامیده می‌شود. پراکندگی یکنواخت نقاط، حول محور افقی که بیانگر خطای باقیمانده‌ی صفر است، نشان‌دهنده‌ی آن است که خطای سیستماتیکی در مدل‌سازی وجود ندارد. نمودار خطای باقیمانده بر حسب مقادیر تجربی، برای مدل‌های ذکر شده در شکل‌های (۳-۱۴) تا (۳-۱۶)، نشان داده شده است.

الف)

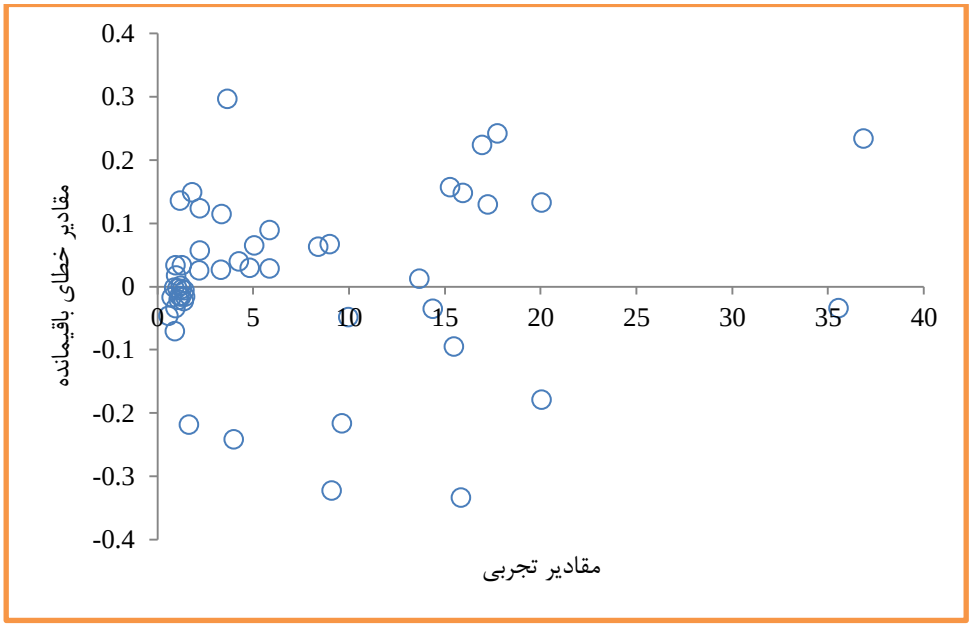


ب)

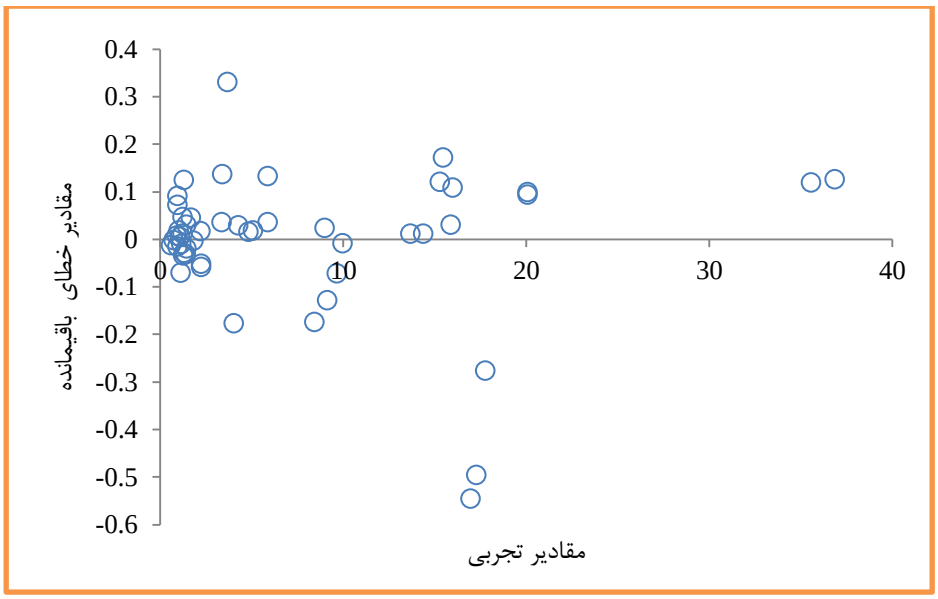


شکل (۳-۱۴)- نمودار مقادیر خطای باقیمانده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN (ب) مدل GA-ANN

الف)

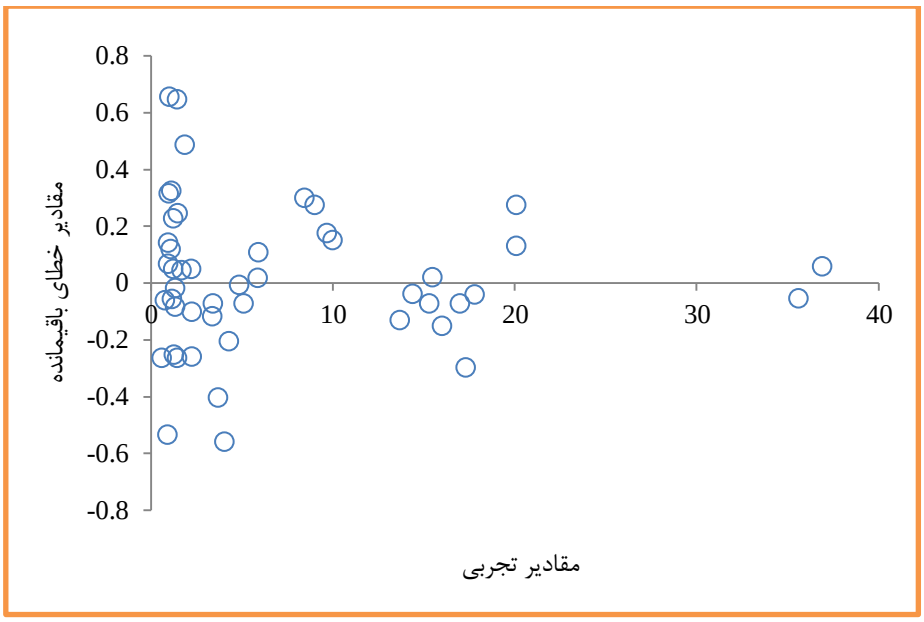


ب)

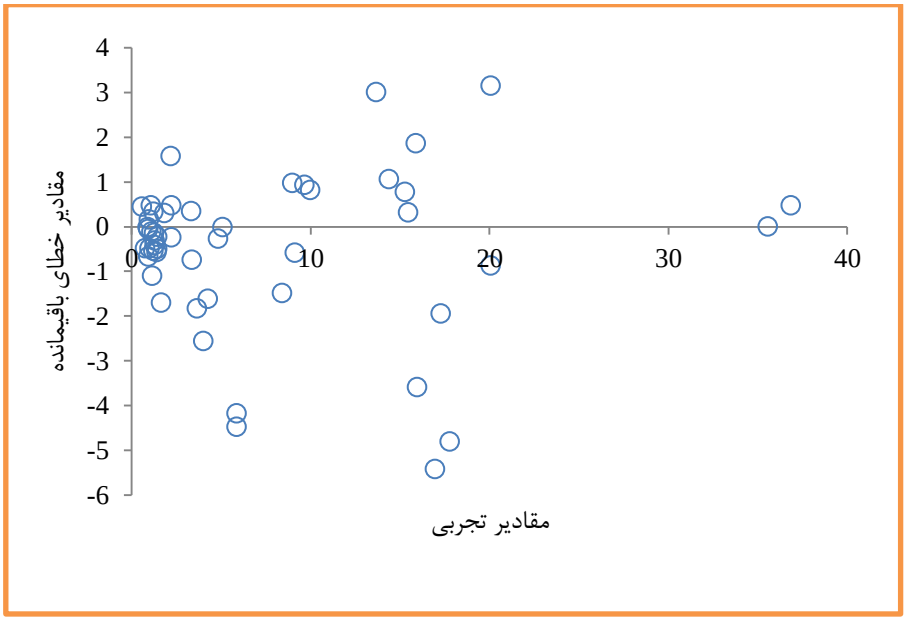


شکل (۳-۱۵)- نمودار مقادیر خطای باقیمانده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری تست (الف) مدل SR-ANN (ب) مدل GA-ANN

الف)



ب)



شکل (۳-۱۶)- نمودار مقادیر خطای باقیمانده γ_{13}^{∞} بر حسب مقادیر تجربی برای سری تست (الف) مدل SR-SVM (ب) مدل GA-SVM

مقایسه‌ی نمودارهای خطای باقیمانده نشان می‌دهد که برای سری ارزیابی مدل SR-ANN مقادیر بالاتر از ۲۰ همگی ترکیبات را با خطای مثبت پیش‌بینی می‌کند ولی مقادیر این خطا ناچیز است. بعلاوه، در مدل SR-ANN مقدار خطای مطلق باقیمانده بین مقادیر تجربی و محاسباتی از هر چهار روش ذکر شده کمتر و در بازه‌ی ۰/۳ تا ۰/۳- است. به همین دلیل روش SR-ANN برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت کارآمدتر می‌باشد.

۳-۲-۷-۳- ارزیابی شبکه ی عصبی مصنوعی با استفاده از آزمون Y تصادفی

در این روش، مقادیر متغیر وابسته به صورت تصادفی تغییر داده شد. سپس مدل QSPR با استفاده از ماتریس متغیرهای اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت و همبستگی متغیرهای مستقل با متغیرهای وابسته توسط شاخص R^2 مورد بررسی قرار گرفت. در جداول (۳-۱۵) و (۳-۱۶) مقادیر ضریب تعیین برای دو مدل SR-ANN وGA-ANN، برای سری ارزیابی و تست آورده شده است. مقادیر پایین R^2 نشان‌دهنده‌ی عدم وجود همبستگی شانسی یا وابستگی ساختاری به سری آموزش در مدل توسعه یافته توسط شبکه در مدل اصلی می‌باشد.

جدول (۳-۱۵)– مقادیر R^2 برای سری ارزیابی و تست با استفاده از آزمون Y- تصادفی در مدل SR-ANN

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 valid	۰/۲۸۷	۰/۰۷۴	۰/۲۶۱	۰/۰۸۲	۰/۰۴۵	۰/۰۱۹	۰/۰۱۵	۰/۰۵۴	۰/۱۴۵	۰/۰۰۶
R^2 test	۰/۰۰۴	۰/۰۸۲	۰/۰۰۵	۰/۳۱۱	۰/۲۶۱	۰/۰۶۷	۰/۳۴۲	۰/۰۷۵	۰/۱۰۳	۰/۰۹۱

جدول (۳-۱۶)– مقادیر R^2 برای سری ارزیابی و تست با استفاده از آزمون Y- تصادفی در مدل GA-ANN

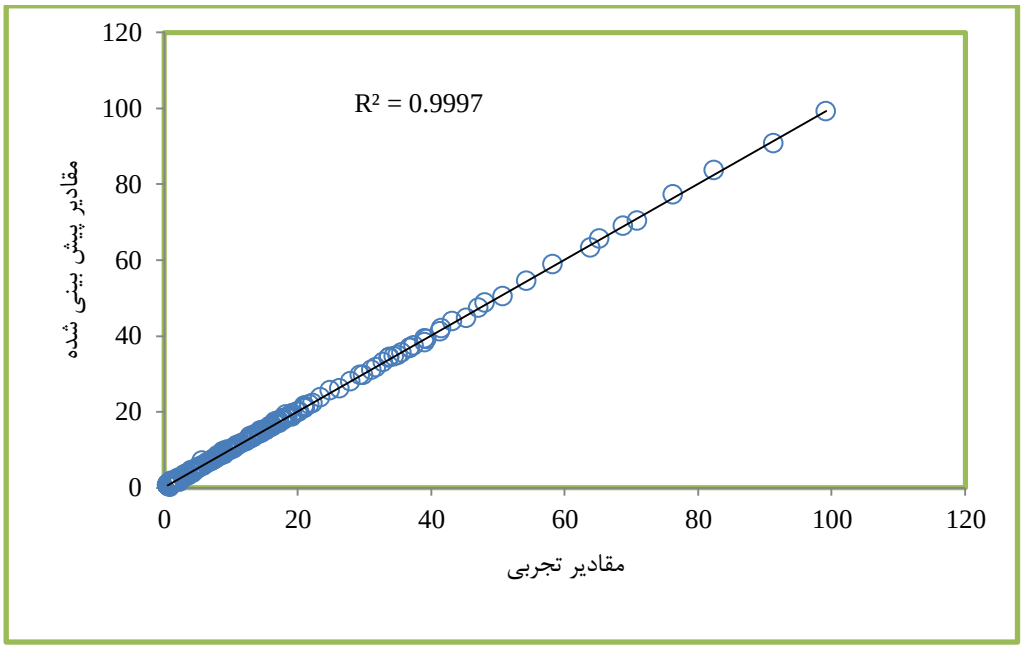
تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 valid	۰/۱۱۴	۰/۰۰۹	۰/۰۵۹	۰/۳۴۵	۰/۲۳۸	۰/۳۶۷	۰/۱۶۵	۰/۰۰۸	۰/۰۱۳	۰/۰۵۶
R^2 test	۰/۰۱۶۷	۰/۱۳۷	۰/۲۰۸	۰/۰۲۳	۰/۳۴۱	۰/۰۰۶۱	۰/۳۴۹	۰/۱۱	۰/۱۹۴	۰/۰۰۷

۳-۲-۷-۴- ارزیابی مدل‌های SR-ANN، GA-ANN، SR-SVM، و GA-SVM به روش رد مرحله‌ای تک‌تک داده‌ها

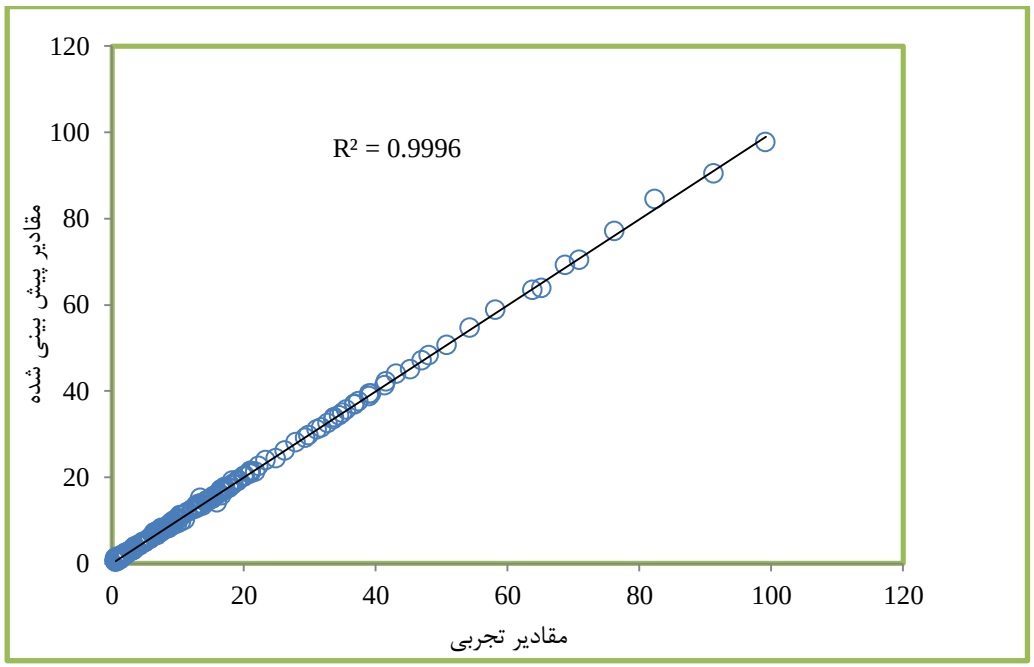
برای ارزیابی مدل‌های بهینه شده توسط دو روش ANN و SVM، روش رد مرحله‌ای تک‌تک داده‌ها مورد استفاده قرار گرفت. برای این منظور، در مدل‌های بهینه شده‌ی SR-ANN، GA-ANN، SR-SVM و GA-SVM هر بار یکی از نقاط در سری تست قرار می‌گیرد و سایر نقاط همگی در سری آموزش قرار داده می‌شوند. سپس برنامه در نقطه‌ی بهینه‌ی مدل‌ها اجرا شده و نقطه‌ی مورد نظر پیش‌بینی می‌گردد. در مرحله‌ی بعد نقطه‌ای که در سری تست قرار داشت به سری آموزش بازگردانده می‌شود و اکنون نقطه‌ی دیگری به عنوان سری تست انتخاب می‌شود و مراحل فوق دوباره تکرار می‌گردد. این کار تا جایی تکرار می‌شود که کل نقاط یک بار در سری تست قرار گیرند.

نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی داده‌ها نیز در ارزیابی چهار مدل SR-ANN، GA-ANN، SR-SVM و GA-SVM به روش رد مرحله‌ای تک‌تک داده‌ها، رسم گردید که نتایج به دست آمده در شکل (۳-۱۷) و (۳-۱۸) آورده شده است. این نتایج تعمیم پذیری بالای مدل طراحی شده توسط هر چهار مدل را بیان می‌کند.

الف)

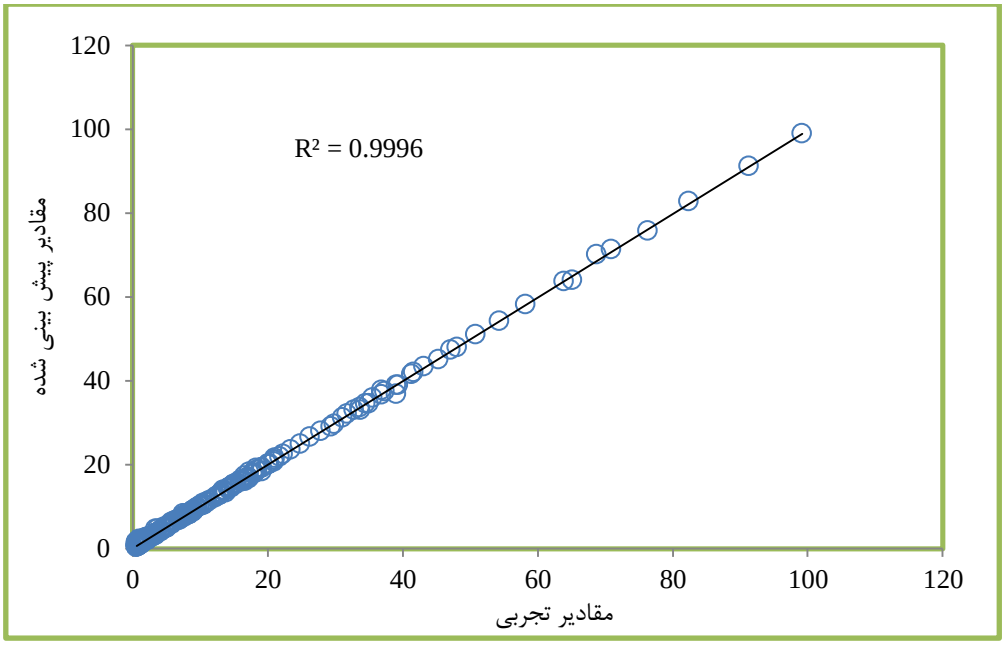


ب)

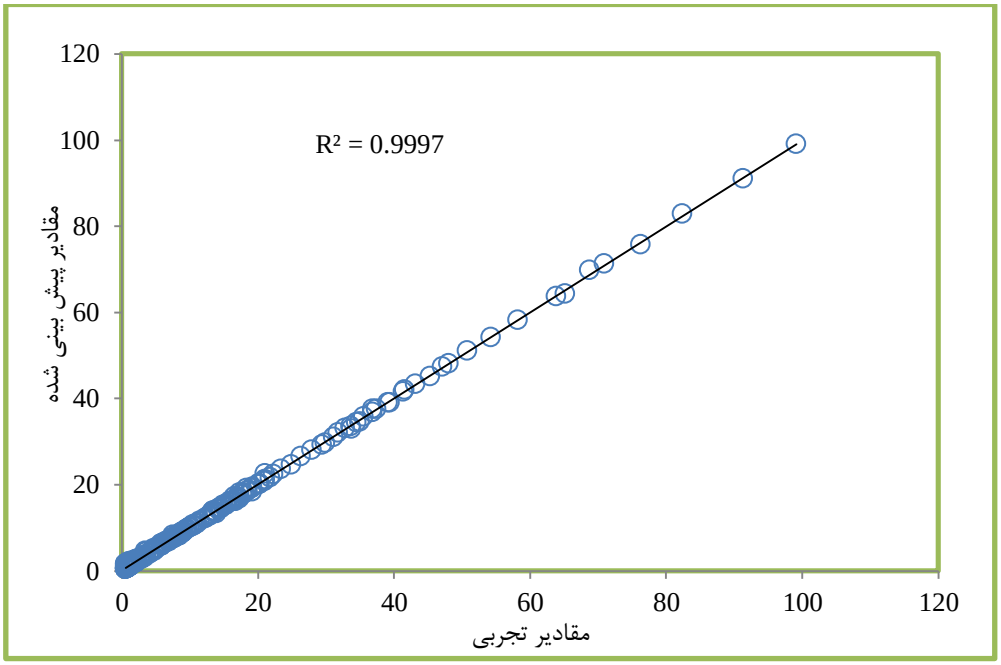


شکل (۳-۱۷)- نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی γ_{13}^{∞} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها (الف) مدل SR-ANN (ب) مدل GA-ANN

الف)

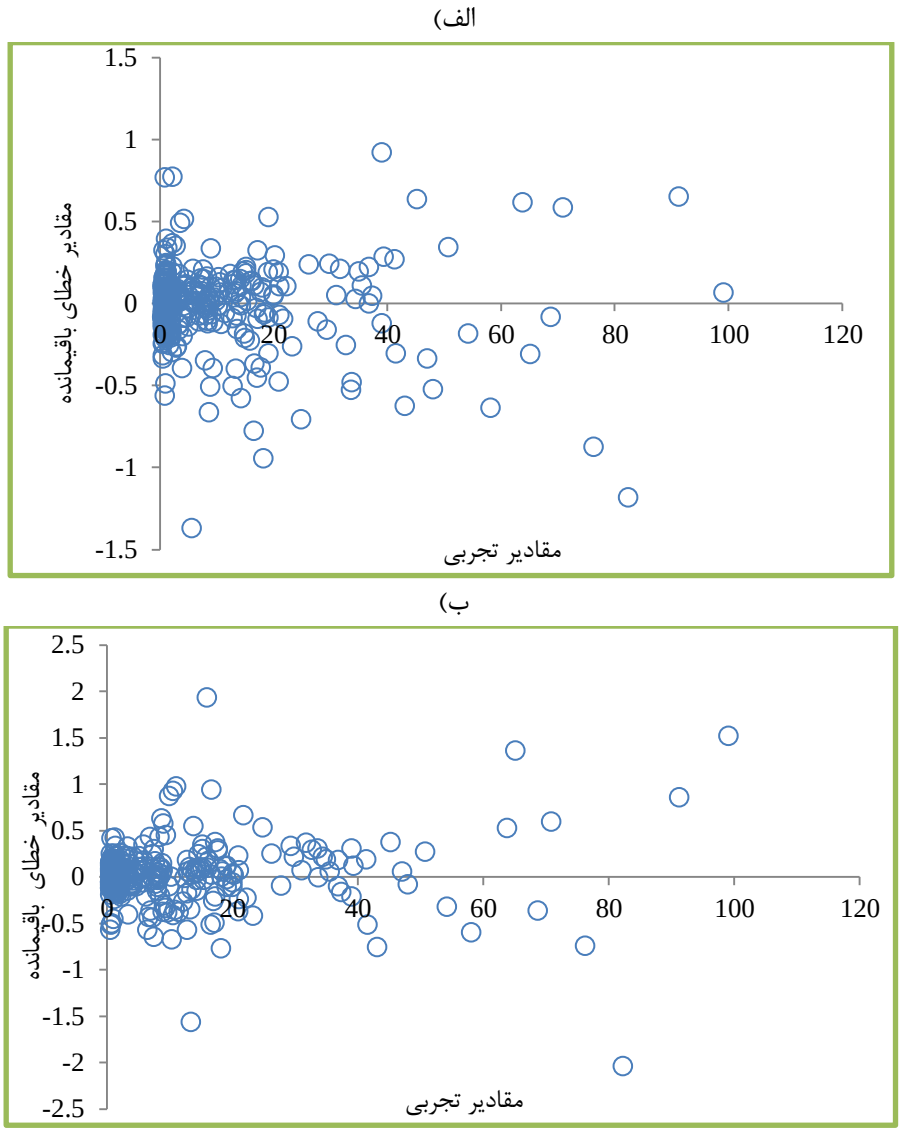


ب)



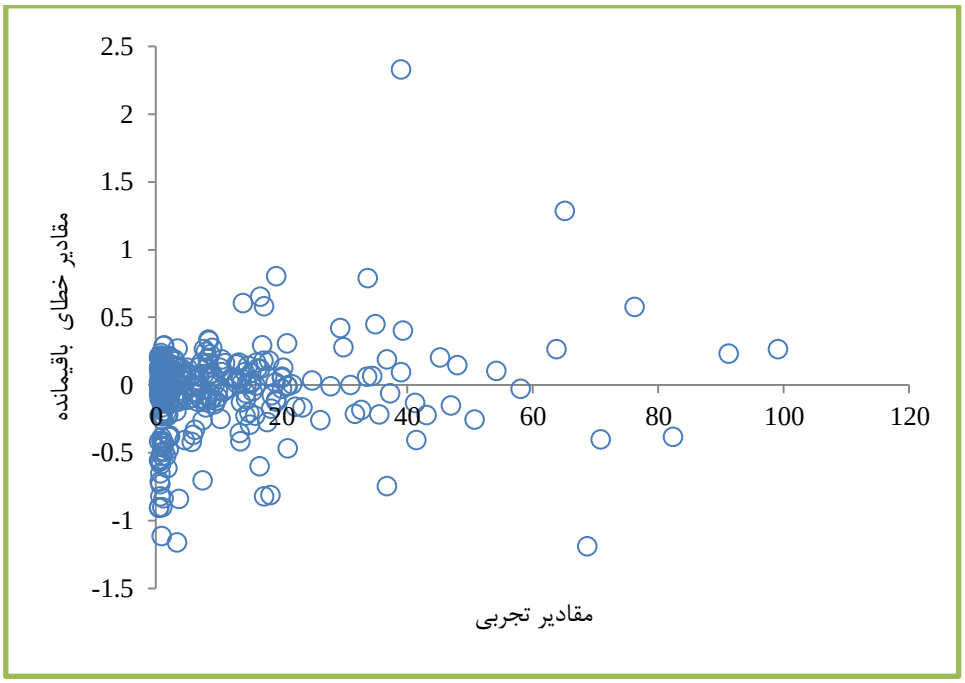
شکل (۳-۱۸)- نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی γ_{13}^{∞} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها (الف) مدل SR-SVM (ب) مدل GA-SVM

همچنین نمودار مقادیر باقی‌مانده‌ی محاسبه شده بر حسب مقـادیر تجربـی ضـریب فعالیـت در رقـت بی‌نهایت به روش رد مرحله‌ای تک‌تک داده‌ها، در ارزیابی چهار مدل در شکل(۳-۱۹) و (۳-۲۰) رسـم شـد. توزیع متقارن داده‌ها حول محور افقی نشان‌دهنده‌ی عدم وجود خطای معین می‌باشد.

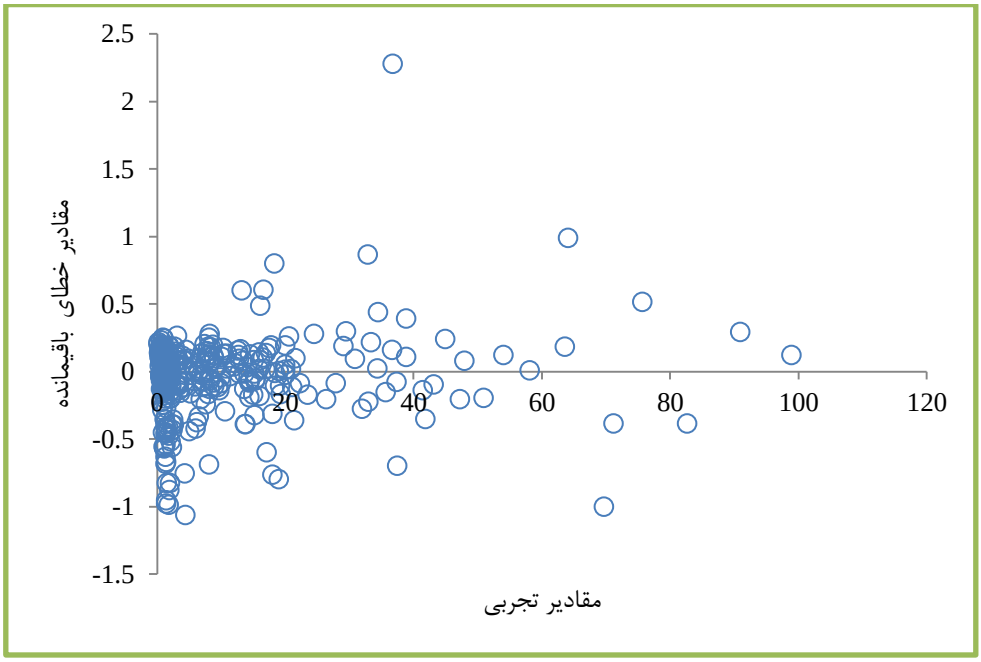


شکل (۳-۱۹)- نمودار مقادیر خطای باقیمانده بر حسب مقادیر تجربی γ_{13}^{∞} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها (الف) مدل SR-ANN (ب) مدل GA-ANN

الف)



ب)



شکل (۳-۲۰)- نمودار مقادیر خطای باقیمانده بر حسب مقادیر تجربی γ_{13}^{∞} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها (الف) مدل SR-SVM (ب) مدل GA-SVM

۳-۲-۷-۵- ارزیابی مدل‌های غیر خطی بهینه شده با استفاده از پارامترهای آماری

هفت پارامتر آماری که قبلاً در بخش (۲-۲-۸) روابط آنها ارائه شده است، جهت ارزیابی توانایی پیش‌بینی مدل‌های ساخته شده با روش‌های غیرخطی SR-ANN، GA-ANN، SR-SVM و GA-SVM برای سری‌های ارزیابی، تست و کل نقاط (LOO) به کار گرفته شد که نتایج به دست آمده در جدول (۳-۱۷) گزارش شده است.

جدول (۳-۱۷)- پارامترهای آماری روش‌های SR-ANN، GA-ANN، SR-SVM و GA-SVM

		SR-ANN	GA-ANN	SR-SVM	GA-SVM
R^2	سری ارزیابی	۰/۹۹۹۹	۰/۹۹۹۸	-	-
	سری تست	۰/۹۹۹۸	۰/۹۹۹۷	۰/۹۹۹۲	۰/۹۵۱۵
	کل نقاط (LOO)	۰/۹۹۹۷	۰/۹۹۹۶	۰/۹۹۹۶	۰/۹۹۹۷
MSE	سری ارزیابی	۰/۰۱۲۲	۰/۰۲۳۶	-	-
	سری تست	۰/۰۱۷۹	۰/۰۲۰۱	۰/۰۶۰۱	۴/۰۳۱
	کل نقاط (LOO)	۰/۰۵۹۳	۰/۰۹۷۹	۰/۰۹۹۸	۰/۱۳۵۶
AAD(%)	سری ارزیابی	۲/۶۷۷۵	۴/۸۳۳	-	-
	سری تست	۲/۴۳۰۶	۳/۴۸۳	۱۰/۱۳۵	۲۸/۸۷۴
	کل نقاط (LOO)	۲/۹۸۱۴	۵/۶۲۱۵	۱۰/۸۰۵۵	۱۱/۱۳۱۰
PRESS	سری ارزیابی	۰/۶۴۷۷	۱/۲۵۵۳	-	-
	سری تست	۰/۸۹۸۵	۱/۰۰۸	۳/۰۸۳۴	۲۰۱/۵۷
	کل نقاط (LOO)	۲۰/۹۹۳	۳۷/۶۷۵	۳۵/۳۳۵	۴۸/۰۱۱
MAE	سری ارزیابی	۰/۰۸۸۰۶	۰/۰۹۱۹۱	-	-
	سری تست	۰/۰۹۴۶	۰/۰۸۶۴	۰/۱۸۷۰	۱/۴۲۹۱
	کل نقاط (LOO)	۰/۱۵۴۹	۰/۱۷۸۷	۰/۲۰۰۶	۰/۲۱۵۵
Bias	سری ارزیابی	-۰/۱۱۶۵	۰/۳۶۰۷	-	-
	سری تست	-۰/۱۶۲۲	-۰/۰۵۶۰	۱/۳۸۶۱	-۱۴/۲۷۱۱
	کل نقاط (LOO)	-۰/۳۹۰۸	-۰/۳۶۸	-۵/۳۳۹۴	-۶/۲۲۸۴
D_{max}	سری ارزیابی	۱۸/۰۵۳	۳۱/۸۲۲	-	-
	سری تست	۱۴/۲۳	۱۶/۰۶۸	۶۴/۱۵۶	۱۰۰/۸۱۵
	کل نقاط (LOO)	۸۳/۸۲۲	۱۰۳/۹۷۸	۱۶۲/۴۲۵	۱۸۰/۷۰۱

نتایج حاصله نشان می‌دهد که مقدار ضریب تعیین (R^2) برای روش SR-ANN از سه روش دیگر بیشتر بوده و سایر پارامترهای آماری ذکر شده برای این روش از سایر روش‌ها کمتر می‌باشد، که نشان از برتری روش SR-ANN نسبت به سایر روش‌ها دارد.

۳-۲-۸- بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل منتخب SR-ANN با ضریب فعالیت در رقت بی‌نهایت

در این بخش به طور خلاصه ارتباط بین توصیف کننده‌های وارد شده در مدل و ضریب فعالیت ترکیبات، مورد بررسی قرار خواهد گرفت. با توجه به نتایج به دست آمده در مدل SR-ANN به عنوان مدل برتر، ۸ توصیف‌کننده انتخاب شد که هر کدام بیانگر خصوصیات متفاوتی از مولکول‌های مورد بررسی است. این توصیف کننده‌ها شامل: C002, T, I_{TH}, RARS, RDF030p، BELm7، RDF065u و H-046 می‌باشد.

۳-۲-۸-۱- توصیف‌کننده‌ی T

ضریب فعالیت در رقت بی‌نهایت به صورت تابعی از دما عمل می‌کند. وابستگی ضریب فعالیت را به دما می‌توان با استفاده از رابطه‌ی خطی موسوم به معادله‌ی وانتهوف[۱] به صورت زیر بیان نمود [۱].

$$\ln \gamma^{\infty} = \frac{\Delta H_1^{\infty}}{RT} - \frac{\Delta S_1^{\infty}}{R} \qquad (۲-۳)$$

بر اساس رابطه‌ی فوق اگر ΔH_1^{∞} گرماگیر باشد ضریب فعالیت در رقت بی‌نهایت با دما رابطه‌ی عکس دارد، یعنی افزایش دما باعث کاهش مقدار γ^{∞} می‌شود و اگر ΔH_1^{∞} گرماده باشد ضریب فعالیت در رقت بی‌نهایت

۱- Van 't Hoff

با دما رابطه‌ی مستقیم دارد، یعنی افزایش دما باعث افزایش مقدار γ^{∞} می‌شود. این موضوع را مـی‌تـوان بـه خوبی از مقادیر γ_{13}^{∞} ترکیبات ارائه شده در جدول (۳-۱۸) بر حسب دما مشاهده نمود.

جدول (۳-۱۸)- مثال‌هایی از مقدار توصیف‌کننده دما بر ضریب فعالیت در رقت بی‌نهایت

Solute	T, K						$\Delta H_1^{E,\infty}$
	۳۱۸/۱۵	۳۲۸/۱۵	۳۳۸/۱۵	۳۴۸/۱۵	۳۵۸/۱۵	۳۶۸/۱۵	۳۳۸/۱۵
Pentane	۱۶/۷	۱۶/۰	۱۵/۳	۱۴/۵	۱۳/۹	۱۳/۴	۴/۴
Hexane	۲۳/۴	۲۲/۳	۲۱/۱	۲۰/۱	۱۹/۲	۱۸/۳	۴/۸
3- Methylpentane	۲۱/۰	۲۰/۱	۱۹/۱	۱۸/۱	۱۷/۳	۱۶/۶	۴/۷
Pen-1-yne	۲/۱۰	۲/۱۷	۲/۲۳	۲/۲۷	۲/۳۰	۲/۳۵	-۲/۱
Hex-1-yne	۲/۸۷	۲/۹۷	۳/۰۳	۳/۰۷	۳/۱۱	۳/۱۶	-۲/۷

۳-۲-۸-۲- توصیف‌کننده‌ی ACFC [۱]

کدهای اجزای با مرکزیت اتمی (ACFC)یک کد با مرکزیت اتمی با دامنه‌ی کوتاه است که هر اتـم را با نوع اتم، انواع پیوند و انواع اتم‌های همسایه توصیف می‌کند. هر مولکول کاملاً با یک کد اجـزا کـه تعـداد اتم‌های غیر هیدروژنی است معرفی می‌شود[۴۶] . این توصیف‌کننـده‌هـا اطلاعـات شـیمیایی زیـادی را در رابطه با گروه‌های عاملی در اطراف یک اتم مرکزی و مولکول ارائه می‌دهد.

ثابت‌های اتمی آب گریزی خوزه-کریپن [۲] (توصیف‌کننده‌های چربی دوستی): چربی دوستی میزانـی از توزیع ترکیبات بین یک فاز آلی و یک فاز آبی است [۴۷] که معمولاً با ضـرایب جـداسـازی P بـه صـورت توزیع غلظت یک ترکیب در فازهای آلی و آبی از تحت شرایط تعادلی زیر تعریف می‌شود:

$$P = \frac{[C]_{org}}{[C]_{aq}} \qquad (۳-۳)$$

۱- Atom-Centred Fragment Code
۲-Ghose-Crippen

که $[C]_{org}$ و $[C]_{aq}$، غلظت‌های حل‌شده در فازهای آلی و آبی هستند.

توصیف‌کننده C002 بیان‌کننده گروه‌های CH_2R_2 می‌باشد که R گروه متصل به اتم کربنی است که در یک مولکول به عنوان اتم مرکزی در نظر گرفته می‌شود. جزء CH_2R_2 شامل پنج اتم با چهار پیوند است که شامل یک اتم کربن مرکزی و اتم‌های متصل به آن می‌باشد [۴۸]. این توصیف‌کننده که از نوع اجزای با مرکزیت اتمی است توسط خوزه-کریپن تعریف گردید. اثر متوسط این توصیف‌کننده منفی بوده بنابراین انتظار می‌رود با افزایش مقدار این توصیف‌کننده ضریب فعالیت کاهش یابد. زیرا افزایش مقدار این توصیف‌کننده به معنای افزایش تعداد گروه‌های CH_2R_2 غیر قطبی و در نتیجه افزایش خاصیت آب گریزی مولکول می‌باشد. این امر باعث می‌شود که نیروی برهمکنش بین حلال و حل‌شونده کمتر شده شود و ضریب فعالیت افزایش یابد. جدول (۳-۱۹) چگونگی ارتباط این توصیف‌کننده‌ها را با ضریب فعالیت در رقت بی‌نهایت برای سه مولکول از سری داده‌ها نشان می‌دهد.

جدول (۳-۱۹)- مثال‌هایی از مقدار توصیف‌کننده C002 بر ضریب فعالیت در رقت بی‌نهایت

نام ترکیب	مقدار C002	γ_{13}^{∞}
Pentane	۳	۱۶/۷
3- Methylpentane	۲	۲۱
2,2,4-trimethylpentane	۱	۳۹/۱

مقدار این توصیف‌کننده در گونه‌های حل‌شونده‌ای که دارای گروه CH_2R_2 در ساختارشان نمی‌باشند صفر است که می‌توان به ترکیباتی مانند: آب، استون، استونیتریل، پیریدین، زایلن‌ها، تیوفن، متانول و اتانول، ترت بوتانول اشاره نمود.

توصیف‌کننده‌ی H-046 (H متصل به کربن با هیبریداسیون sp^3 و عدد اکسایش صفر)، یکی دیگر از توصیف‌کننده‌های از نوع اجزای با مرکزیت اتمی می‌باشد [۴۹] که اولین همسایه اتم کربن، اتم هیدروژن

می‌باشد همچنین، این توصیف‌کننده نمایانگر تعداد اتم‌های هیدروژن متصل به اتم کربن با هیبریداسیون sp^3 می‌باشد. اثر متوسط این توصیف‌کننده مثبت است و نشان دهنده‌ی این است که با افزایش مقدار این توصیف‌کننده، γ_{13}^{∞} افزایش می‌یابد. در جدول (۳-۲۰) در سری آلکان‌ها افزایش مقدار این توصیف‌کننده به معنای افزایش تعداد هیدروژن‌های متصل به اتم کربن است که طول زنجیر بلندتر می‌شود و برهمکنش حلال و حل‌شونده کمتر و نتیجتاً ضریب فعالیت در رقت بی‌نهایت افزایش می‌یابد.

جدول (۳-۲۰)- مثال‌هایی از مقدار توصیف‌کننده H-046 بر ضریب فعالیت در رقت بی‌نهایت

نام ترکیب	مقدار H-046	γ_{13}^{∞}
Pentane	۱۲	۱۶/۷
Hexane	۱۴	۲۳/۴
Heptane	۱۶	۳۳/۷
Octane	۱۸	۴۸/۱
Nonane	۲۰	۶۸/۸
Decane	۲۲	۹۹/۲

مقدار این توصیف‌کننده برای گونه‌های حل‌شونده‌ای که فاقد اتم هیدروژن متصل به اتم کربن با هیبریداسیون sp^3 می‌باشد صفر است ;که ترکیباتی مانند: پیریدین، استونیتریل، بنزن، استایرن،ترت بوتانول، آب، تتراهیدروفوران، ۱و۴-دی اکسان از این دسته می‌باشند.

۳-۲-۸-۳- توصیف‌کننده‌ی BCUT[۱]

این توصیف‌کننده‌های مولکولی مقادیر ویژه‌ای[۲] از یک ماتریس ارتباطی تغییر شکل یافته به نام ماتریس بوردن[۳] هستند. ماتریس بوردن *(B)* یک گراف مولکولی تهی از هیدروژن را ارائه می‌کند که در آن

۱- Burden-CAS-University of Texas Eigenvalues
۲- Eigenvalue
۳- Burden Matrix

ماتریس B_{ii} یعنی عناصر قطری می‌توانند در ارتباط با پارامترهایی مانند اعداد اتمی، الکترونگاتیویته، قطبش‌پذیری و غیره باشند و از طرفی عناصر غیر قطری B_{ij} در ارتباط با مرتبه پیوند دو اتم متصل به هم، هستند [۳۳].

اگر عناصر قطری این ماتریس شامل اعداد اتمی اتم‌های سازنده مولکول باشند، توالی n تایی مرتبی از کوچکترین مقدار ویژه ماتریس بوردن به عنوان توصیف‌کننده‌های مولکولی با قدرت تفکیک بالا در نظر گرفته می‌شود و برای ارزیابی و مرتب‌سازی ساختارهای مولکولی از آنها استفاده می‌شود. فرض اساسی در اینجا این است که کوچکترین ویژه‌مقادیر، محتوی سهم‌هایی برای اتم‌ها و در نتیجه انعکاسی از توپولوژی کل یک مولکول می‌باشند. توصیف‌کننده BELm7 نیز، پایین‌ترین مقدار ویژه n.7 از ماتریس بوردن است که عناصر قطری آن با توجه به جرم اتمی وزن دار شده‌است، از این گروه توصیف‌کننده‌هاست. منفی بودن اثر متوسط برای آن نشان می‌دهد که با افزایش مقادیر این توصیف‌کننده ضریب‌فعالیت کاهش می‌یابد. در سری آلکان‌ها(از پنتان تا دکان) هر چه جرم اتمی مولکول‌ها افزایش می‌یابد مقدار BELm7 افزایش یافته و مولکول غیر قطبی‌تر می‌شوند و برهمکنش با حلال کمتر می‌شود و نیروهای بین مولکولی ضعیف‌تر و در نتیجه ضریب فعالیت افزایش می‌یابد.

۳-۲-۸-۴- توصیف‌کننده‌های گروه GETAWAY [۱]

این توصیف‌کننده‌ها که با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه‌اند با استفاده ماتریس قدرت نفوذ[۲] تعریف می‌شوند و بیان‌کننده ویژگی‌های هندسی و توپولوژیکی مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تاثیر مولکول[۳] (MIM) با رابطه زیر بیان می‌شود.

۱- Geometry,Topology,and Atom-Weights Assembly
۲- Leverage Matrix
۳- Molecular Infiluence Matrix

$$H = M.(M^T.M)^{-1}.M^T \qquad (۴-۳)$$

که M ماتریس مختصات اتمی، T به معنای ماتریس ترنس‌پوز و H ماتریس نفوذ مولکولی است که یک ماتریس $A \times A$ بوده که A نیز تعداد اتم‌هاست. عناصر قطری ماتریس H لوریج‌ها نام دارند که هریک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. بالطبع اتم‌های سطحی اعداد لوریج بزرگتری نسبت اتم‌های مرکزی دارند. همچنین مولکول‌های کروی اعداد لوریج پایین‌تری نسبت به مولکول‌های کشیده دارند[۴۹ و ۴۶].

توصیف‌کننده‌های GETAWAT به دو زیر مجموعه‌ی H و R تقسیم می‌شوند. توصیف‌کننده‌های H-GETAWAY به اندازه و شکل مولکول وابسته بوده به‌گونه‌ای که با افزایش اندازه اتم و فاصله‌ی اتم از مرکز مولکول، مقدار آنها افزایش می‌یابد.

گروه دیگر توصیف‌کننده‌های R-GETAWAY هستند که براساس ماتریس فاصله-نفوذ به صورت زیر تعریف می‌شوند.

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii}.h_{ij}}}{r_{ij}} \right] \qquad i \neq j \qquad (۵-۳)$$

که h_{ii} و h_{jj} لوریج‌های دو اتم i و j و r_{ij} فاصله‌ی آن دو است. عناصر قطری این ماتریس نیز صفر هستند.

این توصیف‌کننده‌ها به ساختار سه بعدی مولکول‌ها حساس بوده و در برگیرنده‌ی اطلاعاتی نظیر شکل، اندازه، تقارن مولکول و نحوه‌ی توزیع اتم‌ها در مولکول می‌باشد.

I_{TH} و RARS با استفاده از برخی عملگرهای[۱] ماتریس و مفاهیم تئوری هر دو نظریه ماتریس نفوذ مولکولی H و ماتریس فاصله-نفوذ R به دست آمده‌اند [۴۶].

توصیف‌کننده‌ی I_{TH} که نام آن محتوای اطلاعات کلی در قدرت نفوذ برابر می‌باشد، جزء توصیف‌کننده‌های GETAWAY بوده که اطلاعاتی در مورد تقارن مولکول‌ها، اندازه‌ی مولکول و پیچیدگی مولکول در اختیار

[۱] Operator

می‌گذارد. اگر همه اتم‌ها دارای نفوذ متفاوت باشند یعنی مولکول هیچ عنصر تقارنی را نشان نداده و اگر همه‌ی مولکول‌ها دارای قدرت نفوذ برابر باشند مولکول متقارن می‌باشد ($I_{TH}=0$)و مقدار این توصیف‌کننده از رابطه‌ی زیر به دست می‌آید [۴۹ و ۳۳]:

$$I_{TH} = A_0 . \log_2 A_0 \qquad (۶-۳)$$

که A_0 تعداد غیر هیدروژن می‌باشد. اثر متوسط I_{TH} منفی است یعنی با افزایش مقدار این توصیف‌کننده ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد. افزایش مقدار توصیف‌کننده زمانی است که تعداد کل عناصر افزایش یافته ، با افزایش آن میزان نفوذ افزایش یافته و در نتیجه برهمکنش حلال-حل‌شونده افزایش و ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد. جدول (۳-۲۱) چگونگی ارتباط این توصیف-کننده را با ضریب فعالیت در رقت بی‌نهایت برای چند مولکول نشان می‌دهد.

مقدار این توصیف‌کننده در گونه‌های حل‌شونده مانند سیکلوهگزان، بنزن و آب که متقارن می‌باشند یعنی دارای مقدار نفوذ یکسان می‌باشند صفر گزارش شده‌است.

جدول (۳-۲۱)- مثال‌هایی از اثر توصیف‌کننده I_{TH} بر ضریب فعالیت در رقت بی‌نهایت

نام ترکیب	مقدار I_{TH}	γ_{13}^{∞}
Hexane	۹/۵۱	۲۳/۴
Cyclohexane	۰	۱۱/۱
Methylcyclohexane	۱۲	۱۵/۹

توصیف‌کننده RARS، بر پایه‌ی مجموع ردیف‌های ماتریس فاصله-نفوذ که توسط وزن اتمی محاسبه شده‌اند .

$$RARS = 1/A \sum_{i=1}^{A} \sum_{j=1}^{A} \frac{\sqrt{h_{ii} . h_{jj}}}{r_{ij}} = 1/A \sum_{i=1}^{A} VS_i(\mathbf{R}) \qquad (۷-۳)$$

که A تعداد اتم‌ها، VS_i نشان دهنده‌ی مجموع ردیف‌های فاصله-نفوذ ماتریس R می‌باشد .

این رمزگذاری‌ها مربوط به اطلاعات مفیدی در مورد حضور گروه‌های استخلافی، اجزاء مهم در مولکول می‌باشد. همچنین این توصیف‌کننده به اندازه مولکول و کنفورماسیون پیچیدگی مولکول مرتبط بوده و مقدار این توصیف‌کننده با افزایش اندازه مولکول، کاهش می‌یابد. جدول (۳-۲۲) چگونگی ارتباط این توصیف‌کننده با خاصیت مورد نظر را برای چندین مولکول نشان می‌دهد.

جدول (۳-۲۲)- مثال‌هایی از اثر توصیف‌کننده RARS بر ضریب فعالیت در رقت بی‌نهایت

نام ترکیب	مقدار RARS	γ_{13}^{∞}
Methanol	۱/۱۷۲	۰/۶۹۹
Ethanol	۱/۱۲۰	۰/۹۶۷
Propan-1-ol	۱/۰۶۴	۱/۱۴
Butan-1-ol	۰/۹۸۷	۱/۳۴

اثر متوسط این توصیف‌کننده منفی بوده، به طوریکه با افزایش مقدار این توصیف‌کننده، ضریب فعالیت کاهش می‌یابد. اعداد جدول فوق نشان می‌دهد که با افزایش اندازه مولکول، ممانعت فضایی بیشتر می‌شود، در نتیجه برهمکنش بین حلال و حل‌شونده کاهش و ضریب فعالیت افزایش می‌یابد.

۳-۲-۸-۵- توصیف‌کننده‌های RDF

می‌توان گفت RDF یا تابع توزیع شعاعی مربوط به یک دسته از اتم‌ها، معادل توزیع احتمال یافتن یک اتم در فضای کروی به شعاع R است. رابطه (۳-۴) نحوه محاسبه تابع RDF را نشان می‌دهد.

$$g(R) = f.\sum_{i}^{N-1}\sum_{j>i}^{N} A_i.A_j.e^{-B.(R-r_{ij})^2} \qquad (۳-۸)$$

که f یک فاکتور مقیاس و N تعداد اتم‌های مولکول است. همچنین، r_{ij} فاصله‌ی بین دو اتم i و j و A یک ویژگی اتمی (وزن دار نشده u، جرم اتمی m، حجم واندروالس v، الکترونگاتیویته e، قطبش‌پذیری p) است.

β یک فاکتور تسهیل‌کننده است که توزیع احتمال فاصله بین اتمی را مشخص می‌کند و می‌توان از آن به فاکتور دما برای تعریف جنبش اتمی تعبیر کرد. $g(R)$ در نقاط گسسته‌ای با فواصل معین محاسبه و با مجموعه‌ای از کدهای RDF با ویژگی‌های مختلف اتمی می‌توان ساختار سه بعدی یک مولکول را به طور واضح توصیف کرد. این توصیف‌کننده‌ها اهمیت توزیع اتم‌ها در مولکول روی ضریب‌فعالیت آن را نشان می‌دهد. همچنین این توصیف‌کننده علاوه بر فاصله بین اتمی، اطلاعاتی با ارزشی در مورد فاصله پیوند، انواع حلقه، سیستم‌های مسطح و غیر مسطح و نوع اتم‌ها فراهم می‌کند [۳۱].

از این گروه، توصیف‌کننده‌های RDF065u و RDF030p در مدل ارائه شده برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت انتخاب شده‌اند. اندیس u نشان می‌دهد که این توصیف‌کننده مربوطه با ویژگی خاصی از اتم‌ها محاسبه نشده‌اند و اندیس p بیان می‌دارد که این توصیف‌کننده براساس قطبش‌پذیری اتمی وزن‌دار شده‌است. اثر متوسط برای هر دو توصیف‌کننده ذکر شده مثبت می‌باشد که این به معنای این است که با افزایش این مقادیر توصیف‌کننده، ضریب فعالیت نیز افزایش می‌یابد.

۳-۲-۹- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه عصبی

میزان مشارکت توصیفگرهای منتخب به صورت زیر تعیین شد:

۱- توصیف‌کننده‌ی مورد نظر به همراه اوزان مربوطه اش از شبکه بهینه شده حذف گردید.

۲- مقدار متغیر وابسته با استفاده از توصیف‌کننده‌های باقیمانده برای هر یک از ترکیبات سری ارزیابی پیش‌بینی گردید.

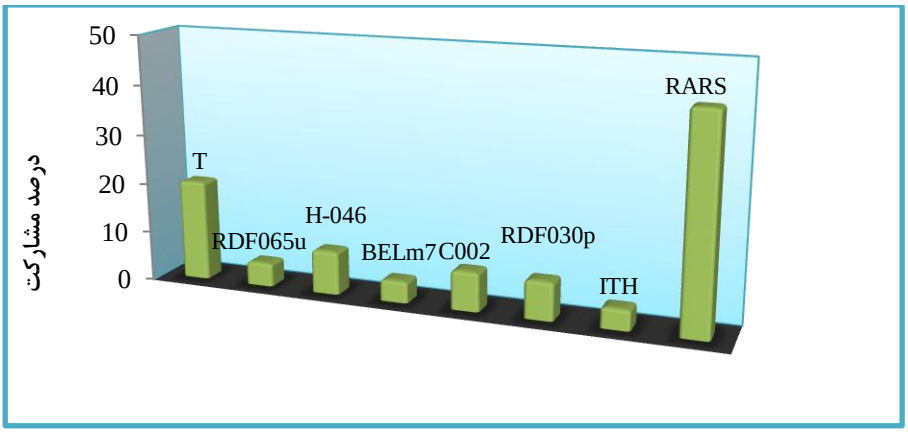
۳- میانگین خطای مطلق (MAE) حاصل از ترکیبات سری ارزیابی محاسبه شد.

۴- مراحل ۱تا ۳ برای دیگر توصیف‌کننده‌های منتخب نیز تکرار شد.

۵- سرانجام درصد مشارکت هر توصیف‌کننده توسط رابطه (۳-۲) برآورد شد [۴۳].

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^{N} \Delta m_i} \qquad (۹-۳)$$

که در این رابطه c_i درصد مشارکت توصیف‌کننده حذف شده iام، N تعداد توصیف‌کننده‌های مدل و Δm_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده‌ی iام را نشان می‌دهد. که بر این اساس درصد مشارکت توصیف‌کننده‌های منتخب در ترکیبات مورد بررسی به شکل زیر می‌باشد.



شکل(۳-۲۱)- درصد مشارکت توصیف‌کننده‌ها در مدل بهینه (SR-ANN)

براساس شکل فوق، توصیف‌کننده RARS بیشترین اثر مشارکت را دارا می‌باشد. با توجه به اینکه این توصیف‌کنندها به اندازه و کنفورماسیون پیچیدگی مولکول وابسته‌اند. این عامل یعنی افزایش اندازه مولکول روی ضریب فعالیت در رقت بی‌نهایت تأثیر معکوس دارد، می‌توان درصد مشارکت این توصیف-کننده را توجیه کرد.

۳-۳- نتیجه‌گیری

در این پایان‌نامه از روش‌های کمومتریکس برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞})، ترکیبات آلی و آب در محیط مایع یونی، ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید [BMPYR][TCM]، در ۶ دمای مختلف استفاده شده‌است. از آنجایی که اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت به صورت تجربی با صرف هزینه، زمان و پیچیدگی زیاد همراه است، دست‌یابی به مقادیر (γ_{13}^{∞})، با روش‌های تجربی مقرون به صرفه نیست. بنابراین، پیش‌بینی آن با استفاده از کمومتریکس از اهمیت بالایی برخوردار است.

برای انتخاب توصیف‌کننده‌های مناسب از روش SR و GA استفاده شد. سپس این توصیف‌کننده‌ها برای مدل‌سازی غیرخطی مورد استفاده قرار گرفت.

نتایج جدول (۳-۱۷) نشان می‌دهد که روش GA-SVM روش مناسبی برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت نیست. از بین دو روش غیرخطی نیز، SR-ANN با مقدار ضریب تعیین بیشتر و مقدار MSE کمتر، نشان می‌دهد که این مدل قادر به پیش‌گویی بهتر مقادیر (γ_{13}^{∞})، برای ترکیبات آلی و آب می‌باشد. همچنین مقایسه‌ی نمودارهای خطای باقیمانده نشان می‌دهد که مقدار خطای مطلق بین مقادیر تجربی و مقادیر محاسباتی برای سری تست با روش غیرخطی SR-ANN بازه‌ی محدودتری دارد و این به معنای خطای کمتر این روش است. در نتیجه شبکه‌ی عصبی مصنوعی روش کارآمدی برای پیش‌بینی (γ_{13}^{∞})، است.

۳-۴- آینده‌نگری

نتایج حاصل از پیش‌بینی توسط مدل‌سازی شبکه‌ی عصبی مصنوعی ترکیبات آلی و آب نشان داد که می‌توان از این روش برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت نیز استفاده نمود.

برای انتخاب توصیف‌کننده‌ی مناسب و معتبر می‌توان از سایر روش‌ها مثل روش‌های الگوریتم جستجوی ممنوع[۱] ، الگوریتم اجتماع مورچگان[۲] ، و الگوریتم بهینه‌سازی جمعیت ذره‌ها[۳] استفاده نمود و نتایج را با این کار مقایسه کرد.

برای مدل‌سازی می‌توان از روش‌های تابع پایه‌ی شعاعی(RBF)[۴] ، حداقل مربعات ماشین‌های بردار پشتیبان (LS-SVM)[۵] و سیستم استنتاج فازی-عصبی سازگار (ANFIS)[۶] استفاده کرد.

۱- Tabu Search
۲- Ant Colony Optimization
۳- Honey-bee Matikg Optimization Algorithm
٤- Radial basis function
۵- Least Square Support Vector Machines
۶- Adaptive neuron-fuzzy inference system

پیوست

جدول (پ-۱)- نام ترکیبات، دماها و ضریب فعالیت در رقت بی‌نهایت سری داده‌ها

شماره	نام	دما	ضریب فعالیت در رقت بی‌نهایت	سری
۱	Pentane	۳۱۸/۱۵	۱۶/۷	train
۲	Hexane	۳۱۸/۱۵	۲۳/۴	train
۳	3-Methylpentane	۳۱۸/۱۵	۲۱/۰	train
۴	2,2-Dimethylbutane	۳۱۸/۱۵	۲۱/۱	valid
۵	Heptane	۳۱۸/۱۵	۳۳/۷	test
۶	Octane	۳۱۸/۱۵	۴۸/۱	train
۷	2,2,4-Trimethylpentane	۳۱۸/۱۵	۳۹/۱	train
۸	Nonane	۳۱۸/۱۵	۶۸/۸	train
۹	Decane	۳۱۸/۱۵	۹۹/۲	train
۱۰	Cyclopentane	۳۱۸/۱۵	۷/۵۳	train
۱۱	Cyclohexane	۳۱۸/۱۵	۱۱/۱	valid
۱۲	Methylcyclohexane	۳۱۸/۱۵	۱۵/۹	test
۱۳	Cycloheptane	۳۱۸/۱۵	۱۳/۵	train
۱۴	Cyclooctane	۳۱۸/۱۵	۱۷/۲	train
۱۵	Pente-1-ne	۳۱۸/۱۵	۷/۵۱	valid
۱۶	Hex-1-ene	۳۱۸/۱۵	۱۰/۶	train
۱۷	Cyclohexene	۳۱۸/۱۵	۵/۱۶	train
۱۸	Hept-1-ene	۳۱۸/۱۵	۱۵/۰	train
۱۹	Oct-1-ene	۳۱۸/۱۵	۲۱/۸	train
۲۰	Dec-1-ene	۳۱۸/۱۵	۴۳/۲	valid
۲۱	Pen-1-yne	۳۱۸/۱۵	۲/۱۰	train
۲۲	Hex-1-yne	۳۱۸/۱۵	۲/۸۷	train
۲۳	Hept-1-yne	۳۱۸/۱۵	۴/۰۳	test
۲۴	Oct-1-yne	۳۱۸/۱۵	۵/۷۳	train
۲۵	Benzene	۳۱۸/۱۵	۱/۰۲	test
۲۶	Toluene	۳۱۸/۱۵	۱/۴۳	test
۲۷	Ethylbenzene	۳۱۸/۱۵	۲/۱۱	train
۲۸	o-Xylene	۳۱۸/۱۵	۱/۷۶	train
۲۹	m-Xylene	۳۱۸/۱۸	۲/۰۶	valid

ادامه جدول (پ-۱)

۳۰	p-Xylene	۳۱۸/۱۵	۲/۰۱	train
۳۱	Styrene	۳۱۸/۱۵	۱/۱۲	train
۳۲	Methanol	۳۱۸/۱۵	۰/۶۶۹	train
۳۳	Ethanol	۳۱۸/۱۵	۰/۷۶۹	test
۳۴	Propan-1-ol	۳۱۸/۱۵	۱/۱۴	train
۳۵	Propan-2-ol	۳۱۸/۱۵	۱/۲۰	train
۳۶	Butan-1-ol	۳۱۸/۱۵	۱/۴۳	train
۳۷	Butan-2-ol	۳۱۸/۱۵	۱/۳۴	train
۳۸	2-Methyl-propan-1-ol	۳۱۸/۱۵	۱/۴۲	train
۳۹	tert-Butanol	۳۱۸/۱۵	۱/۳۰	valid
۴۰	Water	۳۱۸/۱۵	۰/۹۷۳	train
۴۱	Thiophene	۳۱۸/۱۵	۰/۷۰۹	train
۴۲	Tetrahydrofuran	۳۱۸/۱۵	۰/۸۴۸	train
۴۳	1,4-Dioxane	۳۱۸/۱۵	۰/۶۴۲	valid
۴۴	tert-Butyl methyl ether	۳۱۸/۱۵	۳/۴۵	valid
۴۵	Ethyl tert-butyl ether	۳۱۸/۱۵	۸/۴۴	test
۴۶	Diethyl ether	۳۱۸/۱۵	۳/۳۴	train
۴۷	Di-n-propyl ether	۳۱۸/۱۵	۸/۴۳	train
۴۸	Di-iso-propyl ether	۳۱۸/۱۵	۸/۹۹	train
۴۹	Di-n-butyl ether	۳۱۸/۱۵	۱۷/۳	test
۵۰	Acetone	۳۱۸/۱۵	۰/۶۴۲	train
۵۱	Pentan-2-one	۳۱۸/۱۵	۱/۰۳	train
۵۲	Pentan-3-one	۳۱۸/۱۵	۱/۰۰	train
۵۳	Methyl acetate	۳۱۸/۱۵	۱/۰۰	valid
۵۴	Ethyl acetate	۳۱۸/۱۵	۱/۴۱	train
۵۵	Methyl propanoate	۳۱۸/۱۵	۱/۲۴	train
۵۶	Methyl butanoate	۳۱۸/۱۵	۱/۶۹	test
۵۷	Butanal	۳۱۸/۱۵	۰/۹۰۷	train
۵۸	Acetonitrile	۳۱۸/۱۵	۰/۵۵۸	train
۵۹	Pyridine	۳۱۸/۱۵	۰/۵۶۷	train
۶۰	Pentane	۳۲۸/۱۵	۱۶/۰	test
۶۱	Hexane	۳۲۸/۱۵	۲۲/۳	valid
۶۲	3-Methylpentane	۳۲۸/۱۵	۲۰/۱	test
۶۳	2,2-Dimethylbutane	۳۲۸/۱۵	۲۰/۳	train

ادامه جدول (پ-۱)

۶۴	Heptane	۳۲۸/۱۵	۳۱/۸	train
۶۵	Octane	۳۲۸/۱۵	۴۵/۳	train
۶۶	2,2,4-Trimethylpentane	۳۲۸/۱۵	۳۶/۹	valid
۶۷	Nonane	۳۲۸/۱۵	۶۳/۹	train
۶۸	Decane	۳۲۸/۱۵	۹۱/۳	train
۶۹	Cyclopentane	۳۲۸/۱۵	۷/۲۹	train
۷۰	Cyclohexane	۳۲۸/۱۵	۱۰/۶	train
۷۱	Methylcyclohexane	۳۲۸/۱۵	۱۵/۳	train
۷۲	Cycloheptane	۳۲۸/۱۵	۱۲/۹	train
۷۳	Cyclooctane	۳۲۸/۱۵	۱۶/۳	train
۷۴	Pente-1-ne	۳۲۸/۱۵	۷/۴۰	train
۷۵	Hex-1-ene	۳۲۸/۱۵	۱۰/۳	train
۷۶	Cyclohexene	۳۲۸/۱۵	۵/۱۰	test
۷۷	Hept-1-ene	۳۲۸/۱۵	۱۴/۷	valid
۷۸	Oct-1-ene	۳۲۸/۱۵	۲۱/۰	train
۷۹	Dec-1-ene	۳۲۸/۱۵	۴۱/۴	train
۸۰	Pen-1-yne	۳۲۸/۱۵	۲/۱۷	train
۸۱	Hex-1-yne	۳۲۸/۱۵	۲/۹۷	train
۸۲	Hept-1-yne	۳۲۸/۱۵	۴/۱۵	train
۸۳	Oct-1-yne	۳۲۸/۱۵	۵/۸۵	train
۸۴	Benzene	۳۲۸/۱۵	۱/۰۴	train
۸۵	Toluene	۳۲۸/۱۵	۱/۴۷	test
۸۶	Ethylbenzene	۳۲۸/۱۵	۲/۱۵	train
۸۷	o-Xylene	۳۲۸/۱۵	۱/۸۱	train
۸۸	m-Xylene	۳۲۸/۱۵	۲/۱۲	valid
۸۹	p-Xylene	۳۲۸/۱۵	۲/۰۶	train
۹۰	Styrene	۳۲۸/۱۵	۱/۱۶	train
۹۱	Methanol	۳۲۸/۱۵	۰/۶۷۰	train
۹۲	Ethanol	۳۲۸/۱۵	۰/۹۱۸	train
۹۳	Propan-1-ol	۳۲۸/۱۵	۱/۰۸	train
۹۴	Propan-2-ol	۳۲۸/۱۵	۱/۱۳	train
۹۵	Butan-1-ol	۳۲۸/۱۵	۱/۳۴	train
۹۶	Butan-2-ol	۳۲۸/۱۵	۱/۲۷	train

ادامه جدول (پ-۱)

۹۷	2-Methyl-propan-1-ol	۳۲۸/۱۵	۱/۳۲	test
۹۸	tert-Butanol	۳۲۸/۱۵	۱/۲۳	test
۹۹	Water	۳۲۸/۱۵	۰/۹۲۲	train
۱۰۰	Thiophene	۳۲۸/۱۵	۰/۷۳۰	valid
۱۰۱	Tetrahydrofuran	۳۲۸/۱۵	۰/۸۶۸	train
۱۰۲	1,4-Dioxane	۳۲۸/۱۵	۰/۶۶۹	train
۱۰۳	tert-Butyl methyl ether	۳۲۸/۱۵	۳/۵۲	train
۱۰۴	Ethyl tert-butyl ether	۳۲۸/۱۵	۸/۳۹	train
۱۰۵	Diethyl ether	۳۲۸/۱۵	۳/۳۷	valid
۱۰۶	Di-n-propyl ether	۳۲۸/۱۵	۸/۳۳	valid
۱۰۷	Di-iso-propyl ether	۳۲۸/۱۵	۹/۰۲	test
۱۰۸	Di-n-butyl ether	۳۲۸/۱۵	۱۶/۷	train
۱۰۹	Acetone	۳۲۸/۱۵	۰/۶۳۹	train
۱۱۰	Pentan-2-one	۳۲۸/۱۵	۱/۰۵	train
۱۱۱	Pentan-3-one	۳۲۸/۱۵	۱/۰۳	valid
۱۱۲	Methyl acetate	۳۲۸/۱۵	۱/۰۲	train
۱۱۳	Ethyl acetate	۳۲۸/۱۵	۱/۴۳	valid
۱۱۴	Methyl propanoate	۳۲۸/۱۵	۱/۲۷	train
۱۱۵	Methyl butanoate	۳۲۸/۱۵	۱/۷۲	train
۱۱۶	Butanal	۳۲۸/۱۵	۰/۹۳۰	test
۱۱۷	Acetonitrile	۳۲۸/۱۵	۰/۵۶۳	train
۱۱۸	Pyridine	۳۲۸/۱۵	۰/۵۸۰	train
۱۱۹	Pentane	۳۳۸/۱۵	۱۵/۳	train
۱۲۰	Hexane	۳۳۸/۱۵	۲۱/۱	valid
۱۲۱	3-Methylpentane	۳۳۸/۱۵	۱۹/۱	valid
۱۲۲	2,2-Dimethylbutane	۳۳۸/۱۵	۱۹/۲	train
۱۲۳	Heptane	۳۳۸/۱۵	۲۹/۹	train
۱۲۴	Octane	۳۳۸/۱۵	۴۱/۶	train
۱۲۵	2,2,4-Trimethylpentane	۳۳۸/۱۵	۳۵/۰	train
۱۲۶	Nonane	۳۳۸/۱۵	۵۸/۲	train
۱۲۷	Decane	۳۳۸/۱۵	۸۲/۴	train
۱۲۸	Cyclopentane	۳۳۸/۱۵	۷/۰۰	valid
۱۲۹	Cyclohexane	۳۳۸/۱۵	۱۰/۰	train
۱۳۰	Methylcyclohexane	۳۳۸/۱۵	۱۴/۴	test

ادامه جدول (پ-۱)

۱۳۱	Cycloheptane	۳۳۸/۱۵	۱۲/۲	valid
۱۳۲	Cyclooctane	۳۳۸/۱۵	۱۵/۳	test
۱۳۳	Pente-1-ne	۳۳۸/۱۵	۷/۲۹	train
۱۳۴	Hex-1-ene	۳۳۸/۱۵	۱۰/۰	test
۱۳۵	Cyclohexene	۳۳۸/۱۵	۴/۹۹	train
۱۳۶	Hept-1-ene	۳۳۸/۱۵	۱۴/۲	train
۱۳۷	Oct-1-ene	۳۳۸/۱۵	۲۰/۱	train
۱۳۸	Dec-1-ene	۳۳۸/۱۵	۳۹/۴	valid
۱۳۹	Pen-1-yne	۳۳۸/۱۵	۲/۲۳	test
۱۴۰	Hex-1-yne	۳۳۸/۱۵	۳/۰۳	valid
۱۴۱	Hept-1-yne	۳۳۸/۱۵	۴/۲۱	train
۱۴۲	Oct-1-yne	۳۳۸/۱۵	۵/۸۹	test
۱۴۳	Benzene	۳۳۸/۱۵	۱/۰۷	train
۱۴۴	Toluene	۳۳۸/۱۵	۱/۵۱	train
۱۴۵	Ethylbenzene	۳۳۸/۱۵	۲/۱۹	train
۱۴۶	o-Xylene	۳۳۸/۱۵	۱/۸۵	train
۱۴۷	m-Xylene	۳۳۸/۱۵	۲/۱۷	train
۱۴۸	p-Xylene	۳۳۸/۱۵	۲/۱۲	train
۱۴۹	Styrene	۳۳۸/۱۵	۱/۲۰	train
۱۵۰	Methanol	۳۳۸/۱۵	۰/۶۴۶	train
۱۵۱	Ethanol	۳۳۸/۱۵	۰/۸۷۶	train
۱۵۲	Propan-1-ol	۳۳۸/۱۵	۱/۰۳	valid
۱۵۳	Propan-2-ol	۳۳۸/۱۵	۱/۰۸	train
۱۵۴	Butan-1-ol	۳۳۸/۱۵	۱/۲۷	test
۱۵۵	Butan-2-ol	۳۳۸/۱۵	۱/۲۱	train
۱۵۶	2-Methyl-propan-1-ol	۳۳۸/۱۵	۱/۲۴	train
۱۵۷	tert-Butanol	۳۳۸/۱۵	۱/۱۸	train
۱۵۸	Water	۳۳۸/۱۵	۰/۸۸۴	valid
۱۵۹	Thiophene	۳۳۸/۱۵	۰/۷۵۲	train
۱۶۰	Tetrahydrofuran	۳۳۸/۱۵	۰/۸۹۰	train
۱۶۱	1,4-Dioxane	۳۳۸/۱۵	۰/۶۹۷	train
۱۶۲	tert-Butyl methyl ether	۳۳۸/۱۵	۳/۵۶	train
۱۶۳	Ethyl tert-butyl ether	۳۳۸/۱۵	۸/۳۱	train
۱۶۴	Diethyl ether	۳۳۸/۱۵	۳/۳۸	train

ادامه جدول (پ-۱)

۱۶۵	Di-n-propyl ether	۳۳۸/۱۵	۸/۱۶	train
۱۶۶	Di-iso-propyl ether	۳۳۸/۱۵	۸/۹۱	train
۱۶۷	Di-n-butyl ether	۳۳۸/۱۵	۱۶/۱	train
۱۶۸	Acetone	۳۳۸/۱۵	۰/۶۵۰	valid
۱۶۹	Pentan-2-one	۳۳۸/۱۵	۱/۰۸	train
۱۷۰	Pentan-3-one	۳۳۸/۱۵	۱/۰۶	train
۱۷۱	Methyl acetate	۳۳۸/۱۵	۱/۰۴	train
۱۷۲	Ethyl acetate	۳۳۸/۱۵	۱/۴۶	test
۱۷۳	Methyl propanoate	۳۳۸/۱۵	۱/۳۰	train
۱۷۴	Methyl butanoate	۳۳۸/۱۵	۱/۷۶	train
۱۷۵	Butanal	۳۳۸/۱۵	۰/۹۴۹	train
۱۷۶	Acetonitrile	۳۳۸/۱۵	۰/۵۶۶	train
۱۷۷	Pyridine	۳۳۸/۱۵	۰/۵۹۲	train
۱۷۸	Pentane	۳۴۸/۱۵	۱۴/۵	valid
۱۷۹	Hexane	۳۴۸/۱۵	۲۰/۱	test
۱۸۰	3-Methylpentane	۳۴۸/۱۵	۱۸/۱	train
۱۸۱	2,2-Dimethylbutane	۳۴۸/۱۵	۱۸/۴	train
۱۸۲	Heptane	۳۴۸/۱۵	۲۷/۹	train
۱۸۳	Octane	۳۴۸/۱۵	۳۹/۱	train
۱۸۴	2,2,4-Trimethylpentane	۳۴۸/۱۵	۳۲/۸	valid
۱۸۵	Nonane	۳۴۸/۱۵	۵۴/۳	train
۱۸۶	Decane	۳۴۸/۱۵	۷۶/۴	train
۱۸۷	Cyclopentane	۳۴۸/۱۵	۶/۷۱	train
۱۸۸	Cyclohexane	۳۴۸/۱۵	۹/۴۹	valid
۱۸۹	Methylcyclohexane	۳۴۸/۱۵	۱۳/۶	train
۱۹۰	Cycloheptane	۳۴۸/۱۵	۱۱/۵	train
۱۹۱	Cyclooctane	۳۴۸/۱۵	۱۴/۳	train
۱۹۲	Pente-1-ne	۳۴۸/۱۵	۷/۱۴	train
۱۹۳	Hex-1-ene	۳۴۸/۱۵	۹/۶۸	test
۱۹۴	Cyclohexene	۳۴۸/۱۵	۴/۸۶	test
۱۹۵	Hept-1-ene	۳۴۸/۱۵	۱۳/۷	train
۱۹۶	Oct-1-ene	۳۴۸/۱۵	۱۹/۲	valid
۱۹۷	Dec-1-ene	۳۴۸/۱۵	۳۷/۴	train

ادامه جدول (پ-۱)

۱۹۸	Pen-1-yne	۳۴۸/۱۵	۲/۲۷	train
۱۹۹	Hex-1-yne	۳۴۸/۱۵	۳/۰۷	train
۲۰۰	Hept-1-yne	۳۴۸/۱۵	۴/۲۴	train
۲۰۱	Oct-1-yne	۳۴۸/۱۵	۵/۹۰	train
۲۰۲	Benzene	۳۴۸/۱۵	۱/۰۹	test
۲۰۳	Toluene	۳۴۸/۱۵	۱/۵۵	valid
۲۰۴	Ethylbenzene	۳۴۸/۱۵	۲/۲۳	train
۲۰۵	o-Xylene	۳۴۸/۱۵	۱/۹۰	train
۲۰۶	m-Xylene	۳۴۸/۱۵	۲/۲۱	train
۲۰۷	p-Xylene	۳۴۸/۱۵	۲/۱۶	train
۲۰۸	Styrene	۳۴۸/۱۵	۱/۲۴	test
۲۰۹	Methanol	۳۴۸/۱۵	۰/۶۲۵	valid
۲۱۰	Ethanol	۳۴۸/۱۵	۰/۸۳۶	train
۲۱۱	Propan-1-ol	۳۴۸/۱۵	۰/۹۸۳	train
۲۱۲	Propan-2-ol	۳۴۸/۱۵	۱/۰۲	train
۲۱۳	Butan-1-ol	۳۴۸/۱۵	۱/۲۲	train
۲۱۴	Butan-2-ol	۳۴۸/۱۵	۱/۱۶	train
۲۱۵	2-Methyl-propan-1-ol	۳۴۸/۱۵	۱/۱۷	test
۲۱۶	tert-Butanol	۳۴۸/۱۵	۱/۱۴	train
۲۱۷	Water	۳۴۸/۱۵	۰/۸۳۳	train
۲۱۸	Thiophene	۳۴۸/۱۵	۰/۷۷۳	test
۲۱۹	Tetrahydrofuran	۳۴۸/۱۵	۰/۹۱۱	train
۲۲۰	1,4-Dioxane	۳۴۸/۱۵	۰/۷۲۱	train
۲۲۱	tert-Butyl methyl ether	۳۴۸/۱۵	۳/۵۸	train
۲۲۲	Ethyl tert-butyl ether	۳۴۸/۱۵	۸/۱۶	train
۲۲۳	Diethyl ether	۳۴۸/۱۵	۳/۳۷	test
۲۲۴	Di-n-propyl ether	۳۴۸/۱۵	۷/۹۷	valid
۲۲۵	Di-iso-propyl ether	۳۴۸/۱۵	۸/۷۹	train
۲۲۶	Di-n-butyl ether	۳۴۸/۱۵	۱۵/۵	test
۲۲۷	Acetone	۳۴۸/۱۵	۰/۶۶۲	train
۲۲۸	Pentan-2-one	۳۴۸/۱۵	۱/۱۰	train
۲۲۹	Pentan-3-one	۳۴۸/۱۵	۱/۰۸	train
۲۳۰	Methyl acetate	۳۴۸/۱۵	۱/۰۶	valid
۲۳۱	Ethyl acetate	۳۴۸/۱۵	۱/۴۷	train

ادامه جدول (پ-۱)

۲۳۲	Methyl propanoate	۳۴۸/۱۵	۱/۳۲	test
۲۳۳	Methyl butanoate	۳۴۸/۱۵	۱/۷۹	train
۲۳۴	Butanal	۳۴۸/۱۵	۰/۹۶۸	train
۲۳۵	Acetonitrile	۳۴۸/۱۵	۰/۵۶۹	train
۲۳۶	Pyridine	۳۴۸/۱۵	۰/۶۰۳	train
۲۳۷	Pentane	۳۵۸/۱۵	۱۳/۹	train
۲۳۸	Hexane	۳۵۸/۱۵	۱۹/۲	train
۲۳۹	3-Methylpentane	۳۵۸/۱۵	۱۷/۳	train
۲۴۰	2,2-Dimethylbutane	۳۵۸/۱۵	۱۷/۸	test
۲۴۱	Heptane	۳۵۸/۱۵	۲۶/۳	valid
۲۴۲	Octane	۳۵۸/۱۵	۳۶/۹	test
۲۴۳	2,2,4-Trimethylpentane	۳۵۸/۱۵	۳۱/۱	train
۲۴۴	Nonane	۳۵۸/۱۵	۵۰/۸	train
۲۴۵	Decane	۳۵۸/۱۵	۷۰/۹	train
۲۴۶	Cyclopentane	۳۵۸/۱۵	۶/۴۸	train
۲۴۷	Cyclohexane	۳۵۸/۱۵	۹/۱۰	train
۲۴۸	Methylcyclohexane	۳۵۸/۱۵	۱۳/۰	train
۲۴۹	Cycloheptane	۳۵۸/۱۵	۱۰/۹	valid
۲۵۰	Cyclooctane	۳۵۸/۱۵	۱۳/۶	train
۲۵۱	Pente-1-ne	۳۵۸/۱۵	۷/۰۰	train
۲۵۲	Hex-1-ene	۳۵۸/۱۵	۹/۴۲	train
۲۵۳	Cyclohexene	۳۵۸/۱۵	۴/۷۷	train
۲۵۴	Hept-1-ene	۳۵۸/۱۵	۱۳/۳	train
۲۵۵	Oct-1-ene	۳۵۸/۱۵	۱۸/۵	train
۲۵۶	Dec-1-ene	۳۵۸/۱۵	۳۵/۶	test
۲۵۷	Pen-1-yne	۳۵۸/۱۵	۲/۳۰	train
۲۵۸	Hex-1-yne	۳۵۸/۱۵	۳/۱۱	valid
۲۵۹	Hept-1-yne	۳۵۸/۱۵	۴/۲۸	test
۲۶۰	Oct-1-yne	۳۵۸/۱۵	۵/۹۱	test
۲۶۱	Benzene	۳۵۸/۱۵	۱/۱۱	train
۲۶۲	Toluene	۳۵۸/۱۵	۱/۵۷	train
۲۶۳	Ethylbenzene	۳۵۸/۱۵	۲/۲۵	test
۲۶۴	o-Xylene	۳۵۸/۱۵	۱/۹۳	train
۲۶۵	m-Xylene	۳۵۸/۱۵	۲/۲۵	train

ادامه جدول (پ-۱)

۲۶۶	p-Xylene	۳۵۸/۱۵	۲/۲۱	train
۲۶۷	Styrene	۳۵۸/۱۵	۱/۲۷	train
۲۶۸	Methanol	۳۵۸/۱۵	۰/۶۰۷	test
۲۶۹	Ethanol	۳۵۸/۱۵	۰/۸۰۱	valid
۲۷۰	Propan-1-ol	۳۵۸/۱۵	۰/۹۴۵	train
۲۷۱	Propan-2-ol	۳۵۸/۱۵	۰/۹۸۴	test
۲۷۲	Butan-1-ol	۳۵۸/۱۵	۱/۱۶	valid
۲۷۳	Butan-2-ol	۳۵۸/۱۵	۱/۱۲	valid
۲۷۴	2-Methyl-propan-1-ol	۳۵۸/۱۵	۱/۱۲	train
۲۷۵	tert-Butanol	۳۵۸/۱۵	۱/۱۱	train
۲۷۶	Water	۳۵۸/۱۵	۰/۸۰۱	train
۲۷۷	Thiophene	۳۵۸/۱۵	۰/۷۹۴	train
۲۷۸	Tetrahydrofuran	۳۵۸/۱۵	۰/۹۳۴	train
۲۷۹	1,4-Dioxane	۳۵۸/۱۵	۰/۷۴۵	train
۲۸۰	tert-Butyl methyl ether	۳۵۸/۱۵	۳/۶۲	valid
۲۸۱	Ethyl tert-butyl ether	۳۵۸/۱۵	۸/۰۹	valid
۲۸۲	Diethyl ether	۳۵۸/۱۵	۳/۳۹	train
۲۸۳	Di-n-propyl ether	۳۵۸/۱۵	۷/۸۲	train
۲۸۴	Di-iso-propyl ether	۳۵۸/۱۵	۸/۶۳	train
۲۸۵	Di-n-butyl ether	۳۵۸/۱۵	۱۴/۹	valid
۲۸۶	Acetone	۳۵۸/۱۵	۰/۶۷۵	train
۲۸۷	Pentan-2-one	۳۵۸/۱۵	۱/۱۲	train
۲۸۸	Pentan-3-one	۳۵۸/۱۵	۱/۱۱	train
۲۸۹	Methyl acetate	۳۵۸/۱۵	۱/۰۸	train
۲۹۰	Ethyl acetate	۳۵۸/۱۵	۱/۵۰	train
۲۹۱	Methyl propanoate	۳۵۸/۱۵	۱/۳۵	train
۲۹۲	Methyl butanoate	۳۵۸/۱۵	۱/۸۲	train
۲۹۳	Butanal	۳۵۸/۱۵	۰/۹۸۶	train
۲۹۴	Acetonitrile	۳۵۸/۱۵	۰/۵۷۲	train
۲۹۵	Pyridine	۳۵۸/۱۵	۰/۶۱۷	train
۲۹۶	Pentane	۳۶۸/۱۵	۱۳/۴	train
۲۹۷	Hexane	۳۶۸/۱۵	۱۸/۳	test
۲۹۸	3-Methylpentane	۳۶۸/۱۵	۱۶/۶	train
۲۹۹	2,2- Dimethylbutane	۳۶۸/۱۵	۱۷/۰	test

ادامه جدول (پ-۱)

۳۰۰	Heptane	۳۶۸/۱۵	۲۴/۹	train
۳۰۱	Octane	۳۶۸/۱۵	۳۴/۵	train
۳۰۲	2,2,4-Trimethylpentane	۳۶۸/۱۵	۲۹/۴	valid
۳۰۳	Nonane	۳۶۸/۱۵	۴۷/۱	train
۳۰۴	Decane	۳۶۸/۱۵	۶۵/۲	train
۳۰۵	Cyclopentane	۳۶۸/۱۵	۶/۲۷	valid
۳۰۶	Cyclohexane	۳۶۸/۱۵	۸/۷۴	train
۳۰۷	Methylcyclohexane	۳۶۸/۱۵	۱۲/۴	valid
۳۰۸	Cycloheptane	۳۶۸/۱۵	۱۰/۴	train
۳۰۹	Cyclooctane	۳۶۸/۱۵	۱۲/۹	valid
۳۱۰	Pente-1-ne	۳۶۸/۱۵	۶/۹۲	train
۳۱۱	Hex-1-ene	۳۶۸/۱۵	۹/۱۵	test
۳۱۲	Cyclohexene	۳۶۸/۱۵	۴/۶۸	valid
۳۱۳	Hept-1-ene	۳۶۸/۱۵	۱۲/۹	train
۳۱۴	Oct-1-ene	۳۶۸/۱۵	۱۷/۸	train
۳۱۵	Dec-1-ene	۳۶۸/۱۵	۳۳/۸	train
۳۱۶	Pen-1-yne	۳۶۸/۱۵	۲/۳۵	valid
۳۱۷	Hex-1-yne	۳۶۸/۱۵	۳/۱۶	train
۳۱۸	Hept-1-yne	۳۶۸/۱۵	۴/۳۲	train
۳۱۹	Oct-1-yne	۳۶۸/۱۵	۵/۹۳	train
۳۲۰	Benzene	۳۶۸/۱۵	۱/۱۳	train
۳۲۱	Toluene	۳۶۸/۱۵	۱/۶۰	train
۳۲۲	Ethylbenzene	۳۶۸/۱۵	۲/۲۹	train
۳۲۳	o-Xylene	۳۶۸/۱۵	۱/۹۷	train
۳۲۴	m-Xylene	۳۶۸/۱۵	۲/۳۰	train
۳۲۵	p-Xylene	۳۶۸/۱۵	۲/۲۵	test
۳۲۶	Styrene	۳۶۸/۱۵	۱/۳۱	train
۳۲۷	Methanol	۳۶۸/۱۵	۰/۵۹۰	train
۳۲۸	Ethanol	۳۶۸/۱۵	۰/۷۷۰	train
۳۲۹	Propan-1-ol	۳۶۸/۱۵	۰/۹۱۳	train
۳۳۰	Propan-2-ol	۳۶۸/۱۵	۰/۹۵۰	train
۳۳۱	Butan-1-ol	۳۶۸/۱۵	۱/۱۱	train
۳۳۲	Butan-2-ol	۳۶۸/۱۵	۱/۰۹	train
۳۳۳	2-Methyl-propan-1-ol	۳۶۸/۱۵	۱/۰۷	train

ادامه جدول (پ-۱)

۳۳۴	tert-Butanol	۳۶۸/۱۵	۱/۰۹	train
۳۳۵	Water	۳۶۸/۱۵	۰/۷۷۰	train
۳۳۶	Thiophene	۳۶۸/۱۵	۰/۸۲۱	train
۳۳۷	Tetrahydrofuran	۳۶۸/۱۵	۰/۹۵۲	test
۳۳۸	1,4-Dioxane	۳۶۸/۱۵	۰/۷۶۸	valid
۳۳۹	tert-Butyl methyl ether	۳۶۸/۱۵	۳/۶۹	test
۳۴۰	Ethyl tert-butyl ether	۳۶۸/۱۵	۸/۰۲	train
۳۴۱	Diethyl ether	۳۶۸/۱۵	۳/۴۰	test
۳۴۲	Di-n-propyl ether	۳۶۸/۱۵	۷/۷۳	train
۳۴۳	Di-iso-propyl ether	۳۶۸/۱۵	۸/۵۷	train
۳۴۴	Di-n-butyl ether	۳۶۸/۱۵	۱۴/۴	train
۳۴۵	Acetone	۳۶۸/۱۵	۰/۶۸۷	train
۳۴۶	Pentan-2-one	۳۶۸/۱۵	۱/۱۴	test
۳۴۷	Pentan-3-one	۳۶۸/۱۵	۱/۱۳	train
۳۴۸	Methyl acetate	۳۶۸/۱۵	۱/۰۹	train
۳۴۹	Ethyl acetate	۳۶۸/۱۵	۱/۵۳	valid
۳۵۰	Methyl propanoate	۳۶۸/۱۵	۱/۳۸	train
۳۵۱	Methyl butanoate	۳۶۸/۱۵	۱/۸۵	test
۳۵۲	Butanal	۳۶۸/۱۵	۱/۰۱	valid
۳۵۳	Acetonitrile	۳۶۸/۱۵	۰/۵۷۶	train
۳۵۴	Pyridine	۳۶۸.۱۵	۰/۶۲۷	train

جدول (پ-۲)- نتایج حاصل از ارزیابی مدل‌های SR-ANN، GA-ANN، SR-SVM، GA-SVM با استفاده از سری تست

شماره ترکیب	مقدار تجربی	مقادیر پیش‌بینی شده			
		SR-ANN	**GA-ANN**	**SR-SVM**	GA-SVM
۵	۱۵/۹	۱۶/۱۶	۱۵/۵۷	۱۶/۰۷	۱۴/۰۵
۱۲	۴/۰۳	۴/۰۶	۴/۲۷	۴/۵۹	۶/۵۹
۲۳	۱/۰۲	۱/۰۹	۰/۸۵۶	۰/۳۶۵	۱/۵۱
۲۵	۱/۴۳	۱/۵۲	۱/۵۵	۰/۷۸۳	۲/۰۰
۲۶	۰/۹۶۷	۰/۸۴۴	۰/۹۳۵	۰/۸۲۷	۱/۶۴
۳۳	۸/۴۴	۸/۴۵	۸/۴۶	۸/۱۴	۹/۹۲
۴۵	۱۷/۳	۱۷/۴۲	۱۷/۶۷	۱۷	۱۹/۲۴
۴۹	۱/۶۹	۱/۷۹	۱/۷۸	۱/۶۴	۳/۳۹
۵۶	۱۶/۰۰	۱۶/۰۰	۱۵/۹۸	۱۶/۱۵	۱۹/۵۹
۶۰	۲۰/۱	۲۰/۰۰	۱۹/۸۸	۱۹/۸۲	۲۰/۹۷
۶۲	۵/۱	۵/۰۹	۵/۰۸	۵/۱۷	۵/۱۱
۷۶	۱/۴۷	۱/۵۳	۱/۵۴	۱/۲۲	۱/۷۰
۸۵	۱/۳۲	۱/۲۷	۱/۲۷	۱/۴۰	۱/۶۴
۹۷	۱/۲۳	۱/۰۵	۱/۲۳	۱/۱۸	۰/۹۱۱
۹۸	۹/۰۲	۸/۹۶	۸/۷۶	۸/۷۴	۸/۰۵
۱۰۷	۰/۹۳	۰/۹۱۵	۱/۰۴	۱/۴۶	۰/۹۵۴
۱۱۶	۱۴/۴	۱۴/۵۰	۱۴/۳۰	۱۴/۴۴	۱۳/۳۴
۱۲۶	۱۵/۳	۱۵/۱۱	۱۴/۹۹	۱۵/۳۷	۱۴/۵۳
۱۳۰	۱۰/۰۰	۹/۹۹	۹/۹۰	۹/۸۵	۹/۱۸
۱۳۲	۲/۲۳	۲/۲۱	۲/۱۳	۲/۱۸	۰/۶۵۷
۱۳۴	۵/۸۹	۵/۷۹	۵/۸۷	۵/۸۷	۱۰/۳۷
۱۳۹	۱/۲۷	۱/۲۷	۱/۲۵	۱/۵۲	۱/۴۱
۱۴۲	۱/۴۶	۱/۴۵	۱/۳۴	۱/۷۲	۱/۹۶
۱۵۴	۲۰/۱	۲۰/۲۳	۲۰/۱۸	۱۹/۹۷	۱۶/۹۶
۱۷۲	۹/۶۸	۹/۷۲	۹/۶۷	۹/۵۰	۸/۷۶
۱۷۹	۴/۸۶	۴/۸۳	۴/۸۳	۴/۸۶	۵/۱۳
۱۹۳	۱۳/۷	۱۳/۷۱	۱۳/۷۲	۱۳/۸۳	۱۰/۶۹
۱۹۴	۱/۰۹	۱/۰۷	۱/۰۶	۰/۹۷۲	۰/۶۲۳
۱۹۵	۱/۲۴	۱/۲۳	۱/۲۱	۱/۰۱	۱/۷۸
۲۰۲	۱/۱۷	۱/۲۴	۱/۲۲	۱/۲۲	۲/۲۸

ادامه جدول (پ-۲)

۲۰۸	۰/۷۷۳	۰/۷۷۵	۰/۷۸۴	۰/۸۳۵	۱/۲۷
۲۱۵	۳/۳۷	۳/۴۳	۳/۵۰	۳/۴۸	۳/۰۳
۲۱۸	۱۵/۵	۱۵/۴۷	۱۵/۳۸	۱۵/۴۸	۱۵/۱۹
۲۲۳	۱/۳۲	۱/۳۱	۱/۱۹	۱/۳۴	۱/۷۴
۲۲۶	۱۷/۸	۱۷/۴۶	۱۸/۰۰	۱۷/۸۴	۲۲/۶۰
۲۳۲	۳۶/۹	۳۶/۶۱	۳۶/۳۷	۳۶/۸۴	۳۶/۴۴
۲۴۰	۳۵/۶	۵۳/۵۲	۳۵/۵۶	۳۵/۶۵	۳۵/۶۰
۲۴۲	۴/۲۸	۴/۳۱	۴/۱۹	۴/۴۸	۵/۹۰
۲۵۶	۵/۹۱	۵/۸۵	۵/۸۵	۵/۸۰	۱۰/۰۹
۲۵۹	۲/۲۵	۲/۱۳	۲/۰۹	۲/۵۱	۲/۴۹
۲۶۰	۰/۶۰۷	۰/۶۲۵	۰/۶۴۱	۰/۸۷۱	۰/۱۷۴
۲۶۳	۰/۹۸۴	۱/۰۴	۱/۰۱	۰/۶۷۱	۰/۸۳۲
۲۷۱	۱۷	۱۶/۵۲	۱۷/۴۳	۱۷/۰۷	۲۵/۴۲
۲۹۷	۹/۱۵	۹/۲۴	۹/۲۳	۹/۰۵	۹/۷۴
۲۹۹	۲/۲۵	۲/۳۱	۲/۴۰	۲/۳۵	۱/۷۹
۳۱۱	۰/۹۵۲	۰/۹۴۷	۰/۹۴۵	۰/۸۸۵	۱/۰۱
۳۲۵	۳/۶۹	۳/۸۲	۳/۴۹	۴/۰۹	۵/۵۳
۳۳۷	۳/۴	۳/۵۷	۳/۷۱	۳/۴۷	۴/۱۴
۳۳۹	۱/۱۴	۱/۱۷	۱/۰۶	۰/۸۱۸	۱/۲۶
۳۴۱	۱/۸۵	۱/۷۵	۱/۹۹	۱/۳۶۵	۱/۵۵

منابع

[1] Domańska U.and Lukoshko E. V. (2013) “Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-1-1methylpyrrolidinium tricyanomethanid” *J. Chem. Thermodyn.*, 66, pp 144-150.[2] Williams-Wynn M.D., Letcher T. M., Naidoo P., & Ramjugernath D. (2013) “Activity coefficients at infinite dilution of organic solutes in N-formylmorpholine-formylmorpholine and N-methylpyrrolidone from gas–liquid chromatography” *J. Chem. Thermody.*, 61(Complete), pp 154-160.

[۳] اسکوگ د ، وست د و هالر ف ، (۱۹۹۲) " مبانی شیمی تجزیه" جلد دوم، چاپ پنجم، مرکز نشر دانشگاهی، تهران، ص ۷۶۰.

[4] Everett D. H. (1965) “Effect of gas imperfection on GLC measurments: A refined method for determining activity coefficients and second virial coefficients” *Trans. Faraday Soc.*, 61, pp 1637-1645.

[5] Eike D. M., Brennecke J. F., & Maginn E. J. (2004) “Predicting infinite-dilution activity coefficients of organic solutes in ionic liquids” *Ind. Eng. Chem. Res.*, 43(4), pp 1039-1048.

[6] Jiqin Z. H. U., Yanmei Y. U., Jian C. H. E. N., & Weiyang F. E. I. (2007) “ Measurement of activity coefficients at infinite dilution for hydrocarbons in imidazolium-based ionic liquids and QSPR model” *Front. Chem. Eng.*, 1(2), pp 190-194.

[7] Xi L., Sun H., Li J., Liu H., Yao X., & Gramatica P. (2010) “Prediction of infinite-dilution activity coefficients of organic solutes in ionic liquids using temperature-dependent quantitative structure–property relationship method” *J. Chem. Eng.*, 163(3), pp 195-201.

[8] Nami F., Deyhimi F. (2011) ”Prediction of activity coefficients at infinite dilution for organic solutes in ionic liquids by artificial neural network” *J. Chem. Thermody.*,pp 43(1), 22-27.

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

[10] http://www.wordiq.com/definition/chemometrics

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

[12] Jurs P. C. (2005) ”Assessing the reliability of a QSAR models predictions” *J. Mol Graph Model* , 23(6), pp 503-523

[۱۳]. اشرفی م، (۱۳۸۹)، پایان‌نامه‌ی کارشناسی ارشد:"مطالعه‌ی ارتباط کمی ساختار-فعالیت مشتقات تیوکربامات‌ها به عنون دسته‌ی جدیدی از بازدارنده‌های غیرنوکلئوزیدی HIV "، دانشکده‌ی شیمی، دانشگاه شاهرود.

[14] Arab Chamjangali M. (2009) “Modeling of cytotoxicity data (CC_{50}) of anti-HIV 1-[5-chlorophenyl) solfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg- Marquardt artificial neural network” *J. Chem. Bio. Drug. Des*., 73, pp 456-465.

[15] Hyperchem7.0 Toronto, Canada: HyperCube Inc, http:// www.hyper.com.

[۱۶] لواین ای. ان ، اسلامپور غ، پارسافر غ، مقاری ع، نجفی ب، (۱۳۸۷) "شیمی کوانتومی" جلد سوم، چاپ اول، انتشارات فاطمی.

[17] Atkins P.W., Freidman R.S. (1996), “Molecular Quantum Mechanics”, 3rd Ed, Oxford University Press, New York.

[۱۸] عرب چم چنگلی م، (۱۳۸۶) "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیتی) مشتقات ۵-فنیل-۱-فنیل‌آمینو۱-H-ایمیدازول به وسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی.

[19] www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar_introduction.ppt

[20] Habibi A., Danandeh M. (2007), "Prediction of acidity constant for substituted acetic acids in water using artificial neural networks" *Indian J. Chem., Sect B.*, 46, pp 478- 487.

[۲۱] آزمونفر ن، (۱۳۹۲)، پایان نامه کاشناسی ارشد: "مدل‌سازی شبکه‌ی عصبی مصنوعی برای پیش بینی دانسیته‌ی سیالات تجمعی با استفاده از توصیف‌کننده‌های مولکولی"، دانشکده‌ی شیمی، دانشگاه صنعتی

[22] Pauling L. (1960), “The nature of the chemical Bond and the Stricture of Molecules and Crystals”, Vol. 18, Cornell University Press, New York.

[23] Melanie M. (1999), “An introduction to genetic algorithms”, Vol. 3, Cambridge,

Massachusetts London, England, pp 1-144.

[24] Haupt R. L., Haupt, S. E. (2004), "Practical genetic algorithms", John Wiley & Sons.

[۲۵] باهر ا، (۱۳۸۸)، پایان نامه دکتری:"پیش‌بینی و مطالعه QSPRبازداری ترکیبات آلی، دارویی و آلاینده‌ها با استفاده از رگرسیون بردارهای پشتیبان"، دانشکده شیمی، دانشگاه مازندران.

[26] Coley D. A (1999), "An introduction to genetic algorithms for scientists and engineers", Vol. 31, Singapore: World Scientific.

[27] Hassan M. N., Jurs P. C. (1990) "Prediction of gas and liquid chromatographic retention indexes of polyhalogenated biphenyls*" Anal. Chem*., 62, pp 2318-2323.

[28] Kartalopoulos S. V.& Kartakapoulos S. V. (1997), "Understanding neural networks and fuzzy logic: basic concepts and applications". Wiley-IEEE Press.

[۲۹] منهاج م، (۱۳۸۷) "مبانی شبکه‌های عصبی (هوش محاسباتی)" جلد اول، چاپ پنجم، مرکز نشر دانشگاه صنعتی امیر کبیر، تهران

[۳۰] حسن زاده سورشجاعی ح، (۱۳۸۸)، پایان‌نامه ارشد: "پیش‌بینی ویسکوزیته مخلوط چندتایی سیالات با استفاده از شبکه عصبی موجک"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[۳۱] اشرفی م، (۱۳۸۹)، پایان‌نامه ارشد: "مطالعه ارتباط کمی ساختار-فعالیت مشتقات تیوکربومات‌ها به عنوان دسته‌ی جدیدی از بازدارنده‌های غیر نوکلئوزیدی HIV"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

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

[۳۳] رضائی م، (۱۳۸۹) "مطالعه‌ی کمی ساختار-فعالیت برخی از ترکیبات سولفونانیلید به عنوان گروهی جدید از داروهای ضد سرطان و ضد ایدز برخی از ترکیبات"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[۳۴] کلانتر ز، (۱۳۸۵)، پایانامه دکتری: "پیش‌بینی خواص ترمودینامیکی و انتقالی سیالات آلی با استفاده از روش سهم گروه‌ها و شبکه‌ی عصبی موجک"، دانشکده‌ی شیمی، دانشگاه صنعتی اصفهان.

[۳۵] فردوسی م، (۱۳۸۹)، پایان‌نامه ارشد: "پیش‌بینی ثابت‌های هنری بعضی از ترکیبات آلی با استفاده از روش‌های خطی و غیرخطی QSPR"، دانشکده‌ی شیمی، دانشگاه صنعتی شاهرود.

[36] http:// foram.takdownlod.ir/threads/25029

[۳۷] شیرپور ا، (۱۳۸۸)، پایان‌نامه ارشد: "کاربرد مطالعات رابطه کمی ساختار-فعالیت برای پیش‌بینی و مدل‌سازی فعالیت دارویی برخی از داروهای سنتز شده جدید"، دانشکده‌ی شیمی، دانشگاه صنعتی شاهرود.

[38] Martínez-Ramón M.& Christodoulou C. (2005) "Support vector machines for antenna array processing and electromagnetics", Universidad Carlos III De Madrid, Spain, Morgan & Claypool, USA.

[39] Chen N, Yang J, (2004), "Support Vector Machines in chemistry", word scientific publishing Co.

[40] Abe S. (2010), "Support vector machines for pattern classification", Kobe University, Kobe, Japan.

[41] Wang, L. (2005), "Support Vector Machines: theory and applications", Vol. 177, Nanyang Technological University, School of Electrical & Electronic Engineering, Springer Berlin Heidelberg New York.

[42] Cristianini N., Shawe-Taylor J. (2000), "An introduction to support vector machines and other kernel-based learning methods", Cambridge University press, UK.

[43] Moosavi, M. (2012). "Prediction of thermodynamic properties of long chain 1-carboxylic acids and esters using a group contribution equation", *Fluid Phase Equilibria*, *316*, 122-131.

[۴۴] کریمی س، (۱۳۹۱)، پایان‌نامه ارشد: "مدل‌سازی QSPR فاکتور بازداری ترکیبات آلی بر روی فازهای ثابت تری فلوئوروپروپیل"، دانشکده‌ی شیمی، دانشگاه صنعتی شاهرود.

[45] Baumann K. & Clerk J. T. (1997b). "Computer-Assisted IR Spectra Prediction Linked Smilarity Searches for Structures and Spectra". *Anal. Chim. Acta*, 348, pp 327-343.

[46] Todeschini, R., & Consonni, V. (2009)."Molecular descriptors for chemoinformatics". John Wiley & Sons.

[47] Taylor P.J. (1990). "Hydropholic Properties of Drugs" In Quantitative Drug Design. Vol 4(Ramsden, C.A., ed.), Pergamon Press, Oxford (UK), pp. 241-249.

[48] Noorizadeh, H., Farmany, A., & Noorizadeh, M. (2011). "Quantitative structure-

retention relationships analysis of retention index of essential oils" *Química Nova*, 34(2), pp 242-249.

[49] Puzyn, T., Leszczynski, J., & Cronin, M. T. (2010), "Recent advances in QSAR studies". Jackson State University, Jackson, MS, USA.

Abstract

A quantitative structure-property relationship (QSPR) approach was employed to predict activity coefficients at infinite dilution for 58 organic compounds and water at six different temperatures in ionic liquid of [BMPYR][TCM].A large number of descriptors were calculated using Dragon software and the best calculated descriptors was selected from 18 classes of Dragon descriptors by stepwise regression (SR) and genetic algorithm based on partial least square (GA-PLS). 12 descriptors with SR and 10 descriptors with GA-PLS methods were selected. The selected variables by these methods were used as input for artificial neural network (ANN) and support vector machine (SVM) to construct of models to predict of activity coefficients at infinite dilution of these compounds. The performance of each model was investigated by test set. The obtained results show the superiority of the SR-ANN than other methods. The mean square error (MSE) and absolute average percent deviation (AAD) for the test sets of SR-ANN, GA-ANN, SR-SVM and GA-SVM were 0.0179 , 2.4306% and 0.0201 , 3.483% and 0.0601, 10.135% and 4.031, 28/874%, respectively

.

Keywords: activity coefficients at infinite dilution, Stepwise Regression (SR), Genetic Algorithm (GA), artificial neural network (ANN), support vector machine (SVM)

Shahrood University of Tchnology
Faculty of Chemistry

M.Sc. Thesis in Physical Chemistry

Quantitative structure-property relationship study of activity coefficients at infinite dilution for organic Solutes and water in the ionic liquid 1-buthyl-1-methyl pyrrolidinium tricyanomethanide

Maryam Taherzadeh

Supervisors:

Dr. Z. Kalantar

Advisor:

Dr. N. Goudarzi

Feb 2015