

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



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

مورفولینیوم تریسیانومتانید

دانشجو: حسین نعمتی

استاد راهنما:

دکتر زهرا کلانتر

استاد مشاور:

دکتر ناصر گودرزی

پایان نامه جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۹۴

تقدیم به

تجلی گاه حمایت خداوند

پدرم

مظهر لطف و مهربانی

مادرم

تقدیم به همسر عزیزم

به پاس همه مهربانی‌ها و همراهی‌هایش

و شکوفه‌های زندگیم علی‌اکبر و محمدجواد

تشکر و قدردانی

سپاس بی‌کرات 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.

اینک که با لطف خداوند متعال تدوین این پایان‌نامه را به اتمام رسانده‌ام، لازم می‌دانم از استاد برجسته و گرانقدر خانم دکتر کلاتر که با زحمات و راهنمایی‌های فراوان مرا در به پایان رسانیدن این تحقیق یاری نمودند، تشکر و قدردانی نمایم.

با تشکر فراوان از استاد عزیز، آقای دکتر گودرزی که همواره با نظرات و رهنمودهایشان مرا در انجام این پایان‌نامه یاری نمودند.

از اساتید محترم، آقای دکتر نیکوفرد و خانم دکتر موسوی که زحمت داوری این پایان‌نامه را بر عهده داشتند تشکر و قدردانی می‌نمایم.

از جناب آقای دکتر باقریان به عنوان نماینده تحصیلات تکمیلی حاضر در جلسه دفاعیه، سپاسگزاری می‌نمایم.

از خانم مریم طاهرزاده، آقای مهندس میلاد نظامی، آقای مصطفی لطفی به خاطر تمام راهنمایی‌ها و مساعدت‌هایشان در طول انجام پایان‌نامه، تشکر می‌نمایم.

در پایان از دوست عزیزم، آقای مسعود شکفته به خاطر تمام زحمات و حمایت‌هایش در دوران تحصیل تشکر و قدردانی می‌نمایم.

چکیده

ضرب فعاليت در رقت بي نهايت يكي از پارامترهاي كليدي است كه مي تواند براي انتخاب حلال موثر در فرايند جداسازي به كار رود. در اين پايان نامه، يك مطالعه ي ارتباط كمّي ساختار- خاصيت (QSPR)، براي پيش بيني ضرب فعاليت در رقت بي نهايت ۶۰ تركيب آلي مختلف و آب در محيط مايع يوني، ۱-بوتيل ۱-متيل مورفولينوم تري سيانو متانيد [BMMOR][TCM]، در ۶ دماي مختلف انجام شد. تعداد زيادي توصيف كننده توسط نرم افزارهاي هايپر كم و دراگون محاسبه شدند. بهترين توصيف كننده ها، با استفاده از دو روش رگرسيون مرحله اي (SR) و اتحاد جداسازي داده و انتخاب خاصيت (CDFS) انتخاب گرديدند. ۱۲ توصيف كننده به عنوان بهترين توصيف كننده ها توسط روش SR انتخاب شدند. همچنين ۷ توصيف كننده ي مهم مولكولي و يك متغير تجربّي (دما) توسط روش CDFS انتخاب شدند. توصيف كننده هاي انتخابي توسط دو روش به عنوان ورودّي به شبكه ي عصبي مصنوعي (ANN) و ماشين بردار پشتيبان (SVM) وارد شدند. پس از آموزش و بهينه سازي پارامترهاي ANN و SVM، عملکرد هر يك از مدل ها با استفاده از سري تست مورد ارزيابي قرار گرفت. مقدار متوسط مربعات خطا براي داده هاي سري تست از چهار مدل SR-ANN، CDFS-ANN، SR-SVM و CDFS-SVM، به ترتيب مقادير ۰/۰۵۱۵، ۰/۱۳۲۲، ۰/۰۸۳۶ و ۰/۰۹۸۸ به دست آمد. نتايج به دست آمده برتري مدل SR-ANN را نسبت به ساير مدل ها نشان داد.

كلمات كليدي: ضرب فعاليت در رقت بي نهايت، رگرسيون مرحله اي، اتحاد جداسازي داده و انتخاب خاصيت، شبكه ي عصبي مصنوعي، ماشين برداري پشتيبان.

نتایج حاصل از این پایان نامه در دو پوستر تحت عناوین

“Prediction of activity coefficients at infinite dilution for organic solutes and water in ionic liquids [BMMOR] [TCM] using artificial neural network”

“QSPR study of activity coefficient at infinite dilution for organic solutes in the ionic liquid [BMMOR] [TCM] by using CDFS strategy”

در هجدهمین کنگره شیمی سمنان در شهریور ماه ۱۳۹۴ ارائه گردید.

فهرست مطالب

فصل اول: مقدمه

مقدمه	۲
۱-۱- مایعات یونی	۳
۱-۱-۱- ساختار و خواص ترمودینامیکی مایعات یونی	۳
۱-۱-۲- کاربردهای مایعات یونی	۴
۲-۱- کروماتوگرافی	۵
۱-۲-۱- کروماتوگرافی گازی	۶
۲-۲-۱- استفاده از مایعات یونی به عنوان فاز ساکن در کروماتوگرافی گاز-مایع	۷
۱-۲-۳- اندازه گیری ضریب فعالیت در رقت بی نهایت با روش کروماتوگرافی گاز-مایع	۷
۱-۳-۲- بررسی مقدار ضریب فعالیت در رقت بی نهایت برای ترکیبات آلی در دو مایع یونی متفاوت	۸
۱-۳-۳- مروری بر کارهای گذشته برای پیش بینی ضریب فعالیت در رقت بی نهایت	۱۰
۱-۴- هدف تحقیق	۱۲

فصل دوم: کمومتریکس

مقدمه	۱۴
۱-۲- کمومتریکس	۱۴
۲-۲- ارتباط کمی ساختار-خاصیت و ساختار-فعالیت	۱۶
۲-۲-۱- رابطه کمی ساختار-فعالیت	۱۶
۲-۲-۲- رابطه کمی ساختار-خاصیت	۱۶
۲-۳- مراحل انجام QSPR	۱۷
۲-۳-۱- فراهم کردن سری داده ها	۱۸
۲-۳-۲- رسم و بهینه سازی ساختار هندسی مولکول ها	۱۸

۱۹	محاسبه توصیف‌کننده‌ها	۳-۳-۲
۲۱	حذف توصیف‌کننده‌های نامناسب	۳-۳-۴
۲۲	تقسیم بندی داده‌ها	۳-۳-۵
۲۳	انتخاب بهترین توصیف‌کننده‌ها	۳-۳-۶
۲۳	انتخاب بهترین توصیف‌کننده‌ها توسط رگرسیون مرحله‌ای	۳-۳-۶-۱
۲۴	انتخاب بهترین توصیف‌کننده‌ها به روش CDFS	۳-۳-۶-۲
۲۴	ساختن مدل	۳-۳-۷
۲۵	شبکه عصبی مصنوعی	۳-۳-۷-۱
۲۵	سیستم عصبی زیستی	۳-۳-۷-۱-الف
۲۶	ساختار شبکه‌ی عصبی مصنوعی	۳-۳-۷-۱-ب
۲۹	روش‌های آموزش	۳-۳-۷-۱-ج
۳۰	ماشین بردار پشتیبان	۳-۳-۷-۲
۳۱	طبقه بندی خطی دو کلاسه با ماشین های بردار پشتیبان	۳-۳-۷-۲-الف
۳۵	طبقه بندی غیرخطی با ماشین بردار پشتیبان	۳-۳-۷-۲-ب
۳۸	بردار پشتیبانی رگرسیونگر برای سیستم های خطی	۳-۳-۷-۲-ج
۴۰	بردار پشتیبانی رگرسیونگر برای سیستم های غیرخطی	۳-۳-۷-۲-د
۴۱	آموزش ماشین بردار رگرسیون	۳-۳-۷-۲-ه
۴۲	ارزیابی قدرت پیش بینی مدل	۳-۳-۸
۴۲	استفاده از مدل‌های آماری	۳-۳-۸-۱
۴۵	استفاده از نمودار برگشتی	۳-۳-۸-۲
۴۶	استفاده از نمودار خطای باقیمانده	۳-۳-۸-۳
۴۶	استفاده از سری تست خارجی	۳-۳-۸-۴

فصل سوم: پیش بینی ضریب فعالیت در رقت بی نهایت ترکیبات آلی و آب در محیط مایع یونی ۱- بوتیل ۱- متیل مورفولینیوم تری سیانو متانید با استفاده روش های غیرخطی

۴۸	۱-۳-۱- مراحل مدل سازی
۴۸	۱-۳-۱-۱- سری داده ها
۵۱	۱-۳-۲- رسم و بهینه سازی ساختار مولکول ها و محاسبه توصیف کننده ها
۵۱	۱-۳-۳- حذف توصیف کننده های نامناسب
۵۲	۱-۳-۴- تقسیم بندی داده ها
۵۲	۱-۳-۵- انتخاب بهترین توصیف کننده ها
۵۳	۱-۳-۵-۱- انتخاب بهترین توصیف کننده ها به روش رگرسیون مرحله ای
۵۷	۱-۳-۵-۲- انتخاب بهترین توصیف کننده ها به روش CDFS
۶۱	۱-۳-۶- مدل سازی غیرخطی با استفاده از شبکه عصبی مصنوعی
	۱-۳-۶-۱- بهینه سازی شبکه ی عصبی مصنوعی با توصیف کننده های به دست آمده از رگرسیون مرحله ای SR-ANN
	۱-۳-۶-۲- بهینه سازی شبکه ی عصبی مصنوعی با توصیف کننده های به دست آمده از روش CDFS
۷۷	۱-۳-۶-۳- مدل سازی ماشین بردار پشتیبان
۷۹	۳-۲- ارزیابی قدرت پیش بینی مدل های غیر خطی
۷۹	۳-۲-۱- با استفاده از نمودار برگشتی
۸۵	۳-۲-۲- با استفاده از نمودار خطای باقیمانده
۸۹	۳-۲-۳- ارزیابی مدل های غیرخطی بهینه شده با استفاده از پارامتر های آماری
	۳-۳- بررسی ارتباط توصیف کننده های وارد شده در مدل CDFS-ANN با ضریب فعالیت در رقت بی نهایت
۹۱	

۹۱	 Constitutional	۳-۳-۱-توصیف کننده ی گروه
۹۲	 ACFC	۳-۳-۲-توصیف کننده ی
۹۳	 RDF	۳-۳-۳-توصیف کننده ی گروه
۹۴	 2D-Autocorrelation	۳-۳-۴-توصیف کننده ی
۹۵	 GETAWAY	۳-۳-۵-توصیف کننده های
۹۶	 WHIM	۳-۳-۶-توصیف کننده های
۹۶		۳-۴-بررسی میزان مشارکت توصیف کننده های منتخب شبکه ی عصبی
۹۸		۳-۵-نتیجه گیری
۹۹		۳-۶-آینده نگری
۱۰۰		پیوست
۱۱۴		منابع

فهرست اشکال

- شکل (۱-۲) - ساختار یک نرون محاسباتی با چند ورودی ۲۷
- شکل (۲-۲) - یک شبکه‌ی عصبی مصنوعی ۲۸
- شکل (۳-۲) - خط جداساز بهینه با حداکثر مقدار حاشیه ۳۱
- شکل (۴-۲) - نمایش بردارهای پشتیبان روی ابر صفحه‌های موازی مرزی ۳۲
- شکل (۵-۲) - صفحه جداساز و حاشیه‌ها ۳۳
- شکل (۶-۲) - مجموعه داده‌های پیچیده ۳۶
- شکل (۷-۲) - داده‌های ورودی ارجاع داده شده به فضای بالاتر ۳۶
- شکل (۸-۲) - تابع حساسیت به مقدار ε ۳۹
- شکل (۱-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از SR ۶۴
- شکل (۲-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از SR ۶۵
- شکل (۳-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از SR ۶۶
- شکل (۴-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از SR ۶۷
- شکل (۵-۳) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ ۶۹

شکل (۳-۶) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از CDFS ۷۱

شکل (۳-۷) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از CDFS ۷۲

شکل (۳-۸) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین با استفاده از توصیف‌کننده‌های حاصل از CDFS ۷۳

شکل (۳-۹) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده از توصیف‌کننده‌های حاصل از CDFS ۷۴

شکل (۳-۱۰) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ ۷۶

شکل (۳-۱۱) - نمودار مقادیر پیش‌بینی $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN و (ب) مدل CDFS-ANN ۸۰

شکل (۳-۱۲) - نمودار مقادیر پیش‌بینی $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست داخلی (الف) مدل SR-ANN و (ب) مدل CDFS-ANN ۸۱

شکل (۳-۱۳) - نمودار مقادیر پیش‌بینی $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست خارجی (الف) مدل SR-ANN و (ب) مدل CDFS-ANN ۸۲

شکل (۳-۱۴) - نمودار مقادیر پیش‌بینی $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست داخلی (الف) مدل SR-SVM و (ب) مدل CDFS-SVM ۸۳

شکل (۳-۱۵)-نمودار مقادیر پیش‌بینی $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست خارجی الف) مدل SR-SVM و ب) مدل CDFS-SVM ۸۴

شکل (۳-۱۶)-نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری ارزیابی الف) مدل SR-ANN و ب) مدل CDFS-ANN ۸۶

شکل (۳-۱۷)-نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست داخلی الف) مدل SR-ANN و ب) مدل CDFS-ANN ۸۷

شکل (۳-۱۸)-نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست داخلی الف) مدل SR-SVM و ب) مدل CDFS-SVM ۸۸

شکل (۳-۱۹)- درصد مشارکت توصیف‌کننده‌ها در مدل بهینه SR-ANN ۹۷

فهرست جدول‌ها

- جدول (۱-۱)-انواع کروماتوگرافی بر حسب نوع فاز ساکن و متحرک ۶
- جدول (۲-۱)-ضریب فعالیت در رقت بی‌نهایت تعدادی از ترکیبات آلی در دو مایع یونی..... ۹
- جدول (۱-۲)-انواع توصیف‌کننده‌های محاسبه شده توسط نرم‌افزار دراگون..... ۲۱
- جدول (۲-۲)-توابع کرنل در فضای ویژگی ۳۷
- جدول (۱-۳)-نام ترکیبات، محدوده‌ی دمایی و مقدار ضریب فعالیت در رقت بی‌نهایت..... ۴۹
- جدول (۲-۳)-مقایسه‌ی آماره‌های سری ارزیابی مدل‌های به دست آمده از روش رگرسیون
مرحله‌ای (SR) ۵۴
- جدول (۳-۳)-نشان، گروه، نام کامل و اثر متوسط توصیف‌کننده‌های انتخاب شده توسط روش
رگرسیون مرحله‌ای (SR) ۵۵
- جدول (۴-۳)-ماتریس همبستگی بین توصیف‌کننده‌های انتخاب شده توسط رگرسیون
مرحله‌ای ۵۶
- جدول (۵-۳)-توصیف‌کننده‌های به دست آمده در ۱۶ مدل بهینه برای ۱۶ دسته بندی
مختلف..... ۵۸
- جدول (۶-۳)-توصیف‌کننده‌های انتخاب شده توسط CDFS ۵۹
- جدول (۷-۳)-ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط روش CDFS ۶۰
- جدول (۸-۳)-مقادیر همبستگی توصیف‌کننده‌های انتخابی با SR با ضریب $\ln \gamma_{13}^{\infty}$ ۶۲
- جدول (۹-۳)-توابع و پارامترهای بهینه‌ی شبکه با استفاده از توصیف‌کننده‌های حاصل از SR
(SR-ANN) ۶۸
- جدول (۱۰-۳)-توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی (SR-ANN) ۶۹
- جدول (۱۱-۳)-مقادیر همبستگی توصیف‌کننده‌های انتخابی با روش CDFS با ضریب فعالیت
در رقت بی‌نهایت ۷۰
- جدول (۱۲-۳)-توابع و پارامترهای بهینه‌ی شبکه با استفاده از توصیف‌کننده‌های انتخابی به
روش CDFS (CDFS-ANN) ۷۵
- جدول (۱۳-۳)-توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف
کننده‌های حاصل از CDFS ۷۶

جدول (۳-۱۴) - مقادیر پارامترهای بهینه و نتایج به دست آمده برای سری تست داخلی با استفاده از توصیف‌کننده‌های انتخاب شده با روش SR	۷۸
جدول (۳-۱۵) - مقادیر پارامترهای بهینه و نتایج به دست آمده برای سری تست داخلی با استفاده از توصیف‌کننده‌های انتخاب شده با روش CDFS	۷۸
جدول (۳-۱۶) - پارامترهای آماری روش‌های SR-ANN, CDFS-ANN, SR-SVM و CDFS-SVM	۹۰
جدول (۳-۱۷) - مثال هایی از اثر مقدار توصیف‌کننده ی nH بر $\ln \gamma_{13}^{\infty}$	۹۱
جدول (۳-۱۸) - مثال هایی از اثر مقدار توصیف‌کننده ی H050 بر $\ln \gamma_{13}^{\infty}$	۹۲
جدول (۳-۱۹) - مثال هایی از مقدار توصیف‌کننده ی RDF020p بر $\ln \gamma_{13}^{\infty}$	۹۳
جدول (۳-۲۰) - مثال هایی از مقدار توصیف‌کننده ی ATS3p بر $\ln \gamma_{13}^{\infty}$	۹۵
جدول (پ-۱) - نام ترکیبات، دماها و ضریب فعالیت در رقت بی‌نهایت سری داده‌ها.....	۱۰۰
جدول (پ-۲) - نتایج حاصل از ارزیابی مدل‌های SR-ANN, CDFS-ANN, SR-SVM و CDFS-SVM با استفاده از سری تست خارجی	۱۱۲

فصل اول

مقدمه

مقدمه

ضریب فعالیت در رقت بی‌نهایت^۱، γ^∞ ، یکی از پارامترهای مهم در محاسبات تعادلی فازی است. این ضریب که دارای کاربردهای عملی مهمی در مسائل مختلف صنعتی از جمله استخراج، جذب سطحی و تقطیر می‌باشد، به نوع محیط بستگی دارد. یکی از محیط‌هایی که امروزه به طور گسترده در بسیاری از روش‌های شیمیایی استفاده می‌شود، مایعات یونی^۲ است به همین دلیل اندازه‌گیری γ^∞ در محیطی مانند یک مایع یونی از اهمیت ویژه‌ای برخوردار است [۱]. مایعات یونی ترکیبات غیر مولکولی هستند که به عنوان حلال‌های جدید در شیمی به رسمیت شناخته شده‌اند. در حال حاضر جایگزین کردن حلال‌های آلی با مایعات یونی به دلیل اشتعال‌ناپذیری و عدم فراریت، موضوع بسیاری از تحقیقات روز دنیا به شمار می‌آید. این ترکیبات به طور وسیعی جایگزین حلال‌های آلی شده‌اند و در زمینه‌های مختلفی مانند سنتز ترکیبات آلی، الکتروشیمی، استخراج و غیره مورد استفاده قرار گرفته‌اند [۲].

مقدار ضریب فعالیت در رقت بی‌نهایت حل‌شونده‌های آلی در مایع یونی می‌تواند اطلاعات بسیار خوبی را از برهم‌کنش بین مولکولی بین حلال و حل‌شونده در هر مایع یونی در اختیار ما قرار دهد [۳]. این مقدار در طراحی فرایندهای جداسازی مختلف مثل جداسازی ترکیبات آلیفاتیک از آروماتیک و جداسازی تیوفن از آلکان‌ها کاربرد دارد. همچنین در فرایند گوگردزدایی از سوخت، که بر مبنای استخراج ترکیبات گوگردی است نیز این مقادیر از اهمیت خاصی برخوردار است. زیرا این ترکیبات نسبت به هیدروکربن‌های آلیفاتیک در مایعات یونی محلول‌تر هستند [۴].

یکی از روش‌های رایج برای اندازه‌گیری γ^∞ ، روش کروماتوگرافی گاز - مایع^۳ است که در ادامه راجع به این موضوع توضیح بیشتری ارائه می‌شود.

۱ - Activity coefficient at infinite dilution

۲ - Ionic liquid

۳ - Gas-Liquid Chromatography

۱-۱- مایعات یونی

مایعات یونی اولین بار در سال ۱۹۱۴ با سنتز اتیل آمونیوم نیترات شناخته شدند [۵]. این ترکیبات ساختار یونی دارند و به دلیل نداشتن تقارن مولکولی در ساختمان کاتیون آن‌ها، علیرغم ماهیت نمکی دارای نقطه‌ی ذوب پایینی بوده و در اکثر موارد در دمای محیط به شکل مایع هستند. در این ترکیبات که عموماً از یک کاتیون آلی نامتقارن بزرگ و یک آنیون آلی یا غیر آلی تشکیل شده‌اند بار مثبت کاتیون و بار منفی آنیون دور از هم قرار می‌گیرند. در نتیجه، نیروی جاذبه بین آن‌ها کاهش یافته و به همین دلیل دارای نقطه‌ی ذوب بسیار پایین‌تری نسبت به نمک‌های فلزی هستند [۶].

مایعات یونی به عنوان انقلاب شیمی سبز در صنایع شیمیایی شناخته می‌شوند زیرا این گروه شیمیایی جدید می‌توانند استفاده از حلال‌های متداول آلی خطرناک و آلوده‌کننده را که از مشخصه‌های اصلی این حلال‌هاست، کاهش دهند. همچنین این مواد می‌توانند در فرایند ساخت و سنتزهای مختلف نیز سهیم شوند [۷].

۱-۱-۱- ساختار و خواص ترمودینامیکی مایعات یونی

اصلی‌ترین خاصیت مایعات یونی که در بخش قبل نیز به آن اشاره شد دمای ذوب پایین آن‌هاست که از عدم تقارن مولکولی کاتیون آن‌ها ناشی می‌شود. ویژگی دیگر مایعات یونی، حل شدن بسیاری از ترکیبات آلی و معدنی در آن‌هاست، مانند انحلال بالای ترکیبات آروماتیک در مایعات یونی. از طرفی برخی از ترکیبات آلی مانند آلکان‌ها، حلالیت بسیار کمی در این مایعات دارند. مایعات یونی با ایجاد یک محیط قطبی غیرآبی در جداسازی‌های چند فازی بسیار مؤثر هستند. این خاصیت مایعات یونی کاربردهای زیادی را پدید آورده که از آن جمله می‌توان به جداسازی آلکان‌ها از ترکیبات آروماتیک اشاره نمود. یکی

دیگر از خواص مایعات یونی، فشار بخار پایین آن‌هاست. فشار بخار پایین و پایداری حرارتی خوب باعث می‌شود تا مایعات یونی به عنوان حلال‌های سبز شناخته شوند زیرا این عامل باعث کاهش اثرات زیست محیطی شده و نگرانی‌های سلامتی را کاهش می‌دهد. کم بودن آثار منفی مایعات یونی بر روی محیط زیست، یکی از اصلی‌ترین انگیزه‌های جایگزینی آن‌ها به جای حلال‌های آلی است. همچنین این خاصیت باعث کاهش هزینه‌های بازیابی در استخراج نیز می‌گردد.

خواص دیگر مایعات یونی را می‌توان به صورت زیر طبقه‌بندی کرد:

- قطبیت بالا و عدم تشکیل کمپلکس

- هدایت الکتریکی بالا

- غیرقابل اشتعال بودن

لازم به ذکر است که خواص مایعات یونی با تغییر زنجیره‌ی هیدروکربنی متصل به کاتیون کاملاً تغییر می‌کند و این یکی از مزایای مایعات یونی جهت طراحی فرایندهای جداسازی است، زیرا با تغییر ساختار آن‌ها می‌توان در شرایط متفاوت از آن‌ها بهره برد. به همین دلیل به این مایعات لقب حلال‌های قابل طراحی داده شده است [۵].

۱-۲- کاربردهای مایعات یونی

کاربرد جدی مایعات یونی به دو دهه‌ی اخیر برمی‌گردد. در این دو دهه به این ترکیبات به عنوان جایگزین مناسب برای حلال‌های آلی موجود توجه خاصی شد و تحقیقات وسیعی جهت استفاده از مایعات یونی به عنوان حلال در استخراج‌های مایع آغاز شد. علت اصلی این موضوع دمای ذوب پایین این ترکیبات هست که باعث پدید آمدن یک محیط قطبی مایع غیرآبی شده و امکان انجام واکنش‌های متعدد آلی را فراهم می‌سازد [۷].

کاربرد عمده‌ی دیگر مایعات یونی در شیرین سازی گاز طبیعی است. از مایعات یونی می‌توان به عنوان جاذب گازهای اسیدی مانند H_2S و CO_2 استفاده نمود. کاربرد دیگر این مواد جداسازی ترکیبات آروماتیک از سایر هیدروکربن‌ها مانند آلکان‌ها است. مایعات یونی آروماتیک‌ها را به خوبی در خود حل کرده و در مقابل آلکان‌ها را به مقدار خیلی ناچیز در خود حل می‌کنند. لذا می‌توان از آن‌ها در جداسازی ترکیبات آروماتیک از آلکان‌ها استفاده نمود [۵].

همچنین از مایعات یونی می‌توان در استخراج مواد طبیعی، فرایندهای هسته‌ای، فراوری گلوکز و سلولز، صنایع غذایی، استفاده از انرژی خورشیدی، ذخیره‌سازی هیدروژن، تصفیه‌ی پسماندها، ساخت باتری‌هایی با طول عمر زیاد، حذف یون‌های کادمیوم و جیوه از آبهای آلوده، واکنش‌های پلیمر شدن و همچنین به عنوان عامل واسطه در واکنش‌های بیولوژیکی استفاده نمود [۵]. البته لازم به ذکر است که قیمت بالا و کمبود اطلاعات در مورد خصوصیت فیزیکی این ترکیبات، مزیت استفاده از مایعات یونی در فرایندهای شیمیایی را محدود می‌کند.

۱-۲- کروماتوگرافی

کروماتوگرافی یک تکنیک جداسازی است که اغلب در آنالیزهای شیمیایی به کار برده می‌شود. کروماتوگرافی یکی از قدرتمندترین تکنیک‌های موجود و وسیله‌ای چند منظوره برای شیمیدان‌های تجزیه‌ی امروزی است. این روش در یک فرایند تک مرحله‌ای می‌تواند مخلوط را به اجزای جزئی آن جداسازی نماید و همزمان درباره‌ی هر قسمت یک تخمین کمی بدهد.

در کروماتوگرافی دو فاز داریم: یک فاز ساکن (فاز ثابت) که معمولاً آن را در یک لوله ساکن نگه می‌دارند و یک فاز متحرک که از میان فاز ساکن عبور می‌کند. در این روش جداسازی بر مبنای تمایل نسبی هر جزء به فاز ساکن صورت می‌گیرد. گونه‌ای که تمایل بیشتری به فاز متحرک دارد، با سرعت بیشتری حرکت می‌کند و بالعکس گونه‌ای که به فاز

ساکن تمایل بیشتری دارد، با سرعت کمتری در طول ستون حرکت می‌کند. برخی از انواع کروماتوگرافی برحسب نوع فاز ساکن و متحرک در جدول ۱-۱ مشخص شده‌اند [۸].

جدول (۱-۱) - انواع کروماتوگرافی برحسب نوع فاز ساکن و متحرک

نوع فاز ساکن	نوع فاز متحرک	نام روش
کاغذ	مایع یا محلول	کروماتوگرافی کاغذی
جامد فعال	مایع یا محلول	کروماتوگرافی جذب سطحی
جامد فعال	گاز	کروماتوگرافی گاز - جامد
مایع فعال	گاز	کروماتوگرافی گاز - مایع
لایه نازک از جامد فعال	مایع یا محلول	کروماتوگرافی لایه نازک

با توجه به این که یکی از اصلی‌ترین روش‌های اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت روش کروماتوگرافی گازی است، در ادامه این روش به اختصار شرح داده خواهد شد.

۱-۲-۱- کروماتوگرافی گازی

یکی از انواع معمول کروماتوگرافی، کروماتوگرافی گازی است. در این روش، نمونه تبخیر گردیده و سپس به ستون کروماتوگرافی تزریق می‌شود. فاز متحرک یا گاز حامل یک فاز بی‌اثر مثل هلیوم، نئون، نیتروژن، آرگون یا دی‌اکسید کربن است و فاز ساکن یک جسم جامد جاذب و یا لایه‌ی نازکی از یک مایع غیر فرار است که در دیواره‌ی داخلی ستون قرار داده می‌شود. بر اساس نوع فاز ساکن، کروماتوگرافی گازی به دو گروه عمده تقسیم‌بندی می‌شود:

الف) کروماتوگرافی گاز - جامد که فاز ساکن در آن جامد است.

ب) کروماتوگرافی گاز - مایع که فاز ساکن در آن مایع غیر فرار است [۸].

مهم‌ترین ویژگی‌ها و مزایای کروماتوگرافی گازی در مقایسه با سایر روش‌های اندازه‌گیری و جداسازی عبارتند از: توانایی جداسازی نمونه‌های شامل چندین جزء با نقطه‌ی جوش و قطبیت نزدیک به هم، حساسیت، گزینش‌پذیری و هزینه‌ی کم نگهداری و مواد مصرفی نسبت به روش‌های دیگر [۹].

۱-۲-۲- استفاده از مایعات یونی به‌عنوان فاز ساکن در کروماتوگرافی گاز-مایع

خواص غیرمعمول و منحصر به فرد مایعات یونی سبب شده است تا این دسته از ترکیبات، کاربرد گسترده‌ای در زمینه‌های گوناگون کروماتوگرافی به ویژه در کروماتوگرافی گازی پیدا کنند. در کروماتوگرافی گازی برای دستیابی به جداسازی‌هایی با کارایی و گزینش‌پذیری بالا، فاز ساکن باید دارای خصوصیات ویژه‌ای باشد. از جمله این خصوصیات می‌توان به بالا بودن پایداری حرارتی و ویسکوزیته و همچنین پایین بودن کشش سطحی اشاره کرد. مایعات یونی با داشتن این خصوصیات منحصر به فرد، نشان داده‌اند که می‌توانند به عنوان فاز ساکن در کروماتوگرافی گازی به کار برده شوند و جداسازی‌هایی گزینش‌پذیر را با کارایی بالا انجام دهند [۲].

۱-۲-۳- اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت با روش کروماتوگرافی گاز-مایع

با اهمیت‌ترین پارامتر ترمودینامیکی که می‌تواند با روش کروماتوگرافی گاز-مایع به دست آید، ضریب فعالیت در رقت بی‌نهایت γ_{13}^{∞} است. این روش سریع و دقیق است و به مقدار بسیار کمی از ماده نیاز دارد. استفاده از این روش برای اندازه‌گیری مستقیم ضریب فعالیت در رقت بی‌نهایت توسط محققین بسیاری به کار برده شده و ثابت شده است که روش کاملاً قابل اطمینانی است.

مقدار ضریب فعالیت در رقت بی‌نهایت با استفاده از داده‌های کروماتوگرافی گاز-مایع می‌تواند از رابطه‌ی (۱-۱) که توسط کروکشانک^۱ و همکارانش ارائه شده است محاسبه گردد [۴].

$$\ln \gamma_{13}^{\infty} = \ln \left(\frac{n_3 RT}{V_N p_1^*} \right) - \frac{p_1^* (B_{11} - V_1^*)}{RT} + \frac{p_0 J_2^3 (2B_{12} - V_1^{\infty})}{RT} \quad (1-1)$$

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 حل شونده، p_1^* فشار بخار حل‌شده خالص در دمای T ، B_{11} ضریب دوم ویریال حل شونده‌ی خالص، V_1^* حجم مولی حل شونده‌ی خالص، p_0 فشار خروجی ستون، $p_0 J_2^3$ فشار متوسط ستون، B_{12} ضریب دوم ویریال مخلوط حل‌شده 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 حلال ۳ در رقت بی‌نهایت است.

۱-۲-۴- بررسی مقدار ضریب فعالیت در رقت بی‌نهایت برای ترکیبات آلی در دو

مایع یونی متفاوت

مقدار ضریب فعالیت در رقت بی‌نهایت چندین ترکیب آلی در دو محیط مایع یونی ۱-بوتیل-۱-متیل مورفولینیوم تری سیانومتانید [TCM][BMMOR] و ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید [TCM][BMPYR] در جدول ۱-۲ آمده است [۴ و ۱].

جدول (۱-۲) - ضریب فعالیت در رقت بی‌نهایت تعدادی از ترکیبات آلی در دو مایع یونی

نام ترکیب آلی	γ_{13}^{∞} در محیط مایع یونی [TCM] [BMMOR]	γ_{13}^{∞} در محیط مایع یونی [TCM] [BMPYR]
Pentane	۳۹/۷	۱۶/۷
Hexane	۵۶/۵	۲۳/۴
Heptane	۸۴/۰	۳۳/۷
Octane	۱۲۵	۴۸/۱
Nonane	۱۸۷	۶۸/۸
Decane	۲۸۹	۹۲/۲
Hex-1-ene	۲۲/۶	۱۰/۶
Hept-1-ene	۳۴/۱	۱۵/۰
Oct-1-ene	۵۲/۷	۲۱/۸
Hex-1-yne	۵/۴۷	۲/۸۷
Hept-1-yne	۸/۲۸	۴/۰۳
Oct-1-yne	۱۲/۸	۵/۷۳
Benzene	۱/۵۳	۱/۰۲
Toluene	۲/۳۰	۱/۴۳
o-Xylene	۲/۹۶	۱/۷۶
Propan-1-ol	۱/۶۲	۱/۱۴
Pyridine	۰/۷۳۳	۰/۵۶۷

همان‌طور که قبلاً گفته شد مقدار γ^∞ به ساختار مایع یونی بستگی دارد که این موضوع به خوبی در مقادیر داده شده قابل مشاهده است. (مقدار γ_{13}^∞ برای یک ترکیب آلی خاص در دو مایع یونی مختلف با یکدیگر متفاوت است). اما مقدار ضریب فعالیت در رقت بی‌نهایت ترکیبات آلی در مایعات یونی را می‌توان به دو عامل زیر مربوط دانست: اول، تفاوت در اندازه مولکول‌های حلال و گونه حل‌شونده و دوم برهم‌کنش بین مولکول‌ها [۲].

- با توجه به دو عامل ذکرشده و مقادیر داده شده در جدول فوق می‌توان نکات زیر را بیان نمود:
- مقدار ضریب فعالیت در رقت بی‌نهایت برای آلکان‌ها بزرگ‌تر از سایر ترکیب‌ها از جمله آلکین‌ها و ترکیبات آروماتیک می‌باشد که نشان می‌دهد آلکان‌ها (حل‌شونده) برهم‌کنش بسیار کوچکی با حلال (مایع یونی) دارد. آلکین‌ها و ترکیبات آروماتیک برهم‌کنش‌های قوی با مایع یونی دارند زیرا آلکین‌ها از طریق پیوند سه‌گانه و ترکیبات آروماتیک از طریق الکترون‌های غیرمستقر π می‌توانند شدیداً با مایع یونی برهم‌کنش داشته باشند و مقادیر کوچکی از γ_{13}^∞ را نتیجه دهند.
 - ضریب فعالیت در رقت بی‌نهایت برای آلکان‌های خطی با افزایش طول زنجیر کربنی و اختلاف اندازه مولکول حل‌شونده و حلال افزایش پیدا می‌کند.
 - مقدار γ_{13}^∞ برای آلکن‌ها کمتر از آلکان‌های هم‌کربن با آن‌هاست زیرا وجود پیوند دوگانه برهم‌کنش بین‌مولکولی بین حل‌شونده و مایع یونی را افزایش می‌دهد.
 - مقدار γ_{13}^∞ برای الکل‌ها نیز کوچک است که نشان می‌دهد الکل‌ها از طریق گروه OH می‌توانند برهم‌کنش‌های قوی با مایع یونی داشته باشند.

۱-۳- مروری بر کارهای گذشته برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت

همان‌طور که در بخش‌های قبلی ذکر گردید، اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت در محیطی مانند مایعات یونی بسیار حائز اهمیت است. از طرفی اندازه‌گیری این ضرایب به طریق

تجربی و آزمایشگاهی بسیار پرهزینه و وقت‌گیر است. لذا ارائه‌ی یک روش تئوری برای پیش‌بینی این ضرایب در مایعات یونی ارزشمند است.

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 به‌عنوان حل‌شونده در سه مایع یونی [emim][Tf₂N]، [bmpyr][BF₄] و [emmim][Tf₂N] در دمای ۲۹۸ کلوین انجام دادند. آنها توانستند برای $\ln \gamma_{13}^{\infty}$ در هر مایع یونی یک رابطه همبستگی ارائه کنند که مقدار ضریب تعیین^۲ گزارش شده توسط آنها برای سه مایع یونی ذکر شده به ترتیب ۰/۹۵۲، ۰/۹۷۵ و ۰/۹۷۰ بود [۱۰].

در سال ۲۰۰۷ جیکین^۳ و همکارانش ضریب فعالیت در رقت بی‌نهایت حل‌شونده‌های هیدروکربنی در چهار مایع یونی مختلف در دمای ۲۹۸ تا ۳۱۸ کلوین مورد بررسی قرار دادند. آن‌ها برای هر مایع یونی به طور جداگانه با استفاده از رگرسیون خطی مدل‌سازی را انجام دادند و توانستند نتایج خوبی به دست آورند [۱۱].

در سال ۲۰۱۰ هیوجان سان^۴ و همکارانش ضریب فعالیت در رقت بی‌نهایت تعدادی از ترکیبات آلی را در یک مایع یونی در سه دمای مختلف با استفاده از روش حداقل مربعات متداول^۵ پیش‌بینی کردند [۱۲].

در سال ۲۰۱۱ دیهیمی^۶ و نامی^۷ برای اولین بار توانستند با استفاده از روش شبکه‌ی عصبی مصنوعی، ضریب فعالیت در رقت بی‌نهایت را برای چندین ترکیب آلی در ۱۶ مایع یونی در دماهای مختلف پیش‌بینی کنند [۳].

۱- Edward j. Maginn

۲- Determination coefficient

۳- Jiqin

۴- Hui zhang sun

۵- Ordinary Least Squares

۶- Deyhimi

۷- Nami

در سال ۲۰۱۵ نیز کلانتر^۱ و همکارانش توانستند ضریب فعالیت در رقت بی‌نهایت ۵۸ ترکیب آلی و آب را در مایع یونی ۱-بوتیل ۱-متیل پیرولیدینیوم تری سیانو متانید در ۶ دمای مختلف از ۳۱۸/۱۵ تا ۳۶۸/۱۵ درجه کلون با استفاده از شبکه‌ی عصبی مصنوعی پیش‌بینی کنند [۱۳].

۱-۴-هدف تحقیق

هدف این پایان‌نامه ارائه‌ی یک روش ارتباط کمی ساختار - خاصیت (QSPR) برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت برای ۶۰ ترکیب آلی و آب در محیط مایع ۱-بوتیل ۱-متیل-مورفولینیوم تری سیانومتانید [TCM][BMMOR] است. این مایع یونی دارای قدرت انتخاب بسیار بالایی برای جداسازی ترکیبات گوگردی یا نیتروژن از آلکان‌هاست. این موضوع در فرایند گوگردزدایی از سوخت از اهمیت بسزایی برخوردار است.

فصل دوم

کمو متریکس

با ورود فن‌آوری‌های جدید به دنیا و پیشرفت‌های شگرف در علوم مختلف، شیمی هم دچار تحولات زیادی شده است. ورود دستگاه‌های با توانمندی بالا جهت تولید داده‌ها، مشکل به دست آوردن اطلاعات را که روزگاری نه چندان دور، معضلی برای شیمیدانان تلقی می‌شد آسان کرده است. اما سؤالی که مطرح می‌شود این است که چگونه می‌توان ماکزیمم اطلاعات لازم را از این طیف وسیع داده‌ها استخراج کرد؟ همچنین باید به این نکته اشاره نمود که گاهی اوقات شیمیدان‌ها با موادی سر و کار دارند که سمی، گران و خطرناک بوده و یا در مواردی به راحتی قابل دسترس نیستند. کمومتریکس^۱ به عنوان یک علم جدید و جوان در شیمی با بهره‌گیری از روش‌های ریاضی و آمار کمک شایانی به حل این مسائل نموده است [۱۴].

۲-۱- کمومتریکس

ریشه‌ی کمومتریکس به سال ۱۹۶۹ بازمی‌گردد. هنگامی که جرز^۲، کووالسکی^۳ و آیزنهاور^۴ مقالاتی را در ارتباط با کاربرد ماشین فراگیر خطی^۵ در شیمی تجزیه منتشر کردند. انتشار این مقالات روش نوینی در زمینه‌ی تبدیل حجم وسیع داده‌های شیمیایی به اطلاعات شیمیایی معنی‌دار بود. انگیزه‌ی این تحقیق در شیمی آن بود که سال‌ها دانشمندان دفترچه‌های یادداشت آزمایشگاهی خود را با داده‌هایی پرکرده بودند که اغلب به دلیل فقدان روش‌های مناسب، بررسی داده‌ها مورد توجه قرار نمی‌گرفت [۱۵]. نام کمومتریکس نیز برای نخستین بار توسط اسوانت ولد^۶، دانشمند جوان سوئدی که در زمینه شیمی- فیزیک آلی فعالیت داشت

۱- Chemometrics

۲- Jurs

۳- Kowalski

۴- Isenhour

۵- Linear Learning machine

۶- Svante Wold

پیشنهاد گردید. همکاری ولد با کووالسکی که در دانشگاه واشنگتن آمریکا بر روی روش‌های الگوشناسی در شیمی تجزیه مطالعه می‌کرد، منجر به تأسیس انجمن بین‌المللی کمومتریکس^۱ (ICS) در سال ۱۹۷۴ گردید [۱۶]. با تأسیس انجمن بین‌المللی کمومتریکس، رشد و توسعه‌ی کمومتریکس سرعتی بیش از پیش به خود گرفت و امروزه به طور گسترده‌ای در زمینه‌های مختلف شیمی کاربرد دارد. اسوانت ولد با طرح سؤال زیر یک تعریف کلی برای کمومتریکس ارائه داد. چگونه می‌توان اطلاعات شیمیایی را از داده‌های اندازه‌گیری شده، به دست آورد و آن را ارائه داد. وی در پاسخ به این سوال گفت کمومتریکس مجموعه‌ای از روش‌های ریاضی، اصول آماری و دیگر روش‌های مبتنی بر منطق است که به استخراج اطلاعات شیمیایی کمک می‌کند. در واقع کمومتریکس یک زمینه‌ی بین رشته‌ای است که شامل آمارهای چند متغیره، مدل‌های ریاضی، علم کامپیوتر و شیمی است [۱۴]. تعریف کامل‌تری از کمومتریکس را می‌توان به صورت زیر بیان کرد: «کمومتریکس شاخه‌ای از شیمی است که از روش‌های آماری و ریاضی به منظور طراحی یا انتخاب فرایندهای بهینه و فراهم کردن بیشترین اطلاعات شیمیایی با آنالیز داده‌های شیمیایی، استفاده می‌کند» [۱۷].

تکامل کمومتریکس توسط هاوری^۲ و هیرش^۳ و همچنین بعدها توسط براون^۴ صورت گرفت. در این دوره توجه پژوهشگران عموماً روی روش‌های چند متغیره متمرکز بود. زیرا جهان اطراف ما ذاتاً چند متغیره است، بنابراین عملیات آنالیز داده‌ها در اندازه‌گیری‌های چندگانه قابل درک است. با بررسی و در نظر گرفتن مناسب توزیع متغیرهای چندگانه به طور همزمان می‌توان اطلاعات بیشتری نسبت به حالتی که هر متغیر را به طور مجزا بررسی می‌شود، به دست آورد.

۱-International Chemometrics Society

۲-Howery

۳-Hirsch

۴-Brown

۲-۲- ارتباط کمی ساختار - خاصیت و ساختار - فعالیت

۲-۲-۱- رابطه کمی ساختار - فعالیت^۱ (QSAR)

یکی از زمینه‌های مهم کمومتریکس، مطالعاتی است که فعالیت بیولوژیکی ترکیبات را به ویژگی‌های ساختاری آن‌ها ارتباط می‌دهد. این نوع از مطالعات به رابطه‌ی کمی ساختار - فعالیت معروف هستند [۱۸]. درواقع QSAR شامل تمام روش‌های آماری است که فعالیت‌های بیولوژیکی را به ویژگی‌های فیزیکوشیمیایی ترکیب مربوط می‌کند.

با توجه به این که مولکول‌های مشابه با تغییر کوچکی در ساختارشان می‌توانند فعالیت بیولوژیکی متفاوتی داشته باشند. می‌توان گفت QSAR سعی در پیدا کردن رابطه‌ای بین فعالیت بیولوژیکی و ویژگی‌های مولکولی دارد تا بتواند از این قواعد برای ارزیابی فعالیت ترکیبات جدید استفاده کند [۱۹]. برای این منظور باید QSAR به دو سوال زیر پاسخ دهد:

الف) چه جنبه‌هایی از مولکول روی فعالیت آن تأثیر می‌گذارد؟

ب) برای افزایش آن فعالیت چه عواملی باید بهینه شوند؟

۲-۲-۲- رابطه کمی ساختار - خاصیت^۲ (QSPR)

یکی دیگر از زمینه‌های مهم کاربرد کمومتریکس، مطالعاتی است که خواص مولکول‌ها را به ویژگی‌های ساختاری آن‌ها نسبت می‌دهد. از نظر شیمیدانان خاصیت یک مولکول ناشی از ویژگی‌های ساختاری آن‌هاست. این نوع از مطالعات به بررسی کمی ارتباط ساختار - خاصیت معروف می‌باشد. هدف از این مطالعات پیدا کردن رابطه‌ای است که بین رفتار فیزیکوشیمیایی یک مولکول با پارامترهای ساختاری آن وجود دارد. مانند یافتن رابطه‌ای بین ساختار مولکول‌ها

۱- Quantitative Structure-Activity Relationship

۲- Quantitative Structure-Property Relationship

با خواصی نظیر نقطه‌ی جوش و ذوب، فشار بخار و غیره. نتایج این مطالعات علاوه بر شفاف سازی نحوه‌ی ارتباط خواص مولکول‌ها و ویژگی‌های ساختمانی آن‌ها با خاصیت مورد نظر، به پژوهشگران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند.

از روش‌هایی که به منظور مطالعه‌ی ارتباط خطی ساختار خاصیت مورد استفاده قرار می‌گیرند می‌توان به روش‌های خطی مثل حداقل مربعات متداول (OLS)، تحلیل مؤلفه‌های اساسی^۱ (PCA)، رگرسیون خطی چندگانه^۲ (MLR)، رگرسیون اجزای اصلی^۳ (PCR) و حداقل مربعات جزئی^۴ (PLS)، اشاره نمود. روش‌های دیگری مانند شبکه عصبی^۵ (ANN) و ماشین بردار پشتیبان^۶ (SVM)، ارتباط غیر خطی میان ساختار و خواص ترکیبات را مورد مطالعه قرار می‌دهند.

۲-۳- مراحل انجام QSPR

به‌طور کلی مطالعات QSPR و همچنین QSAR شامل مراحل زیر است:

۱. فراهم کردن سری داده‌ها
۲. رسم و بهینه‌سازی ساختار هندسی مولکول‌ها
۳. محاسبه توصیف‌کننده‌ها
۴. حذف توصیف‌کننده‌های نامناسب
۵. تقسیم‌بندی داده‌ها به سری‌های آموزش، ارزیابی و تست

۱- Principle Component Analysis

۲- Multiple Linear Regression

۳- Principal Component Regression

۴- Partial Least Squares

۵- Artificial Neural Networks

۶- Support Vector Machines

۶. انتخاب بهترین توصیف‌کننده‌ها

۷. ساختن مدل

۸. ارزیابی اعتبار مدل‌های انتخاب شده

۹. پیش‌بینی خاصیت مورد نظر

۲-۳-۱- فراهم کردن سری داده‌ها

در یک مطالعه QSPR اولین مرحله جمع‌آوری و انتخاب سری داده‌هاست. داده‌های این مطالعه به دو دسته‌ی اصلی متغیرهای مستقل و متغیرهای وابسته تقسیم می‌شوند. معمولاً فعالیت‌های زیستی و سایر کمیت‌هایی که بیانگر خاصیت ویژه‌ای از یک ترکیب و یا دسته‌ای از ترکیبات هستند و تحت عنوان متغیر وابسته شناخته می‌شوند، جهت مدل‌سازی انتخاب می‌شوند. مقدار تجربی این کمیت‌ها باید برای تعدادی از آن‌ها اندازه‌گیری شده و یا اینکه توسط محققین دیگر در مجلات علمی گزارش شده باشد. البته ترکیب‌های موجود در سری داده‌ها باید مشابهت ساختاری داشته و کمیت مورد نظر آن‌ها در شرایط یکسان تعیین شده باشند.

۲-۳-۲- رسم و بهینه‌سازی ساختار هندسی مولکول‌ها

ایجاد توصیف‌کننده‌های هندسی بر مبنای ساختار و هندسه‌ی دقیق مولکولی استوار است. اگر ساختارها به صورت صورت‌بندی با حداقل انرژی نباشند مقادیر غیر صحیحی برای توصیف‌کننده‌ها ایجاد می‌شود. بنابراین ساختار هندسی ترکیبات مورد مطالعه با استفاده از روش‌های محاسباتی در نرم‌افزارهای مختلفی چون هایپرکم^۱ و گوسین^۲ بهینه می‌گردند. در این تحقیق ساختار مولکول‌ها پس از رسم شدن در نرم‌افزار هایپرکم با استفاده از روش نیمه تجربی

۱-Hyperchem

۲-Gaussian

کوانتومی PM3^۱ در این نرم افزار بهینه شدند.

۲-۳-۳- محاسبه‌ی توصیف‌کننده‌ها

توصیف‌کننده‌ها پارامترهای عددی هستند که ویژگی‌های متفاوتی از یک مولکول را بیان می‌کنند. توصیف‌کننده‌های مولکولی به دو دسته مهم تقسیم می‌شوند:

(الف) توصیف‌کننده‌های حاصل از اندازه‌گیری‌های آزمایشگاهی (مثل دما و فشار)

(ب) توصیف‌کننده‌های نظری: این توصیف‌کننده‌ها با استفاده از ساختار مولکول و بدون نیاز به داده‌های تجربی محاسبه می‌شوند که این باعث صرفه‌جویی در وقت و هزینه‌ی مواد و تجهیزات خواهد شد و برای هر مولکول واقعی یا فرضی که هنوز سنتز نشده است نیز در دسترس می‌باشد [۱۳]. این توصیف‌کننده‌ها از بیان نمادین مولکول ناشی می‌شوند و تعداد اطلاعات شیمیایی حاصل از آن‌ها به نوع نمایش بستگی دارد. ساده‌ترین نمایش مولکولی فرمول شیمیایی است که فهرستی از اتم‌ها می‌باشند. این بیان مستقل از نیاز به دانستن ساختار مولکولی است. توصیف‌کننده‌های مولکولی حاصل از فرمول شیمیایی، توصیف‌کننده‌های 0D^۲ نامیده می‌شوند. عدد اتمی، وزن مولکولی و تعداد نوع اتم مثال‌هایی از این گروه توصیف‌کننده می‌باشند.

زیر ساختارها به عنوان نمایش یک بعدی مولکول شناخته می‌شوند و در واقع فهرستی از قطعات ساختمانی مولکول هستند که این فهرست شامل گروه‌های عاملی یا استخلاف‌های موجود در مولکول می‌باشند. بنابراین کاملاً نیاز به دانستن ساختار مولکولی ندارند. این توصیف‌کننده‌ها به عنوان توصیف‌کننده‌های 1D^۳ شناخته شده و در آنالیز زیر ساختارها و مدل‌سازی کاربرد زیادی دارند.

۱-Parameterized Model number 3

۲-zero Dimentional Descriptors

۳-one Dimentional Descriptors

نمایش دو بعدی مولکول نشان‌دهنده‌ی اتصالات در مولکول بوده و نشان می‌دهد که چگونه اتم‌ها به هم متصل شده‌اند و طبیعت پیوندهای شیمیایی را نشان می‌دهد. توصیف‌کننده‌هایی که بر اساس گراف مولکولی نمایش دوبعدی مولکول را نشان می‌دهند توصیف‌کننده‌های 2D^۱ نامیده می‌شوند. توصیف‌کننده‌های توپولوژیکی جزء این دسته به شمار می‌آیند.

نمایش سه بعدی، مولکول را به‌عنوان جسم هندسی سخت در فضا در نظر می‌گیرد. در نتیجه این نمایش نه تنها به پیوند اتم‌ها بستگی دارد بلکه به صورت‌بندی کلی مولکول نیز وابسته است. این نمایش مولکول در واقع نمایش هندسی بوده و توصیف‌کننده‌های مولکولی حاصل از آن توصیف‌کننده‌های 3D^۲ نام دارند. مثال‌هایی از این توصیف‌کننده‌ها، توصیف‌کننده‌های فضایی و توصیف‌کننده‌های اندازه هستند [۲۰].

در این تحقیق ۱۴۸۱ توصیف‌کننده‌ی ساختاری توسط نرم‌افزار دراگون^۳ محاسبه شد که این توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به ۱۸ دسته متفاوت تقسیم می‌شوند، که در جدول (۱-۲) نام این توصیف‌کننده‌ها آورده شده است [۲۱].

۱-Two Dimentional Descriptors

۲-Three Dimentional Descriptors

۳-Dragon

جدول (۲-۱) - انواع توصیف‌کننده‌های محاسبه شده توسط نرم‌افزار دراگون

۱	توصیف‌کننده‌های زیرساختاری ^۱	۲	توصیف‌کننده‌های توپولوژی ^۲
۳	شمارنده‌های مولکولی ^۳	۴	توصیف‌کننده‌های BCUT ^۴
۵	شاخص بار توپولوژیکی GALVEZ ^۵	۶	خود ارتباطی‌های دو بعدی ^۶
۷	توصیف‌کننده‌های بار ^۷	۸	شاخص‌های آروماتیسیته ^۸
۹	پروفایل‌های مولکولی راندیک ^۹	۱۰	توصیف‌کننده‌های هندسی ^{۱۰}
۱۱	توصیف‌کننده‌های RDF ^{۱۱}	۱۲	توصیف‌کننده‌های سه بعدی مورس ^{۱۲}
۱۳	توصیف‌کننده‌های WHIM ^{۱۳}	۱۴	توصیف‌کننده‌های GETAWAY ^{۱۴}
۱۵	گروه‌های عاملی ^{۱۵}	۱۶	اجزای میان اتمی ^{۱۶}
۱۷	توصیف‌کننده‌های تجربی ^{۱۷}	۱۸	خصوصیات مولکولی ^{۱۸}

۲-۳-۴- حذف توصیف‌کننده‌های نامناسب

تا این مرحله تعداد زیادی توصیف‌کننده به دست آمده است. برخی از این توصیف‌کننده‌ها اطلاعات کمتری در مورد تغییرات خواص شیمیایی و یا بیولوژیکی مواد مورد نظر دارند، و برای مدل‌سازی مناسب نیستند، پس باید حذف شوند. توصیف‌کننده‌هایی که حذف می‌شوند معمولاً

-
- ۱- Constitutional Descriptors
 - ۲- Topological Descriptors
 - ۳- Molecular Walk Counts
 - ۴- BCUT Descriptors
 - ۵- Galvez topol Charge Indices
 - ۶- 2D Autocorrelations
 - ۷- Charge Descriptors
 - ۸- Aromatic Indices
 - ۹- Randic Molecular Profiles
 - ۱۰- Geometrical Descriptors
 - ۱۱- Radial Distribution Function Descriptors
 - ۱۲- 3D-MORSE Descriptors
 - ۱۳- Weighted Holistic Invariant Molecular Descriptors
 - ۱۴- Geometry, Topology and Atom Weighted Assembly
 - ۱۵- Functional Groups
 - ۱۶- Atom- Central Fragments
 - ۱۷- Empirical Descriptors
 - ۱۸- Molecular Properties

دارای یک یا چند ویژگی زیر هستند.

- ۱- توصیف‌کننده‌هایی که دارای بیش از ۹۰ درصد مقادیر یکسان باشند.
 - ۲- توصیف‌کننده‌هایی که با توصیف‌کننده‌های دیگر همبستگی بیش از ۰/۹ دارند. برای این منظور می‌توان از نرم‌افزار SPSS استفاده کرد. اما انجام این کار با استفاده از این نرم‌افزار کاری بسیار وقت‌گیر است. اخیراً با استفاده از برنامه‌ی نوشته شده در محیط نرم‌افزار MATLAB، همبستگی توصیف‌کننده‌ها بررسی شده و توصیف‌کننده‌های حذفی مشخص می‌شوند.
 - ۳- توصیف‌کننده‌هایی که با متغیر وابسته همبستگی کمی دارند و ارتباط آن‌ها با خاصیت مورد نظر مشکل است.
- پس از حذف توصیف‌کننده‌های نامناسب، تعداد توصیف‌کننده‌ها کاهش یافته و به این ترتیب معادلات نهایی ساده‌تر شده و پیچیدگی محاسبات به میزان زیادی کاهش می‌یابد.

۲-۳-۵- تقسیم‌بندی داده‌ها

قبل از ساختن مدل، سری داده‌ها به سه زیر مجموعه تقسیم می‌شوند. اولین زیر مجموعه سری آموزش^۱ می‌باشد که اکثریت داده‌ها را در بر گرفته و از آن برای ساخت مدل‌های خطی و غیرخطی استفاده می‌شود. دومین زیر مجموعه، سری ارزیابی^۲ است که مدل‌های به دست آمده از سری آموزش توسط آن ارزیابی شده و به این ترتیب توصیف‌کننده‌های بهینه برای ساخت مدل خطی و غیرخطی مشخص می‌شوند. زیر مجموعه‌ی سوم سری تست^۳ است که در طول فرایند مدل‌سازی هیچ‌گونه دخالتی نداشته و از آن برای مقایسه مدل‌های مختلف استفاده می‌شود.

۱- Training set

۲- Validation set

۳- Test set

۲-۳-۶- انتخاب بهترین توصیف‌کننده‌ها

تا این مرحله توصیف‌کننده‌های نامناسب حذف شده‌اند، اما هنوز تعداد توصیف‌کننده‌های باقی‌مانده برای مدل‌سازی بسیار زیاد است. هرچه در مدل نهایی تعداد توصیف‌کننده‌ها کمتر باشند، مدل ساده‌تر و کارآمدتر خواهد شد. بنابراین انتخاب بهترین توصیف‌کننده‌ها قبل از آغاز مرحله‌ی مدل‌سازی بسیار با اهمیت می‌باشد. در این مرحله باید توصیف‌کننده‌هایی که اطلاعات مفیدتری برای افزایش قدرت پیش‌بینی مدل را دارند، پیدا کرد. برای این منظور می‌توان از روش‌های خطی و غیرخطی گوناگونی استفاده نمود. از جمله این روش‌ها می‌توان به روش‌های رگرسیون مرحله‌ای یا گام به گام^۱ (SR)، روش اتحاد جداسازی داده و انتخاب خاصیت^۲ (CDFS) و روش الگوریتم ژنتیک اشاره نمود که در این پایان‌نامه دو روش SR و CDFS مورد استفاده قرار گرفته‌اند که در ادامه توضیح داده خواهند شد.

۲-۳-۶-۱- انتخاب بهترین توصیف‌کننده‌ها توسط رگرسیون مرحله‌ای

برای انتخاب بهترین توصیف‌کننده‌ها به روش رگرسیون مرحله‌ای بایستی با استفاده از توصیف‌کننده‌ها مدل‌های اولیه‌ای ساخته شوند تا از مقایسه پارامترهای آماری این مدل‌ها بتوان بهترین توصیف‌کننده‌ها را انتخاب نمود.

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 همبستگی، متغیر اولیه انتخاب می‌شود. سپس مدل دو متغیری ایجاد شده و پارامترهای آماری مدل حاصل شامل ضریب تعیین، مقدار میانگین مربعات خطا^۳ (MSE) و آماره‌ی F برای سری ارزیابی مورد بررسی قرار می‌گیرند. سپس مدل سه متغیری ساخته می‌شود و دوباره پارامترهای آماری آن برای سری ارزیابی مورد بررسی قرار می‌گیرد. این کار تا

۱- Stepwise Regression

۲- Combined Data Splitting-Feature Selection Strategy

۳- Mean Square Error

جایی ادامه پیدا می‌کند که تعداد توصیف‌کننده‌های انتخابی از نظر مدل‌سازی ارزش محاسباتی داشته باشند. مدلی که ضریب تعیین و آماره‌ی F آن بیشتر باشد و مقدار میانگین مربعات خطای آن کمتر باشد به عنوان بهترین مدل انتخاب شده و توصیف‌کننده‌های موجود در مدل به عنوان توصیف‌کننده‌های بهینه با روش رگرسیون مرحله‌ای در نظر گرفته می‌شوند [۱۳].

۲-۳-۶-۲- انتخاب بهترین توصیف‌کننده‌ها به روش CDFS

همان‌طور که قبلاً اشاره شد برای ایجاد یک مدل QSPR داده‌ها باید دسته‌بندی شوند. با توجه به این‌که تمام مراحل ساخت مدل، به ویژه انتخاب بهترین توصیف‌کننده‌ها با استفاده از این دسته‌بندی انجام می‌شود لذا عملکرد مدل نهایی شدیداً به دسته‌بندی اولیه‌ی داده‌ها وابسته است. برای رفع این مشکل و به دست آوردن مدلی که مستقل از تأثیر نحوه‌ی دسته‌بندی داده‌ها باشد، روش CDFS ارائه شده است [۲۲ و ۲۳]. در این روش ابتدا داده‌های سری تست، کنار گذاشته می‌شوند و مابقی داده‌ها چندین بار به طور تصادفی به دو دسته‌ی آموزش و ارزیابی تقسیم‌بندی می‌شوند. با هر بار دسته‌بندی باید تمام مراحل مدل‌سازی انجام گیرد و بهترین مدل که شامل توصیف‌کننده‌های بهینه است، انتخاب شود. بنابراین اگر تعداد دفعات دسته‌بندی داده‌ها n باشد، n مدل بهینه به دست می‌آید. در نهایت توصیف‌کننده‌هایی که در بیش از ۵۰ درصد مدل‌ها تکرار شده‌اند، به عنوان بهترین توصیف‌کننده‌ها انتخاب می‌شوند.

۲-۳-۷- ساختن مدل

پس از انتخاب بهترین توصیف‌کننده‌ها، نوبت به ساخت مدل می‌رسد. مدل، یک رابطه‌ی ریاضی است که بیان‌کننده‌ی ارتباط بین خاصیت مورد نظر به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیر مستقل است. با کمک مدل ساخته شده می‌توان با داشتن مقادیر متغیرهای

مستقل، متغیر وابسته را برای ترکیبات جدید پیش‌بینی کرد.

برای مدل‌سازی روش‌های خطی و غیرخطی مختلفی وجود دارد که رگرسیون خطی چندگانه جزء روش‌های خطی بوده و روش‌های شبکه‌ی عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) از جمله روش‌های غیرخطی هستند. در این پایان نامه از دو روش غیرخطی ANN و SVM برای مدل‌سازی γ_{13}^{∞} در مایع یونی [TCM][BMMOR] استفاده شده است. بنابراین این دو روش در ادامه به اختصار شرح داده خواهد شد.

۲-۳-۷-۱- شبکه‌ی عصبی مصنوعی

در بسیاری از کاربردهای شیمیایی رابطه‌ی بین متغیرهای مستقل و پاسخ را نمی‌توان به‌صورت یک رابطه‌ی ساده‌ی خطی در نظر گرفت. بنابراین باید از روش‌های مدل‌سازی غیرخطی استفاده نمود. در این روش‌ها داده‌های اولیه به وسیله‌ی یک تابع انتقال غیرخطی از فضای اولیه به فضایی با ابعاد بیشتر منتقل می‌شوند که در فضای جدید، ساختن مدل آسان‌تر صورت می‌گیرد [۲۴].

شبکه‌ی عصبی مصنوعی یک برنامه نرم‌افزاری است که می‌تواند همانند مغز انسان عمل نماید. در واقع شبکه‌ی عصبی مصنوعی ایده‌ای برای پردازش اطلاعات است که از سیستم‌های عصبی زیستی الهام گرفته شده و مانند مغز انسان به پردازش اطلاعات می‌پردازد [۱۸]. شبکه‌های عصبی با پردازش روی داده‌های تجربی، قانون نهفته در ورای آن‌ها را کشف می‌کنند و قوانین کلی را فرا می‌گیرند. به همین دلیل به این سیستم‌ها، هوشمند گفته می‌شود.

۲-۳-۷-۱-الف- سیستم عصبی زیستی

مغز انسان یک سیستم پردازش اطلاعات با ساختار موازی و کاملاً پیچیده است که از حدود

^{۱۱} ۱۰ (صد میلیارد) سلول عصبی به نام نرون^۱ تشکیل شده است که هر نرون به ۱۰۰۰۰ (ده هزار) نرون دیگر مرتبط می‌باشد [۲۵]. نرون اصلی‌ترین عنصر پردازش^۲ اطلاعات به شمار می‌آید که متشکل از دندریت^۳، بدنه‌ی سلول (شامل هسته و قسمت‌های حفاظتی دیگر) و اکسون^۴ است. دندریت‌ها به عنوان مناطق دریافت سیگنال‌ها، شبکه‌های متشکل از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بی‌شماری می‌باشند. به همین علت آن‌ها را شبکه‌های دریافتی «درخت گونه» نیز می‌گویند. دندریت‌ها پس از دریافت اطلاعات به شکل سیگنال، آن را به هسته‌ی سلول که در بدنه‌ی سلول قرار دارد، هدایت می‌کنند. بدنه‌ی سلول ضمن فراهم کردن انرژی لازم برای فعالیت نرون، بر روی سیگنال‌های دریافتی عملیات لازم را انجام داده که این عملیات شبیه یک عمل جمع ساده بر روی کل سیگنال‌ها و مقایسه‌ی آن با یک سطح آستانه می‌باشد. اکسون که برخلاف دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتر ولی طول بیشتر برخوردار است، سیگنال‌های الکتروشیمیایی دریافتی از بدنه‌ی سلول را توسط سیناپس‌ها که ارتباط دهنده‌ی نرون‌ها هستند به نرون‌های دیگر منتقل می‌کند و به این ترتیب فعالیت‌های مغزی انجام می‌شود.

۲-۳-۷-۱-ب- ساختار شبکه‌ی عصبی مصنوعی

شبکه‌های عصبی مصنوعی می‌توانند با روشی مشابه با مغز، اطلاعات را پردازش کنند. این شبکه‌ها بی‌آن‌که سعی در حفظ پیچیدگی مغز داشته باشند، مشابه مغز از تعداد زیادی سلول عصبی (نرون) تشکیل شده‌اند که از طریق ارتباطاتی به هم متصل شده‌اند (سیناپس^۵). هر نرون محاسباتی شامل ورودی، جسم سلولی و یک خروجی است. همان‌طور که در شکل (۲-۱) نشان

^۱-Neuron

^۲-Processing Element

^۳-Dendrite

^۴-Axon

^۵-Synapse

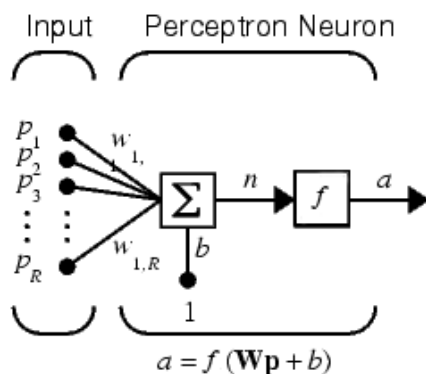
داده شده است هر ورودی قبل از ورود به هسته‌ی پردازشگر اصلی، وزن‌دار می‌شود و سپس ورودی‌های وزن‌دار شده با هم جمع می‌شوند. علاوه بر تمام ورودی‌های وزن‌دار شده، ورودی دیگری نیز در نظر گرفته شده و آن عدد یک است. این عدد در بایاس ضرب شده و با حاصل ضرب ورودی‌ها در وزن‌ها جمع می‌شود تا ورودی خالص را ایجاد کند.

$$n = \mathbf{WP} + b \quad (1-2)$$

ورودی خالص پس از عبور از تابع انتقال^۱ به سیگنال خروجی نرون تبدیل می‌شود.

$$a = f(n) = f(\mathbf{WP} + b) \quad (2-2)$$

هدف اصلی شبکه‌های عصبی این است که با تغییر مقادیر وزن و بایاس شبکه، عملکرد شبکه را بهینه کند.



شکل (۱-۲) - ساختار یک نرون محاسباتی با چند ورودی

توابع انتقالی که بیشترین استفاده را در شبکه‌های عصبی دارند عبارتند از: تابع انتقال آستانه‌ای حدی^۲، تابع انتقال خطی^۳ (Purline)، تابع انتقال لگاریتم سیگموئیدی^۴ (Logsig) و

^۱-Transfer function

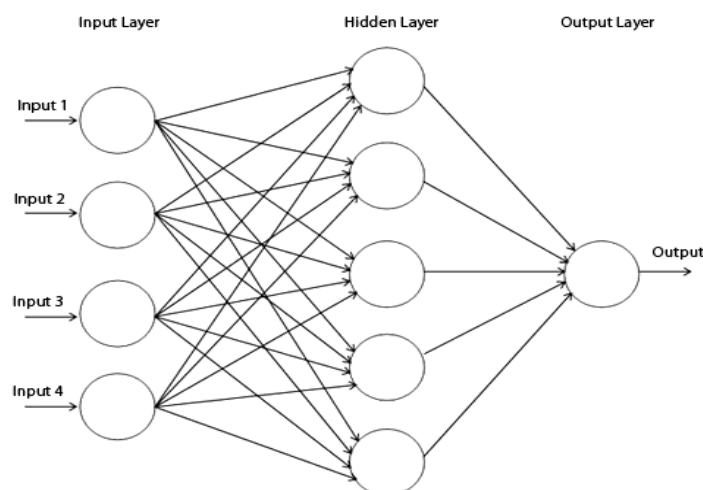
^۲- Hard Limit Transfer Function

^۳-Linear Transfer Function

^۴-Log-Sigmoid Transfer Function

تابع انتقال تانژانت سیگموئیدی^۱ (Tansig). در این پایان نامه از دو تابع انتقال Logsig و Tansig برای لایه‌ی پنهان و از تابع انتقال Purrline برای لایه خروجی استفاده شده است. در تقسیم‌بندی ساختاری، شبکه‌های عصبی را می‌توان به دو دسته‌ی تک لایه و چند لایه تقسیم کرد. در شبکه‌ی تک لایه نرون‌های خروجی و ورودی یکی هستند پس به تعداد سیگنال‌های خروجی، نرون در شبکه وجود دارد اما در شبکه‌های چند لایه، خروجی نرون‌های لایه‌ی قبل به ورودی نرون‌های لایه‌ی بعد منتقل می‌شود. در این شبکه‌ها، لایه‌ای که خروجی شبکه را ایجاد می‌کند یا همان لایه‌ی آخر به عنوان لایه‌ی خروجی، لایه‌ی اول به عنوان لایه‌ی ورودی و بقیه لایه‌ها تحت عنوان لایه‌های مخفی نام‌گذاری می‌شوند. شکل (۲-۲) نمای ساده‌ای از یک شبکه‌ی عصبی چند لایه را نشان می‌دهد.

اکثر شبکه‌های مورد استفاده ۲ و ۳ لایه می‌باشند و به ندرت از شبکه‌های با ۴ لایه و یا بیشتر استفاده می‌شود. استفاده از بایاس در شبکه دلخواه می‌باشد. اما استفاده از بایاس با فراهم آوردن یک پارامتر بیشتر منجر به قوی‌تر شدن شبکه می‌شود [۲۶].



شکل (۲-۲) - یک شبکه‌ی عصبی مصنوعی چند لای

۱- Hyperbolic Tangent Sigmoid Transfer Function

۲-۳-۷-۱-ج- روشهای آموزش

شبکه‌های عصبی مصنوعی می‌توانند آموزش ببینند و یاد بگیرند. به طور کلی دو نوع یادگیری برای شبکه‌ها وجود دارد: آموزش با ناظر^۱ و آموزش بدون ناظر^۲.

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.

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 [۲۷].

برای حداقل شدن خطای بین مقادیر تجربی و پاسخ شبکه باید عملکرد شبکه را بهینه نمود. برای بهینه‌سازی عملکرد شبکه، ابتدا باید یک شاخص عملکرد معرفی شود. 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) 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 [۱۳].

۱-Supervised learning

۲-Unsupervised learning

از رایج‌ترین الگوریتم‌های آموزشی، الگوریتم‌های لونی‌گ - مارکوات و تنظیم بایزین است. این الگوریتم‌ها روش مناسبی برای آموزش شبکه‌هایی هستند که شاخص عملکرد آن‌ها حداقل متوسط مربعات خطا است. این روش‌ها به عنوان سریع‌ترین و مؤثرترین روش‌های آموزش شبکه‌های عصبی شناخته شده‌اند. سرعت همگرایی این روش صدها برابر بیشتر از شیوه‌های استاندارد مشابه است. در این تحقیق نیز از این دو الگوریتم برای بهینه‌سازی عملکرد شبکه استفاده شده است [۲۰].

۲-۳-۷-۲- ماشین بردار پشتیبان

ماشین‌های بردار پشتیبان (SVM) یکی دیگر 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. الگوریتم ماشین بردار پشتیبان اولیه در سال ۱۹۶۳ توسط وپنایک^۱ ابداع شد و در سال ۱۹۹۵ توسط وپنایک و کورینا^۲ برای حالت غیر خطی تعمیم داده شد. از الگوریتم‌های آموزش دیده توسط ماشین‌های برداری پشتیبان می‌توان برای موارد زیر استفاده 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

۱- Vapnik

۲- Corinna

است.

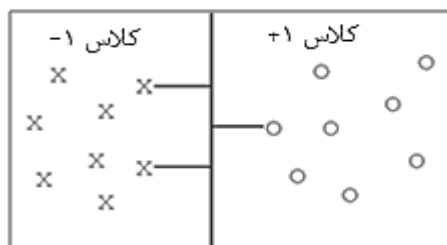
۲- در آموزش شبکه‌های عصبی از حداقل‌سازی خطای تجربی برای یافتن بهترین مدل استفاده می‌کنند؛ یعنی، شبکه خطای داده‌های آموزشی را حداقل می‌کند اما در SVM از حداقل‌سازی خطای ساختاری برای یافتن بهترین مدل استفاده می‌شود که در آن خطای تعمیم حداقل می‌شود.

۲-۳-۷-۲-الف- طبقه‌بندی خطی دو کلاسه با ماشین‌های بردار پشتیبان

فرض کنید دو دسته داده‌ی آموزشی که به صورت خطی از هم جداپذیر باشند در اختیار داریم. برای این داده‌ها تعداد زیادی خط جداکننده وجود دارد که قادر هستند داده‌ها را به صورت خطی جدا کنند اما همان‌طور که در شکل (۳-۲) 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:

$$D = \left\{ (X_i, y_i) \mid X_i \in R^n, y_i \in [-1, 1] \right\}_{i=1}^n \quad (3-2)$$

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 نقاط x_i به کدام طبقه تعلق دارند. هدف پیدا کردن ابرصفحه‌ی جداکننده^۱ با بیشترین فاصله از نقاط حاشیه‌ای است که نقاط با $y_i = +1$ را از نقاط با $y_i = -1$ جدا کند.



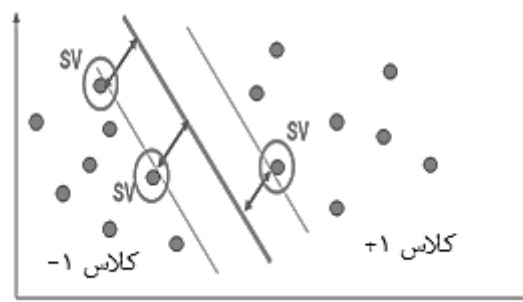
شکل (۳-۲): خط جداساز بهینه با حداکثر مقدار حاشیه

^۱-Separating hyperplane

به نزدیک‌ترین داده‌های آموزشی به ابرصفحه‌ی جداساز بهینه که در واقع روی ابرصفحه‌های موازی مرزی قرار گرفته‌اند، بردارهای پشتیبان^۱ می‌گویند. این داده‌ها که در شکل (۴-۲) با SV نشان داده شده‌اند، معیاری هستند که ماشین برداری پشتیبان از آن‌ها برای طبقه‌بندی صحیح داده‌ها استفاده می‌کند. در حالت دو بعدی معادله‌ی این خط به صورت زیر نوشته می‌شود:

$$W_1X_1 + W_2X_2 + b = 0 \quad (۴-۲)$$

ابر صفحه‌ی جداساز بهینه



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

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:

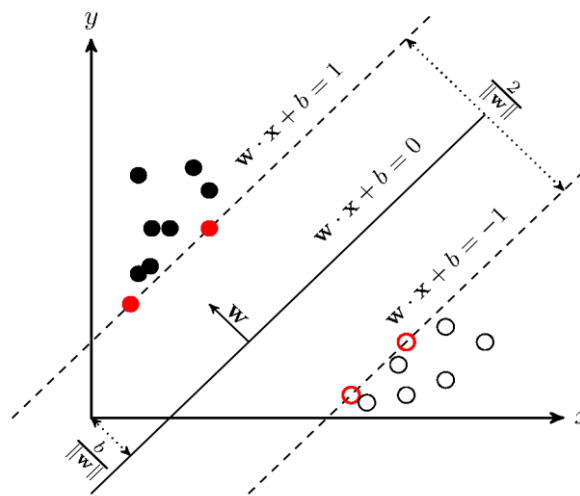
$$\mathbf{W}^T \cdot \mathbf{X} + b = 0 \quad (۵-۲)$$

با توجه به رابطه‌ی (۵-۲) که در آن $\mathbf{W}$ بردار نرمال ابر صفحه‌ی جداساز بهینه و b عرض از مبدأ آن است. علامت نقطه، " . "، بین بردارهای $\mathbf{X}$ و $\mathbf{W}$ به معنی ضرب داخلی این دو بردار بوده و علامت T بالای بردار $\mathbf{W}$ ، ترانهاده‌ی این بردار می‌باشد.

^۱-Support vectors (SV)

باید $\mathbf{W}$ 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 بین می‌شود :

$$\mathbf{W}^T \cdot \mathbf{X} + b = \pm 1 \quad 1 \leq i \leq n \quad (۶-۲)$$



شکل (۵-۲) - صفحه‌ی جداساز و حاشیه‌ها

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

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

$$d = \frac{|(\mathbf{W}^T \cdot \mathbf{X} + b - 1) - (\mathbf{W}^T \cdot \mathbf{X} + b + 1)|}{\|\mathbf{W}\|} = \frac{2}{\|\mathbf{W}\|} \quad (۷-۲)$$

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

$$y_i (\mathbf{W}^T \cdot \mathbf{X} + b) \geq 1 \quad 1 \leq i \leq n \quad (۸-۲)$$

حل این مسئله‌ی بهینه‌سازی به دلیل وابستگی به $\|\mathbf{W}\|$ مشکل است، لذا برای سادگی کار می‌توان بدون تغییر در مسئله، به جای تابع $\frac{2}{\|\mathbf{W}\|}$ ، تابع $\frac{1}{2}\|\mathbf{W}\|^2$ را با در نظر گرفتن قید (۸-۲) مینیمم کرد که این تابع به صورت $\frac{1}{2}\mathbf{W}^T \cdot \mathbf{W}$ نوشته می‌شود. برای حل این مسئله از روش ضرایب لاگرانژ استفاده می‌شود. تابع لاگرانژ برای این مسئله به صورت رابطه‌ی (۹-۲) نوشته می‌شود:

$$L_p(\mathbf{W}, b, \alpha) = \frac{1}{2} \mathbf{W}^T \cdot \mathbf{W} - \sum_{i=1}^n \alpha_i [y_i (\mathbf{W}^T \cdot \mathbf{X}_i + b) - 1] \quad (9-2)$$

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

(۱۰-۲)

$$\frac{\partial L}{\partial \mathbf{W}} = 0 \Rightarrow \mathbf{W} = \sum_{i=1}^N \alpha_i \mathbf{X}_i y_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (11-2)$$

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

$$L_d(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j \mathbf{X}_i^T \mathbf{X}_j \quad (12-2)$$

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

$$\alpha_i \geq 0 \quad (13-2)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (14-2)$$

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

$$\mathbf{w}^T \cdot \mathbf{X}_i + b = \frac{1}{y_i} = y_i \Rightarrow b = y_i - \mathbf{w}^T \cdot \mathbf{X}_i \quad (۱۵-۲)$$

البته بهتر است که با متوسط‌گیری روی همه‌ی بردارهای پشتیبان، مقدار بهینه‌ی b را

محاسبه کرد تا در عمل الگوریتم مقاوم‌تری به دست آید:

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (y_i - \mathbf{w}^T \cdot \mathbf{X}_i) \quad (۱۶-۲)$$

با حل مسئله‌ی بهینه‌سازی (۱۲-۲) و استفاده از رابطه‌ی (۱۶-۲) می‌توان به بهینه‌ترین

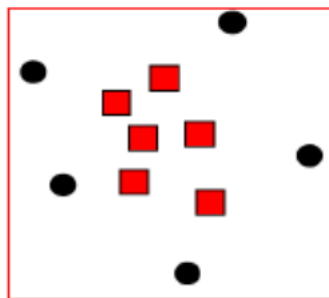
ابرفضه‌ی جداساز دست یافت و سپس از این ابرصفحه جداساز برای طبقه‌بندی نمونه‌های

جدید استفاده نمود.

۲-۳-۷-۲-ب- طبقه بندی غیر خطی^۱ با ماشین بردار پشتیبان

گاهی با مجموعه داده های پیچیده ای روبرو هستیم که به طور خطی قابل جداسازی نیستند.

مثلا در شکل (۶-۲) مجموعه داه ها را نمی توان به سادگی با یک خط دسته بندی کرد.



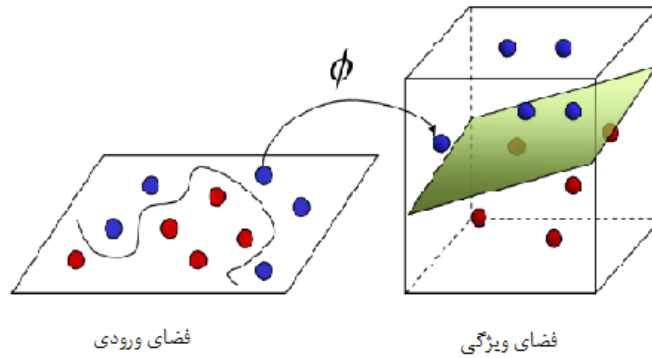
شکل (۶-۲) - مجموعه داده های پیچیده

در این گونه موارد ابتدا بردارهای ورودی به فضایی با ابعاد بالاتر که فضای ویژگی^۱ نام دارد،

نگاشته می‌شود. در این فضای جدید، نگاشت بردار $\mathbf{X}_i$ به صورت $\phi(\mathbf{X}_i)$ نمایش داده می‌شود.

^۱-Nonlinear Classification

سپس در این فضای بالاتر ماشین بردار پشتیبان می‌تواند به جداسازی داده‌ها بپردازد در حالی که فضای ورودی در همان حالت غیرخطی باقی مانده است. این فرایند در شکل (۷-۲) نشان داده شده است.



شکل (۷-۲) - داده‌های ورودی ارجاع داده شده به فضای بالاتر

بنابراین روابط موجود در سیستم‌های خطی را باید در فضای بالاتر به صورت زیر بازنویسی

کرد:

(۱۷-۲)

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \phi(\mathbf{X}_i)$$

$$b = y_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i y_i \phi(\mathbf{X}_i)^T \cdot \phi(\mathbf{X}_j) \quad (18-2)$$

به منظور کم کردن حجم محاسبات به دلیل زیاد شدن ابعاد، تئوری مرسر^۲ در سال ۱۹۹۹

ضرب داخلی $\phi(\mathbf{X}_i) \cdot \phi(\mathbf{X}_j)$ را به صورت تابع کرنل $K(\mathbf{X}_i, \mathbf{X}_j)$ در فضای ویژگی معرفی کرد

که به آن کرنل مرسر می‌گویند. در نتیجه رابطه‌ی (۱۸-۲) به صورت زیر تغییر می‌یابد :

$$b = y_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i y_i K(\mathbf{X}_i, \mathbf{X}_j) \quad (19-2)$$

۱-Feature space

۲-Mercer

با استفاده از ترفند کرنل می‌توان معادله‌ی ابرصفحه‌ی جداساز بهینه در سیستم‌های

غیرخطی را به صورت معادله‌ی زیر نوشت:

$$d(x) = \sum_{i=1}^N \sum_{j=1}^N y_i \alpha_i K(\mathbf{X}_i, \mathbf{X}_j) + b \quad (2-20)$$

همان‌گونه که مشاهده می‌شود، مقدار معادله‌ی صفحه‌ی جداساز را می‌توان بدون محاسبه

$\phi(\mathbf{X}_i)$ در سیستم‌های غیرخطی به دست آورد و تنها کافیست که از کرنل مناسب برای حل

معادله‌ی صفحه استفاده شود.

مزایای تابع کرنل عبارتند از:

- حل مسائل محاسباتی که دارای ابعاد زیادی هستند، با سرعت زیادی امکان‌پذیر

می‌شود

- از لحاظ زمانی و حافظه کارآمد هستند.

- استفاده از ابعاد نامتناهی را ممکن می‌سازد

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 (2-2) مشخص شده‌اند.

جدول (2-2): توابع کرنل در فضای ویژگی

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

۲-۳-۷-۲-ج- بردار پشتیبانی رگرسیونگر برای سیستم های خطی

استفاده از SVM برای آنالیز رگرسیون در سال ۱۹۹۶ توسط وپنایک و همکارانش پیشنهاد شد که این روش بردار پشتیبانی رگرسیونگر^۱ (SVR) نامیده می شود. هدف SVR پیدا نمودن تابعی است که بر داده با کمترین انحراف از کمیتی مانند ε برای هر جفت X_i, y_i برازش یابد. به عبارت دیگر، تابع رگرسیون در این حالت یک ابر صفحه جداساز است که بر روی داده ها با در نظر گرفتن کمترین مربع خطا بین ابر صفحه و داده ها، برازش می یابد. بنابراین تابع رگرسیون به صورت زیر بیان می شود.

$$f(x) = \mathbf{W}^T \cdot \mathbf{X} + b \quad (2-21)$$

در این جا نیز تلاش می شود که کمترین مقدار ممکن برای $\|\mathbf{W}\|$ با توجه به دو قید (2-22) و (2-23) به دست آید.

$$y_i - \mathbf{W}^T \cdot \mathbf{X} - b \leq \varepsilon \quad (2-22)$$

$$y_i + \mathbf{W}^T \cdot \mathbf{X} + b \leq \varepsilon \quad (2-23)$$

برخی مواقع، برخی خطاها ممکن است از مقدار ε بیشتر باشد، بنابراین برای این حالت تابع

حساسیت 2ε یا متغیر ξ_i معرفی می شود که به صورت زیر بیان می شود:

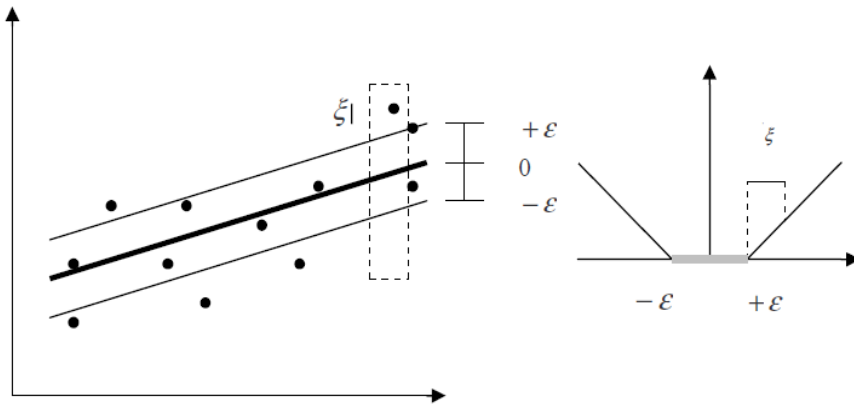
$$(2-24)$$

$$|\xi_i| = \begin{cases} 0 & \text{if } |\xi_i| < \varepsilon \\ |\xi_i| - \varepsilon & \text{otherwise} \end{cases}$$

^۱-Support Vector Regression

^۲- ε - sensitivity

این حالت در شکل (۸-۲) نمایش داده شده است.



شکل (۸-۲): تابع حساسیت به مقدار ε .

بنابراین رگرسیون خطی برای ماشین‌های برداری می‌تواند به صورت تابع اولیه (۲۵-۲) نوشته شود که در آن هدف، مینیمم‌سازی خطای کلی و $\|W\|$ با توجه به قیدهای (۲۶-۲)، (۲۷-۲) و (۲۸-۲) است.

$$L_p = \frac{1}{2} \|W\| + C \sum_{i=1}^N (\xi_i + \xi'_i) \quad (25-2)$$

$$y_i - W^T \cdot X - b \leq \xi_i + \varepsilon \quad (26-2)$$

$$y_i + W^T \cdot X + b \leq \xi'_i + \varepsilon \quad (27-2)$$

$$\xi_i, X_i \geq 0 \quad (28-2)$$

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.

در رابطه (۲۵-۲) متغیرهای ξ_i و ξ'_i جهت نشان دادن داده‌هایی است که خارج از محدوده مجاز در نظر گرفته شده حول خط تابع قرار گرفته‌اند. داده‌هایی که بالاتر از $\varepsilon +$ قرار دارند با ξ_i و داده‌هایی که پایین‌تر از $\varepsilon -$ قرار دارند با ξ'_i نشان داده می‌شوند. C ثابت

تنظیم است و مصالحه بین خطای تجربی و پیچیدگی مدل را برقرار می کند. همچنین ε بیانگر فاصله $\varepsilon -$ و $\varepsilon +$ حول خط رگرسیون می باشد، که مشخص کننده میزان صحت داده ها است. به طوری که داده هایی که در این فاصله قرار بگیرند، خطا محاسبه نشده و داده های خارج از این محدوده به عنوان مقادیر خطا محسوب می شوند [۲۸].

برای حل مسئله ی بهینه سازی دارای محدودیت (۲-۲۵)، باید از بهینه سازی لاگرانژ برای تبدیل این معادله به یک معادله ی بدون محدودیت استفاده نمود. با در نظر گرفتن تابع لاگرانژ و مشتق گیری از تابع هدف بدون محدودیت نسبت به دو پارامتر W و b ، دو معادله به صورت (۲-۲۹) و (۲-۳۰) به دست می آیند.

$$W = \sum_{i=1}^N (\alpha_i - \alpha'_i) X_i \quad (2-29)$$

$$\sum_{i=1}^N (\alpha_i - \alpha'_i) = 0 \quad (2-30)$$

با قرار دادن رابطه ی به دست آمده از معادله ی (۲-۲۹) برای مقدار W ، در معادله ی به دست آمده از تابع لاگرانژ، معادله ی اساسی ماشین های برداری رگرسیونگر به صورت معادله ی زیر نوشته خواهد شد:

$$L_d = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha'_i) X_i^T \cdot X_j (\alpha_i - \alpha'_i) + \sum_{i=1}^N ((\alpha_i - \alpha'_i) y_i - (\alpha_i + \alpha'_i) \varepsilon) \quad (2-31)$$

که در آن :

$$0 \leq (\alpha_i - \alpha'_i) \leq C \quad (2-32)$$

۲-۳-۷-۲-د-بردار پشتیبانی رگرسیونگر برای سیستم های غیر خطی

همان گونه که در بحث برداری پشتیبان رگرسیونگر در سیستم های خطی بیان شد، مقدار

بهینه ی W از رابطه ی $W = \sum_{i=1}^N (\alpha_i - \alpha'_i) X_i$ به دست می آید که این رابطه در سیستم های

غیرخطی تبدیل به رابطه‌ی $W = \sum_{i=1}^N (\alpha_i - \alpha'_i) \phi(X_i)$ 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 $\phi(X_i)$ است که مقدار آن ناشناخته است. در این شرایط مانند حالت طبقه‌بندی داده‌ها از حقه‌ی کرنل استفاده می‌شود تا بتوان بدون محاسبه‌ی مقدار $\phi(X_i)$ 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 $y_i = \sum_{i=1}^N \mathbf{w}^T \phi(X_i) + b$ بیان می‌گردد. با قراردادن رابطه‌ی به دست آمده برای W در رابطه‌ی صفحه‌ی جداساز، می‌توان معادله‌ی اساسی (۳۳-۲) را برای ماشین‌های برداری رگرسیونگر در سیستم‌های غیرخطی به صورت زیر معرفی نمود.

$$y_i = \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha'_i) \phi(X_i)^T \phi(X_j) + b = \sum_{i=1}^N \sum_{j=1}^N (\alpha - \alpha'_i) K(X_i, X_j) + b \quad (33-2)$$

با در نظر گرفتن رابطه‌ی (۳۳-۲) نیازی به محاسبه مقدار $\phi(X_i)$ نیست و مقدار b از این رابطه محاسبه می‌شود.

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 کار گرفته شده، خصوصاً در سیستم‌های غیرخطی است.

۲-۳-۷-۲-۵-آموزش ماشین بردار رگرسیون

نوع تابع کرنل و همچنین پارامترهای موازنه‌ی C و تابع حساسیت ϵ که قبلاً به آنها اشاره شد، باید به دقت توسط کاربر بهینه و انتخاب شوند. چون کارایی تعمیم پذیری ماشین بردار رگرسیون تا حد زیادی بستگی به انتخاب مناسب این سه پارامتر دارد.

مقدار خیلی زیاد ε موجب می شود تعداد داده های بیشتری در محدوده $\varepsilon -$ تا $\varepsilon +$ قرار گیرند. در نتیجه تابع رگرسیونی با پیچیدگی کمتری ایجاد می شود. پارامتر دیگری که باید بهینه شود، پارامتر موازنه C است که پارامتر ظرفیت یا تنظیم کننده نیز نامیده می شود. توانمندی مدل رگرسیون به انتخاب صحیح مقدار C نیز بستگی دارد. با توجه به رابطه ی (۳۲-۲) چون مقدار $(\alpha_i - \alpha'_i)$ مترادف با مقدار C است، این بدان معنی است که انتخاب مقدار C متاثر از اهمیت هر یک از داده ها است. به طور مثال از یک طرف، مقدار زیاد C منجر به اختلاف زیاد بین مقدار $(\alpha_i - \alpha'_i)$ در بردارهای پشتیبان خواهد شد که در این مورد بردارهای پشتیبان با مقدار بالای $(\alpha_i - \alpha'_i)$ در مدل رگرسیون ایجاد شده تعیین کننده خواهد بود. از طرف دیگر، مقدار کم C می تواند منجر به اختلاف کمی در بردارهای پشتیبان شده یا حتی مقدار $(\alpha_i - \alpha'_i)$ یکسان شود که در این مورد داده های انتخاب شده به عنوان بردارهای پشتیبان در تابع رگرسیون توزیع یکسانی خواهند داشت [۲۸].

جهت دستیابی به یک مدل ماشین بردار پشتیبان قدرتمند با قابلیت پیش بینی خوب نیاز است که تابع کرنل مناسب به همراه پارامترهای قابل تنظیم مربوط به آن و همچنین پارامترهای موازنه C و پارامتر تابع حساسیت ε به درستی بهینه شوند.

۳-۳-۸- ارزیابی قدرت پیش بینی مدل

۳-۳-۸-۱- استفاده از مدل های آماری

برای اطمینان از این که مدل به دست آمده توانایی پیش بینی نمونه های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی با محاسبه پارامترهای آماری صورت می گیرد. روابط ریاضی پارامترهای آماری استفاده شده در این پایان نامه، در ادامه توضیح داده خواهد شد.

ضریب تعیین (R^2): به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان دهنده‌ی نسبت تغییرات متغیر وابسته نسبت به متغیر مستقل است. به عنوان مثال R^2 برابر ۰/۹۶۶۰ نشان می‌دهد ۹۶/۶۰ درصد تغییرات در متغیر وابسته می‌تواند توسط متغیر مستقل توضیح داده شود. رابطه‌ی ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (۳۳-۲)$$

که در آن SSR^1 طبق رابطه‌ی (۳۴-۲) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است:

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (۳۴-۲)$$

و SST^2 طبق رابطه (۳۵-۲) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (۳۵-۲)$$

که در این روابط $\hat{y}_i$ مقدار پیش‌بینی شده‌ی متغیر وابسته، y_i مقدار واقعی متغیر وابسته و $\bar{y}$ در هر رابطه، میانگین مقادیر متغیر وابسته است. همچنین SSE^3 مبین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است و طبق رابطه‌ی زیر تعریف می‌شود.

$$SSE = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (۳۶-۲)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

^۱-Sum Square Regression

^۲- Sum Square Total

^۳- Sum Square Error

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (37-2)$$

طبق رابطه‌ی (37-2) اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشند، یعنی به ازای تمام نقاط $y_i = \hat{y}_i$ باشد، مقدار R^2 برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار R^2 از یک کوچکتر شود.

آماره‌ی F : آماره‌ی F یا آزمون فیشر در واقع آزمون معنی‌دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین مربعات رگرسیون (MSR) به میانگین مربعات خطاها (MSE) است. بیان ریاضی آماره F به صورت زیر می‌باشد:

$$F = \frac{MSR}{MSE} = \frac{SSR / df_m}{SSE / df_{res}} \quad (38-2)$$

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

میانگین مربع خطاها (MSE): طبق رابطه‌ی زیر به دست می‌آید.

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (39-2)$$

مجموع مربع باقیمانده‌ها^۱ ($PRESS$):

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (40-2)$$

میانگین خطای مطلق^۱ (MAE):

^۱ - Predictive Residual Sum of Squares

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}|}{n} \quad (۴۱-۲)$$

متوسط مطلق درصد خطای نسبی^۲ (MAPE):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad (۴۲-۲)$$

متوسط درصد خطای نسبی^۳ (MPE):

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{y_i - \hat{y}}{y_i} \quad (۴۳-۲)$$

ماکزیمم مطلق درصد خطای نسبی^۴ (D_{Max}):

$$D_{Max} = \text{Max} \left(100 \left| \frac{y_i - \hat{y}}{y_i} \right| \right) \quad (۴۴-۲)$$

در روابط (۳۹-۲) تا (۴۴-۲) (y_i) و ($\hat{y}_i$) به ترتیب مقدار واقعی و مقدار پیش‌بینی شده‌ی متغیر وابسته و n تعداد داده‌های وارد شده به مدل است.

۲-۳-۸-۲- استفاده از نمودار برگشتی

۱- Mean Absolute Error

۱- Mean Absolute Percentage Error

۲- Mean Percentage Error

۳- Maximum Absolute Percentage Error

در نمودار برگشتی مقادیر پیش بینی شده‌ی کمیت مورد نظر برحسب 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 ضریب تعیین به دست آمده بین ۰ و ۱ است. هرچه قدر به ۱ نزدیک‌تر باشد، نتایج پیش‌بینی شده به مقادیر واقعی نزدیک‌ترند. اما اگر این مقدار برابر صفر باشد، بین داده‌ها هیچ گونه همبستگی وجود ندارد.

۲-۳-۸-۳- 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 امر نشان دهنده‌ی تصادفی بودن خطاست. ولی اگر عمده‌ی نقاط در این نمودار، در یک طرف خط صفر باشند، این بدان معناست که خطای جهت‌داری رخ داده است.

۲-۳-۸-۴- استفاده از سری تست خارجی

اعتبار و اهمیت مدل‌های پیش‌بینی کننده وقتی مشخص می‌شود که بتوانند خاصیت مورد نظر را برای مولکول‌های جدیدی که در دسته‌ی آموزش موجود نیستند پیش‌بینی کنند. بدین منظور از ابتدای کار تعدادی از مولکول‌ها به عنوان سری تست خارجی کنار گذاشته می‌شود و در نهایت از مدل منتخب بهینه شده، جهت پیش‌بینی خاصیت مورد نظر مولکول‌های موجود در سری تست خارجی که در مدل‌سازی استفاده نشده‌اند، استفاده می‌شود.

فصل سوم

پیش‌بینی ضریب فعالیت در رقت به‌نهایت

ترکیبات آله و آب در محیط مایع یونک

1- بوتیل 1- متیل مورفولینیوم تری سیانو متانید

با استفاده از روش‌های غیرخطی

در دو فصل قبل کمیت ضریب فعالیت در رقت بی‌نهایت و اهمیت پیش‌بینی آن و همچنین مراحل و روش‌های مدل‌سازی معرفی شدند. در این فصل از تکنیک‌های رگرسیون مرحله‌ای و CDFS به عنوان روش‌های انتخاب توصیف‌کننده و از روش‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان جهت مدل‌سازی غیرخطی برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت ترکیبات آلی و آب در محیط مایع یونی ۱-بوتیل ۱-متیل مورفولینیوم تری سیانومتانید [TCM][BMMOR]، در ۶ دمای مختلف استفاده گردید. در ادامه به بررسی مراحل طراحی، بهینه‌سازی و ارزیابی مدل‌های ساخته‌شده خواهیم پرداخت.

۳-۱- مراحل مدل‌سازی

۳-۱-۱- سری داده‌ها

ضریب فعالیت متوسط در رقت بی‌نهایت ۶۰ ترکیب آلی شامل: آلکان‌ها، آلکن‌ها، آلکین‌ها، سیکلوآلکان‌ها، ترکیبات آروماتیک، الکل‌ها، اترها، کتون‌ها، استرها، استونیتریل، نیتروپروپان و آب در مایع یونی [TCM][BMMOR] در شش دمای مختلف به عنوان سری داده‌ها جمع‌آوری گردید که این مجموعه داده‌ها شامل ۳۶۶ نقطه داده‌ی ضریب فعالیت در رقت بی‌نهایت می‌باشد. نام این ترکیبات به همراه مقدار ضریب فعالیت در رقت بی‌نهایت، γ_{13}^{∞} ، در شش دما در جدول (۳-۱) نشان داده شده است.

جدول (۱-۳)- مقدار ضریب فعالیت در رقت بی‌نهایت (γ_{13}^{∞}) برای حل‌شونده‌های مختلف در [TCM][BMMOR]

Solute γ_{13}^{∞}	T, K					
	۳۱۸/۱۵	۳۲۸/۱۵	۳۳۸/۱۵	۳۴۸/۱۵	۳۵۸/۱۵	۳۶۸/۱۵
Pentane	۳۹/۷	۳۵/۵	۳۲/۲	۲۹/۸	۲۷/۷	۲۶/۲
Hexane	۵۶/۵	۵۱/۱	۴۶/۷	۴۳/۲	۴۰/۳	۳۷/۲
3- Methylpentane	۵۱/۳	۴۶	۴۲/۳	۳۸/۹	۳۶/۳	۳۳/۵
2,2- Dimethylbutane	۵۶/۷	۴۹/۸	۴۴/۵	۴۰/۹	۳۷/۷	۳۴/۸
Heptane	۸۴	۷۶/۸	۶۹/۶	۶۳/۳	۵۸/۷	۵۳/۷
Octane	۱۲۵	۱۱۳	۱۰۲	۹۳/۴	۸۶/۵	۷۸/۶
2,2-4-trimethylpentane	۱۰۳	۹۳/۴	۸۴/۲	۷۶/۹	۷۰/۸	۶۴/۵
Nonane	۱۸۷	۱۶۸	۱۵۱	۱۳۷	۱۲۶	۱۱۵
Decane	۲۸۹	۲۵۱	۲۲۴	۲۰۲	۱۸۳	۱۶۶
Cyclopentane	۱۴/۷	۱۳/۷	۱۲/۸	۱۲/۱	۱۱/۴	۱۰/۹
Cyclohexane	۲۲/۶	۲۰/۸	۱۹/۳	۱۸	۱۶/۹	۱۶
Methylcyclohexane	۳۵/۲	۳۲/۳	۲۹/۷	۲۷/۶	۲۵/۷	۲۴
Cycloheptane	۲۸/۱	۲۶	۲۴	۲۲/۵	۲۱	۱۹/۷
Cyclooctane	۳۷	۳۴/۱	۳۱/۲	۲۹/۱	۲۷	۲۵/۱
Hex-1-ene	۲۲/۶	۲۱/۳	۲۰	۱۹/۱	۱۸/۲	۱۷/۴
Cyclohexene	۹/۴۴	۹/۰۸	۸/۷۱	۸/۴۶	۸/۱۶	۷/۹۲
Hept-1-ene	۳۴/۱	۳۲	۳۰/۲	۲۸/۷	۲۷/۲	۲۵/۷
Oct-1-ene	۵۲/۷	۴۹/۱	۴۵/۷	۴۳	۴۰/۶	۳۸/۱
Dec-1-ene	۱۱۸	۱۱۰	۱۰۱	۹۴/۵	۸۷/۸	۸۱/۴
Pent-1-yne	۳/۶۹	۳/۷۳	۳/۷۵	۳/۸۰	۳/۸۱	۳/۸۳
Hex-1-yne	۵/۴۷	۵/۵۰	۵/۵۰	۵/۵۴	۵/۵۳	۵/۵۱
Hept-1-yne	۸/۲۸	۸/۲۶	۸/۲۰	۸/۲۰	۸/۱۳	۸/۰۵
Oct-1-yne	۱۲/۸	۱۲/۶	۱۲/۴	۱۲/۲	۱۲	۱۱/۸
Benzene	۱/۵۳	۱/۵۵	۱/۵۷	۱/۵۸	۱/۶۱	۱/۶۲
Toluene	۲/۳۰	۲/۳۳	۲/۳۶	۲/۳۸	۲/۴۱	۲/۴۳
Ethylbenzene	۳/۷۴	۳/۷۴	۳/۷۵	۳/۷۵	۳/۷۵	۳/۷۵
o-Xylene	۲/۹۶	۲/۹۹	۳/۰۲	۳/۰۴	۳/۰۷	۳/۰۸
m-Xylene	۳/۵۴	۳/۵۶	۳/۶۱	۳/۶۱	۳/۶۲	۳/۶۲
p-Xylene	۳/۴۹	۳/۵۲	۳/۵۴	۳/۵۷	۳/۵۹	۳/۶۳
Styrene	۱/۸۲	۱/۸۶	۱/۹۰	۱/۹۳	۱/۹۶	۱/۹۹
α -Methylstyrene	۲/۸۰	۲/۸۹	۲/۹۹	۳/۰۷	۳/۱۵	۳/۲۲

Methanol	۰/۸۲۵	۰/۷۹۰	۰/۷۶۲	۰/۷۲۹	۰/۷۰۴	۰/۶۷۷
----------	-------	-------	-------	-------	-------	-------

ادامه جدول (۱-۳)

Solute γ_{13}^{∞}	T, K					
	۳۱۸/۱۵	۳۲۸/۱۵	۳۳۸/۱۵	۳۴۸/۱۵	۳۵۸/۱۵	۳۶۸/۱۵
Ethanol	۱/۲۶	۱/۱۹	۱/۱۲	۱/۰۶	۱/۰۱	۰/۹۶۵
Propan-1-ol	۱/۶۲	۱/۵۲	۱/۴۳	۱/۳۵	۱/۲۹	۱/۲۳
Propan-2-ol	۱/۷۱	۱/۵۹	۱/۵۰	۱/۴۰	۱/۳۳	۱/۲۵
Butan-1-ol	۲/۱۸	۲/۰۴	۱/۹۱	۱/۷۹	۱/۷۰	۱/۶۱
Butan-2-ol	۲/۰۵	۱/۹۳	۱/۸۱	۱/۷۱	۱/۶۴	۱/۵۸
2-Methyl-propan-1-ol	۲/۱۹	۲/۰۳	۱/۸۷	۱/۷۴	۱/۶۵	۱/۵۶
tert-Butanol	۲	۱/۸۸	۱/۷۷	۱/۶۹	۱/۶۳	۱/۵۸
Water	۱/۰۳	۰/۹۷۱	۰/۹۱۶	۰/۸۶۰	۰/۸۱۹	۰/۷۸۰
Thiophene	۱/۰۱	۱/۰۴	۱/۰۶	۱/۰۸	۱/۱۰	۱/۱۲
Tetrahydrofuran	۱/۱۲	۱/۱۵	۱/۱۷	۱/۱۸	۱/۲۱	۱/۲۲
1,4- Dioxane	۰/۷۰۷	۰/۷۴۰	۰/۷۷۲	۰/۷۹۸	۰/۸۳۱	۰/۸۵۳
tert-Buthyl methyl ether	۵/۷۴	۵/۷۷	۵/۷۴	۵/۷۶	۵/۷۷	۵/۷۸
Ethyl tert-butyl ether	۱۶/۸	۱۶/۲	۱۵/۶	۱۵/۳	۱۴/۹	۱۴/۶
Methyl tert-pentyl ether	۸/۶۱	۸/۵۲	۸/۴۲	۸/۳۷	۸/۲۹	۸/۲۱
Diethyl ether	۵/۳۹	۵/۳۳	۵/۲۷	۵/۲۴	۵/۲۱	۵/۱۶
Di-n-propyl ether	۱۶/۴	۱۵/۹	۱۵/۴	۱۴/۹	۱۴/۴	۱۳/۹
Di-iso-propyl ether	۱۷/۸	۱۷/۴	۱۶/۸	۱۶/۵	۱۶	۱۵/۶
Di-n-butyl ether	۳۹/۷	۳۷/۳	۳۵/۱	۳۳/۳	۳۱/۵	۲۹/۷
Acetone	۰/۷۴۱	۰/۷۵۰	۰/۷۶۵	۰/۷۷۴	۰/۷۹۱	۰/۷۹۸
Pentan-2-one	۱/۴۴	۱/۴۷	۱/۴۹	۱/۵۰	۱/۵۳	۱/۵۴
Pentan-3-one	۱/۴۲	۱/۴۴	۱/۴۷	۱/۴۹	۱/۵۱	۱/۵۴
Methyl acetate	۱/۲۰	۱/۲۳	۱/۲۴	۱/۲۶	۱/۲۹	۱/۳۰
Ethyl acetate	۱/۸۸	۱/۹۰	۱/۹۲	۱/۹۳	۱/۹۶	۱/۹۸
Methyl propionoate	۱/۶۴	۱/۶۷	۱/۷۰	۱/۷۳	۱/۷۶	۱/۷۹
Methyl butyrate	۲/۴۲	۲/۴۶	۲/۴۹	۲/۵۱	۲/۵۴	۲/۵۶
Butyraldehyde	۱/۲۰	۱/۲۴	۱/۲۶	۱/۳۰	۱/۳۳	۱/۳۵
Acetonitrile	۰/۶۱۷	۰/۶۲۰	۰/۶۲۴	۰/۶۲۴	۰/۶۲۹	۰/۶۳۰
Pyridine	۰/۷۳۳	۰/۷۴۸	۰/۷۵۸	۰/۷۶۶	۰/۷۷۶	۰/۷۸۸
1-Nitropropane	۰/۹۸۲	۰/۹۹۰	۰/۹۹۸	۱	۱/۰۱	۱/۰۱

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

ابتدا ساختار مولکول‌ها با نرم‌افزار هایپرکم رسم گردید. سپس با استفاده از روش نیمه تجربی PM3 ساختار سه بعدی ترکیبات بهینه شد. در این مرحله ۲۰ توصیف‌کننده برای هر مولکول به دست آمد. در ادامه ساختار بهینه‌شده‌ی مولکول‌ها به عنوان ورودی به نرم‌افزار دراگون داده شد و برای هر مولکول ۱۴۸۱ توصیف‌کننده‌ی دیگر به دست آمد. با توجه به این که ضریب فعالیت متوسط به دما وابسته است، لذا یک متغیر تجربی دما هم به توصیف‌کننده‌های به دست آمده اضافه گردید. به این ترتیب برای هر مولکول ۱۵۰۲ توصیف‌کننده به دست آمد. البته لازم به تذکر است که با توجه به رابطه‌ی (۱-۱) بهتر است به جای مقدار ضریب فعالیت در رقت بی‌نهایت از $\ln \gamma_{13}^{\infty}$ برای مدل‌سازی استفاده گردد، که در اکثر تحقیقات از این روش استفاده می‌شود [۲۹ و ۱۱].

۳-۱-۳- حذف توصیف‌کننده‌های نامناسب

تا این مرحله تعداد زیادی توصیف‌کننده به دست آمد که زیاد بودن تعداد توصیف‌کننده‌ها موجب طولانی و دشوار شدن فرایند مدل‌سازی می‌شود، پس باید توصیف‌کننده‌های نامناسب حذف شوند. روش مورد استفاده برای حذف توصیف‌کننده‌های نامناسب 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 صفر یا مقادیر یکسانی داشتند حذف شدند. سپس توصیف‌کننده‌هایی که برای بیش از ۹۰ درصد مولکول‌ها، مقدار یکسانی داشتند، نیز حذف شدند. در ادامه به وسیله برنامه‌ی نوشته شده در محیط MATLAB، توصیف‌کننده‌هایی که همبستگی بالای ۰/۹

داشتند، نیز حذف شدند. در پایان این مرحله ۲۷۱ توصیف‌کننده از کل توصیف‌کننده‌ها باقی ماندند که از میان این تعداد، توصیف‌کننده‌های Dipole moment، Heat of formation، Hydration Energy، total energy و softness با نرم‌افزار هایپرکم بدست آمدند.

۳-۱-۴- تقسیم‌بندی داده‌ها

پیش از شروع مرحله‌ی انتخاب توصیف‌کننده‌های مناسب، داده‌ها باید به گروه‌های آموزش، ارزیابی و تست تقسیم شوند که در این تحقیق هم از سری تست خارجی و هم تست داخلی استفاده گردید. بنابراین داده‌های مربوط به ۶ مولکول هپتان، سیکلوپنتان، ۱-هپتن، سیکلوهگزن، ۱-بوتانول و دی بوتیل اتر در شش دمای داده شده (۳۶ نقطه) در سری تست خارجی قرار گرفتند. سپس تعداد ۳۳۰ داده به طور تصادفی در سه سری آموزش، ارزیابی و تست داخلی دسته بندی شدند: ۲۳۰ نقطه در سری آموزش، ۵۰ نقطه در سری ارزیابی و ۵۰ نقطه در سری تست داخلی. داده‌های سری آموزش به منظور ساخت مدل‌های مختلف و داده‌های سری ارزیابی برای بهینه‌سازی پارامترهای مؤثر بر قدرت پیش‌بینی مدل‌ها استفاده می‌شوند. سپس از داده‌های سری تست داخلی برای ارزیابی ابتدایی و مقایسه توانایی مدل‌های مختلف و از داده‌های سری تست خارجی برای ارزیابی نهایی مدل انتخاب شده در پیش‌بینی ضریب فعالیت متوسط در رقت بی‌نهایت ترکیبات استفاده می‌شوند. این تقسیم‌بندی در جدول (پ-۱) در بخش پیوست آمده است.

۳-۱-۵- انتخاب بهترین توصیف‌کننده‌ها

تا این مرحله ۲۷۱ توصیف‌کننده وجود دارد که پس از تقسیم‌بندی داده‌ها باید از بین این تعداد، بهترین توصیف‌کننده‌ها برای ساخت مدل انتخاب شوند. در این تحقیق از روش

رگرسیون مرحله‌ای (SR) و روش اتحاد جداسازی داده و انتخاب خاصیت (CDFS)، برای انتخاب بهترین توصیف‌کننده‌ها استفاده شد.

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

برای انتخاب بهترین توصیف‌کننده‌ها با روش SR، از داده‌های سری آموزش توسط نرم افزار SPSS، رگرسیون مرحله‌ای گرفته شد. به این ترتیب که با انتخاب سربرگ Analyze در این نرم افزار، گزینه Regression و سپس گزینه‌ی Linear انتخاب و ۳۴ مدل مختلف به دست آمد. سپس ۱۴ مدل اول توسط سری ارزیابی مورد بررسی قرار گرفتند. میانگین مربعات خطا (MSE) و ضریب تعیین (R^2) و آماره‌ی F برای این مدل‌ها محاسبه شدند. مقادیر به دست آمده برای ۱۴ مدل در جدول (۲-۳) نشان داده شده است. با توجه به نتایج موجود در جدول، مدل ۱۲ که تعداد توصیف‌کننده‌های کمتر، مقدار ضریب تعیین و آماره‌ی F بیشتر و مقدار میانگین مربعات خطای کمتری برای سری ارزیابی داشت، به عنوان بهترین مدل انتخاب گردید.

برای ارزیابی اهمیت و سهم هر یک از توصیف‌کننده‌ها در مدل، باید اثر متوسط^۱ (ME) برای هر یک از توصیف‌کننده‌ها از رابطه‌ی زیر محاسبه شود.

$$ME_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j \beta_j \sum_{i=1}^n d_{ij}} \quad (1-3)$$

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

^۱-Mean Effect

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 (۴-۳) ارائه شده که عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

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

مدل	توصیف‌کننده	R^2	MSE	F
۱	nH	۰/۹۰۶۶	۰/۲۷۷۰	۱۶۱۴/۲۶۱
۲	nH , GATS1e	۰/۹۴۹۵	۰/۱۴۷۱	۲۳۹۷/۶۲۱
۳	nH , GATS1e, H050	۰/۹۶۹۴	۰/۰۸۸۹	۲۸۶۱/۳۷
۴	nH , GATS1e, H050, RDF020p	۰/۹۷۲۲	۰/۰۸۲۵	۲۳۷۵/۴۰۳
۵	nH , GATS1e, H050, RDF020p, ATS3p	۰/۹۸۲۶	۰/۰۵۰۹	۲۲۹۲/۳۶۶
۶	nH , GATS1e, H050, RDF020p, ATS3p, De	۰/۹۸۲۹	۰/۰۵۱۵	۲۶۸۳/۵۹۶
۷	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u	۰/۹۸۶۴	۰/۰۴۰۰	۳۲۲۳/۲۴۴
۸	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T	۰/۹۸۶۳	۰/۰۴۹۲	۲۴۱۵/۵۵۴
۹	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e	۰/۹۸۷۵	۰/۰۴۷۱	۲۳۹۸/۴۰۱
۱۰	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e, ATS3m,	۰/۹۸۹۷	۰/۰۴۹۴	۲۰۸۶/۴۸
۱۱	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e, ATS3m, H4m	۰/۹۹۰۶	۰/۰۵۶۹	۱۶۳۰/۱۳۷
*۱۲	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e, ATS3m, H4m, HATS0p	۰/۹۹۰۷	۰/۰۳۹۶	۲۱۳۲/۳۷۴
۱۳	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e, ATS3m, H4m, HATS0p, softness	۰/۹۹۱۰	۰/۰۴۳۳	۱۷۷۴/۱۹۴

۱۴	nH , GATS1e, H050, RDF020p, ATS3p, De, R4u, T, MATS1e, ATS3m, H4m, HATS0p, softness, RDF055u	۰/۹۹۰۹	۰/۰۴۹۶	۱۴۳۹/۴۶۶
----	--	--------	--------	----------

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	nH	Constitutional descriptor	number of Hydrogen atoms	۵۰/۹۴۳
۲	GATS1e	2D autocorrelations	Geary autocorrelation of lag 1 weighted by atomic Sanderson electronegativities	۳۰/۶۱۹
۳	H050	Atom-centred fragment	H attached to heteroatom	-۲۲/۶۵۶
۴	RDF020p	RDF descriptor	Radial Distribution Function -2.0 Weighted by atomic polarizabilities	-۲۱/۸۶۵
۵	ATS3p	2D autocorrelations	Broto –Moreau autocorrelation of a topological structure of lag 3 weighted by atomic polarizabilities	-۱۹/۳۵۳
۶	De	WHIM descriptors	D total accessibility index weighted by atomic Sanderson electronegativities	-۱۱/۳۵۶
۷	R4u	GETAWAY descriptors	R autocorrelation of lag 4 unweighted	-۸/۸۰۳
۸	T	Experimental	Temperature	-۶/۶۵۵
۹	MATS1e	GETAWAY descriptors	Moran autocorrelation - lag 1 Weighted by atomic Sanderson electronegativities	۶/۵۵۷
۱۰	ATS3m	2D autocorrelations	Broto-Moreau autocorrelation of a topological structure -lag3 weighted by atomic masses	۸/۳۶۶
۱۱	H4m	GETAWAY descriptors	H autocorrelation of lag 4 Weighted by atomic masses	-۶/۱۴۱

۱۲	HATS0p	GETAWAY descriptors	Leverage-weighted autocorrelation of lag 0 weighted by atomic polarizabilities	۴/۵۵۸
----	--------	---------------------	---	-------

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

	nH	GATS1e	H050	RDF020p	ATS3p	De	HATS3m	H2m	H3m	MATS3e	GATS4e	T
nH	۱											
GATS1e	۰/۶۰۵	۱										
H050	-۰/۳۶۷	۰/۰۳۹	۱									
RDF020p	۰/۵۸۰	۰/۳۳۸	-۰/۲۹۱	۱								
ATS3p	۰/۲۳۶	-۰/۰۰۵	۰/۶۰۶	۰/۱۳۰	۱							
De	۰/۴۸۹	۰/۲۶۹	-۰/۱۳۷	۰/۴۵۱	-۰/۳۶۲	۱						
R4u	۰/۱۴۲	-۰/۰۳۶	۰/۰۱۰	-۰/۱۰۹	۰/۲۴۴	-۰/۰۸۵	۱					
T	۰	۰	۰	۰	۰	۰	۰	۱				
MATS1e	۰/۱۴۴	-۰/۴۵۷	-۰/۶۷۹	۰/۱۶۲	۰/۵۱۲	-۰/۰۰۳	۰/۱۸۳	۰	۱			
ATS3m	-۰/۰۲۳	-۰/۳۷۰	-۰/۴۷۱	۰/۰۹۴	۰/۷۵۲	-۰/۳۰۵	۰/۱۵۴	۰	۰/۴۵۲	۱		
H4m	۰/۵۳۳	۰/۰۰۴	-۰/۲۷۰	۰/۲۸۷	۰/۰۸۸	۰/۳۲۵	۰/۱۰۳	۰	۰/۱۷۱	۰/۲۴۹	۱	
HATS0p	-۰/۷۴۲	-۰/۴۵۱	۰/۲۵۲	-۰/۴۶۸	-۰/۲۶۲	-۰/۳۸۱	-۰/۲۹۷	۰	-۰/۱۲۸	-۰/۱۶۲	-۰/۳۸۲	۱

۳-۱-۵-۲- انتخاب بهترین توصیف‌کننده‌ها به روش CDFS

در انتخاب توصیف‌کننده‌ها به روش رگرسیون مرحله‌ای، مدل نهایی به شدت تحت تأثیر دسته‌بندی داده‌ها است. روش CDFS نسبت به روش رگرسیون مرحله‌ای این مزیت را دارد که اثر دسته‌بندی بر مدل نهایی را کاهش می‌دهد. در این روش، ابتدا نقاطی که به عنوان گروه تست خارجی انتخاب شده بودند، کنار گذاشته شدند. سپس داده‌های باقی مانده با ۱۶ دسته‌بندی مختلف به دو دسته‌ی آموزش و ارزیابی تقسیم شدند. در مرحله‌ی بعد روش رگرسیون مرحله‌ای برای مدل‌سازی سری‌های آموزش هر یک از دسته‌بندی‌ها انجام و مدل بهینه بر اساس پارامترهای آماری سری ارزیابی مربوطه انتخاب گردید. در مجموع ۱۶ مدل بهینه به دست آمد که توصیف‌کننده‌هایی که در بیش از ۵۰٪ مدل‌ها تکرار شده بودند به عنوان توصیف‌کننده‌های نهایی انتخاب شدند. توصیف‌کننده‌های منتخب هر دسته‌بندی در جدول (۳-۵) آمده‌اند که با توجه به آن‌ها تعداد ۸ توصیف‌کننده انتخاب شدند. نشان، نام کامل و اثر متوسط این توصیف‌کننده‌ها در جدول (۳-۶) ارائه شده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۷) گزارش داده شده است. مقادیر درج شده در این جدول عدم همبستگی بین توصیف‌کننده‌های انتخابی را نشان می‌دهد.

جدول (۳-۵) - توصیف‌کننده‌های به‌دست آمده به روش SR در ۱۶ مدل بهینه برای ۱۶ دسته بندی مختلف

تعداد توصیف‌کننده	توصیف‌کننده‌ها	شماره
۱۱	nH, GATS1e, H050, RDF020u, ATS3p, De, Mor11e, Mor13p, Mor30m, Mor07p, T	۱
۱۲	nH, GATS1e, H050, RDF020p, Mor15v, Mor11e, RDF035m, Mor25p, LBW, Jhetp, T, H2m	۲
۱۱	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF055u, H047, T	۳
۱۲	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF035m, Mor25p, GATS4p, T, BELm1	۴
۱۳	nH, GATS1e, H050, HATS0m, GATS1m, HATS3m, E2u, Mor11e, Q2, GATS5e, T, HATS4p, H2u	۵
۱۲	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF055u, T, MATS4e, Mor25e, RDF035p	۶
۱۲	nH, GATS1e, H050, RDF020p, Mor15v, Mor11e, RDF035m, H4v, LBw, Jhetp, Mor25p, T	۷
۱۲	nH, GATS1e, H050, HATS0m, GATS1m, Mor16u, RDF040v, Mor13m, T, Mor20p, H4m, Mor23v	۸
۱۱	nH, GATS1e, H050, RDF020p, GATS3p, ATS4p, De, RDF055u, T, Jhetp, H2m	۹
۱۰	nH, GATS1e, H050, RDF020p, GATS3p, RDF055u, E2u, HATS1m, RDF040v, T	۱۰
۱۳	nH, GATS1e, H050, RDF020p, GATS3p, RDF055u, RDF040v, di, ZM1v, T, H2m, RDF040u, GATS5e	۱۱
۱۲	nH, GATS1e, H050, HATS0m, GATS1m, Mor11e, Q2, RDF040v, softness, RDF020u, Mor13e, T	۱۲
۱۲	nH, GATS1e, H050, RDF020p, Mor15v, Mor11e, RDF035m, Mor20p, LBw, Jhetp, T, H2m	۱۳
۱۲	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF055u, Mor25e, di, Mor28u, T	۱۴
۱۲	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF035m, Mor25p, ATS4p, T, H2m	۱۵
۱۳	nH, GATS1e, H050, RDF020p, GATS3p, Mor15v, Mor11e, RDF055u, Mor25e, Mor04p, T, ASP, H4m	۱۶

جدول (۳-۶) - توصیف‌کننده‌های مولکولی انتخاب‌شده با روش CDFS

شماره	نشان	گروه	نام کامل	اثر متوسط
۱	T	experimental	Temperature	-۵/۵۰۶
۲	nH	Constitutional descriptor	number of Hydrogen atoms	۵۱/۵۰۸
۳	GATS1e	2D autocorrelations	Geary autocorrelation of lag 1 weighted by atomic Sanderson electronegativities	۴۷/۶۷۶
۴	Mor11e	3D-MORSE descriptors	3D-MORSE-Signal 11 weighted by atomic Sanderson electronegativities	۹/۹۴۱
۵	RDF020p	RDF descriptor	Radial Distribution Function -2.0 Weighted by atomic polarizabilities	-۱۹/۷۵۶
۶	H050	Atom-centred fragment	H attached to heteroatom	-۳۳/۳۲۰
۷	Mor15v	3D-MORSE descriptors	3D-MORSE-Signal 15 weighted by atomic vanderwaals volume	-۱۷/۲۵۷
۸	GATS3p	2D autocorrelations	Geary autocorrelation of lag 3 weighted by atomic polarizabilities	-۱۰/۶۲۹

جدول (۳-۷) - ماتریس همبستگی توصیف‌کننده‌های انتخاب‌شده توسط روش CDFS

	nH	GATS1e	H050	RDF020p	Mor11e	GATS3p	Mor15v	T
nH	۱							
GATS1e	۰/۶۰۵	۱						
H050	-۰/۳۶۷	۰/۰۳۹	۱					
RDF020p	۰/۵۸۰	۰/۳۳۸	-۰/۲۹۱	۱				
Mor11e	-۰/۶۷۶	-۰/۴۷۱	-۰/۲۳۸	-۰/۱۹۱	۱			
GATS3p	۰/۳۲۱	۰/۰۷۴	-۰/۳۱۸	-۰/۰۷۱	-۰/۲۸۵	۱		
Mor15v	-۰/۲۷۶	-۰/۲۷۷	-۰/۰۶۰	-۰/۰۹۷	۰/۶۵۹	۰/۰۷۲	۱	
T	•	•	•	•	•	•	•	۱

۳-۱-۶- مدل سازی غیرخطی با استفاده از شبکه‌ی عصبی مصنوعی

یکی از راه‌های یافتن رابطه‌ی غیرخطی بین متغیرهای مستقل و متغیر وابسته، استفاده از شبکه‌ی عصبی مصنوعی برای مدل سازی است. مقدار عددی توصیف‌کننده‌های انتخاب شده توسط دو روش SR و CDFS به عنوان ورودی و مقادیر تجربی $\ln \gamma_{13}^{\infty}$ 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 لونیبرگ-مارکوات (trainlm) و تنظیم 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 است 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 شامل تعداد توصیف‌کننده‌های ورودی (K)، تعداد گره در لایه پنهان (n)، تعداد دور آموزش (Epoch) و پارامتر ممثوم (μ) برای دستیابی به بهترین نتیجه، باید بهینه شوند. تابع کارایی شبکه نیز میانگین مربع خطا (MSE) می‌باشد.

۳-۱-۶-۱- بهینه‌سازی شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های به‌دست آمده از

رگرسیون مرحله‌ای (SR-ANN)

برای ساخت مدل غیرخطی با استفاده از روش شبکه‌ی عصبی مصنوعی، ابتدا توصیف‌کننده‌های انتخاب شده با روش رگرسیون مرحله‌ای، به عنوان ورودی به شبکه عصبی مصنوعی داده می‌شوند. با توجه به این که شبکه، توصیف‌کننده‌ها را به‌صورت مرحله‌ای به مدل وارد می‌کند، لذا ترتیب

چیدمان توصیف‌کننده‌ها در ماتریس ورودی شبکه از اهمیت ویژه‌ای برخوردار است. در حقیقت توصیف‌کننده‌ها باید برحسب میزان اهمیت، در ماتریس ورودی مرتب شوند. یک روش مرتب‌سازی توصیف‌کننده‌ها، مرتب‌سازی برحسب میزان همبستگی با متغیر وابسته است که به آن رتبه‌بندی همبستگی^۱ (CR) گفته می‌شود. برای این منظور میزان همبستگی هر توصیف‌کننده با خاصیت مورد نظر یا همان متغیر وابسته سنجیده می‌شود و توصیف‌کننده‌ای که بیشترین مقدار همبستگی را با متغیر وابسته داشته باشد، در ستون اول ماتریس ورودی شبکه قرار می‌گیرد. سایر توصیف‌کننده‌ها نیز به همین ترتیب در ماتریس ورودی شبکه مرتب می‌شوند. در جدول (۳-۸) مقادیر همبستگی توصیف‌کننده‌های انتخابی به روش SR با متغیر وابسته نشان داده شده است.

پس از مرتب‌سازی توصیف‌کننده‌ها با روش ذکرشده، توصیف‌کننده‌ها به مدل وارد و عملیات مدل‌سازی آغاز می‌گردد. هر شبکه دارای یک لایه‌ی ورودی، یک لایه‌ی خروجی و یک لایه‌ی پنهان می‌باشد. تعداد گره‌های لایه‌ی ورودی برابر توصیف‌کننده‌های بهینه‌ی حاصل از روش رگرسیون مرحله‌ای یعنی تعداد ۱۲ توصیف‌کننده‌ی ذکرشده در جدول (۳-۳) و لایه‌ی خروجی نیز دارای یک نرون است که نشان‌دهنده‌ی $\ln \gamma_{13}^{\infty}$ متناظر با هر ترکیب می‌باشد.

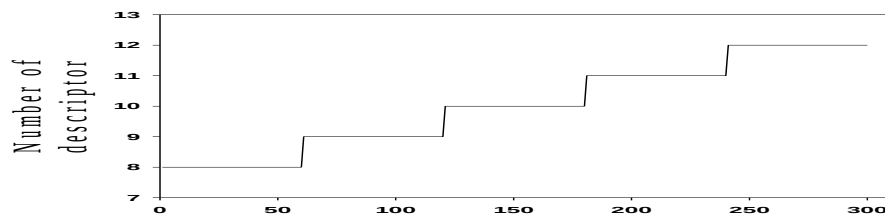
جدول (۳-۸)-مقادیر همبستگی توصیف‌کننده‌های انتخابی با روش SR با $\ln \gamma_{13}^{\infty}$

توصیف‌کننده	nH	GATS1e	H4m	ATS3m	RDF020p	De
همبستگی با مقدار $\ln \gamma_{13}^{\infty}$	۰/۹۲۸	۰/۷۷۶	۰/۳۶۶	-۰/۱۵۴	۰/۴۶۲	۰/۴۲۴
توصیف‌کننده	H050	MATS1e	R4u	ATS3p	T	HATS0p
همبستگی با مقدار $\ln \gamma_{13}^{\infty}$	-۰/۳۷۹	۰/۰۰۸	۰/۰۰۹	۰/۱۶۶	-۰/۰۳۲	-۰/۶۴۶

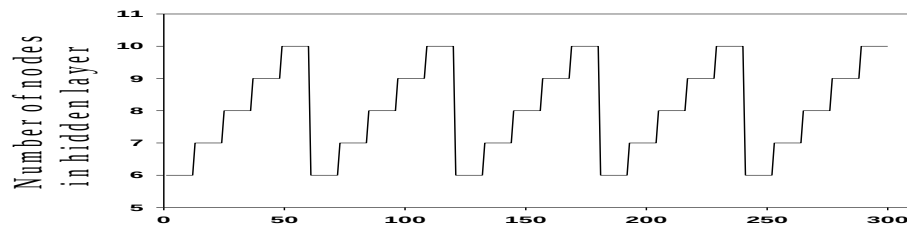
۱-Correlation Ranking

برای بهینه‌سازی تعداد ورودی‌های شبکه، تعداد گره‌های لایه‌ی پنهان و تعداد دور آموزشی از روش بهینه‌سازی همزمان استفاده گردید. برای این منظور، در هر یک از شبکه‌های طراحی شده، تعداد ورودی‌های شبکه از ۱ تا ۱۲ با گام ۱، تعداد گره‌ی لایه‌ی پنهان شبکه از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی شبکه از ۱۰ تا ۴۰۰ با گام ۱۰ تغییر داده شد و به ازای همه‌ی ترکیب‌های ممکن از این سه پارامتر، شبکه‌های 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. به دلیل این که بررسی تغییرات سه پارامتر فوق و مقدار MSE بر روی یک نمودار غیر ممکن است 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 برحسب 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.

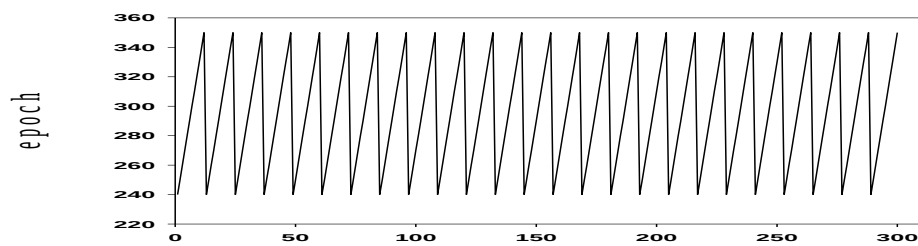
(الف)



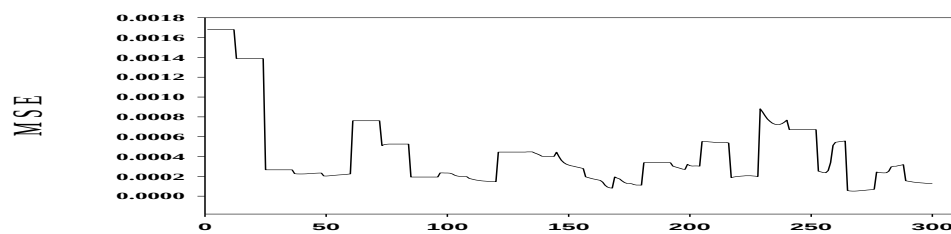
(ب)



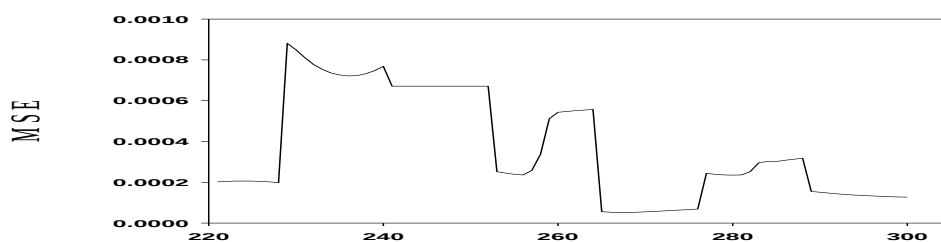
(ج)



(د)



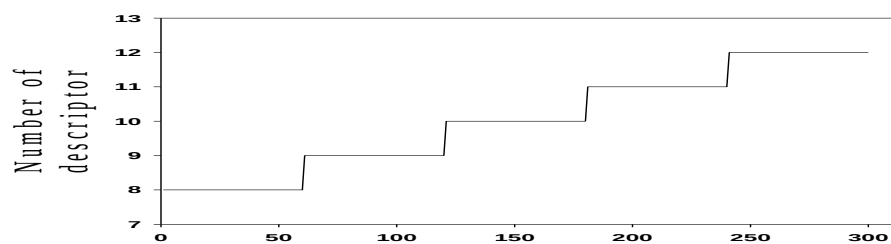
(ه)



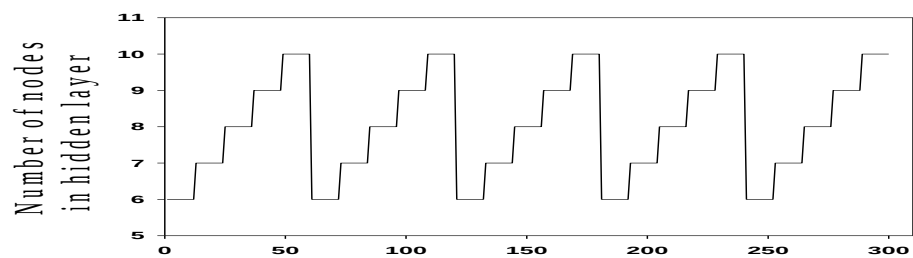
بردار مرجع

شکل (۱-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه 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

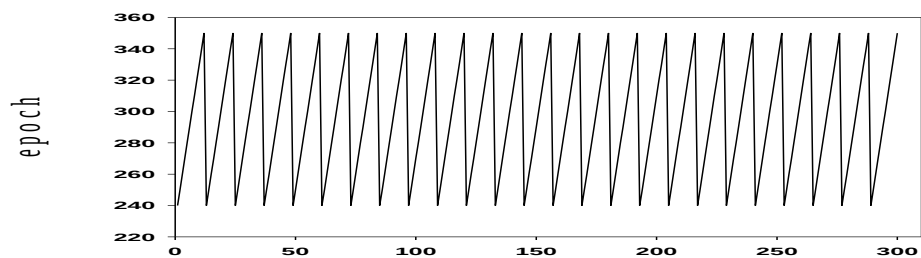
(الف)



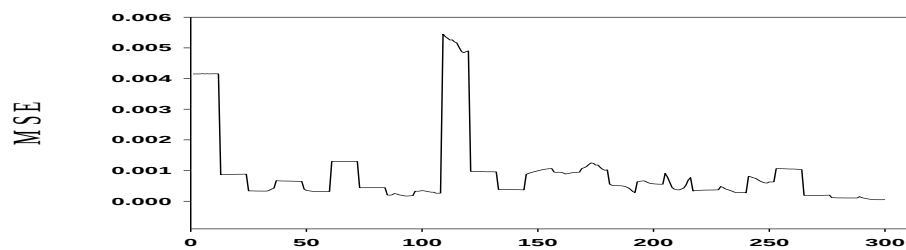
(ب)



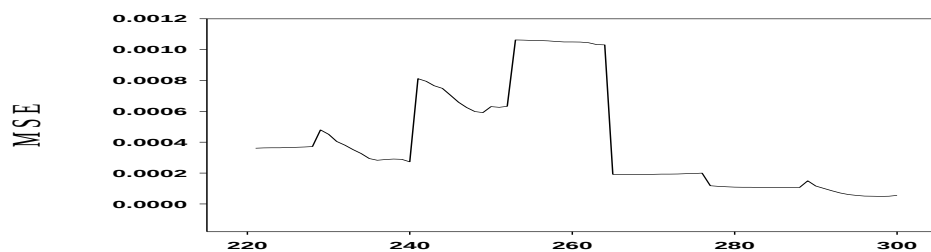
(ج)



(د)



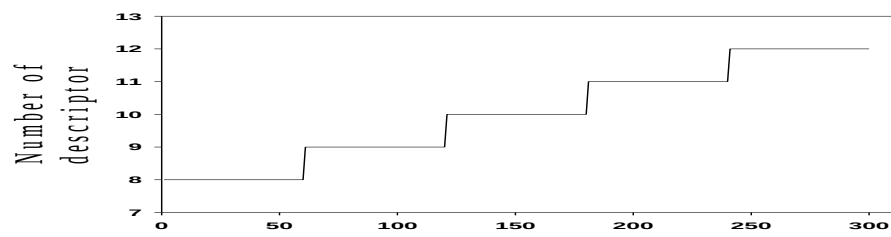
(ه)



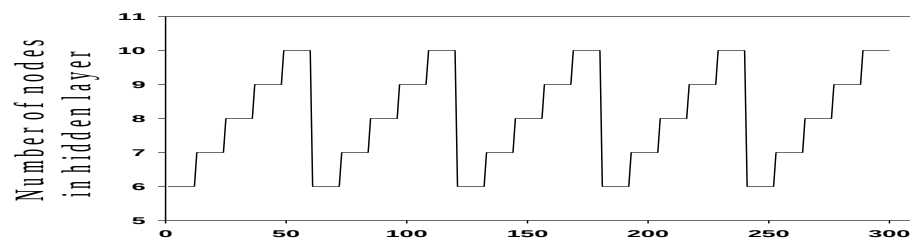
بردار مرجع

شکل (۲-۳) - 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

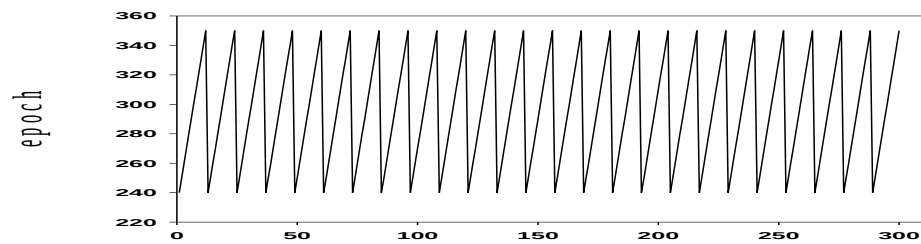
(الف)



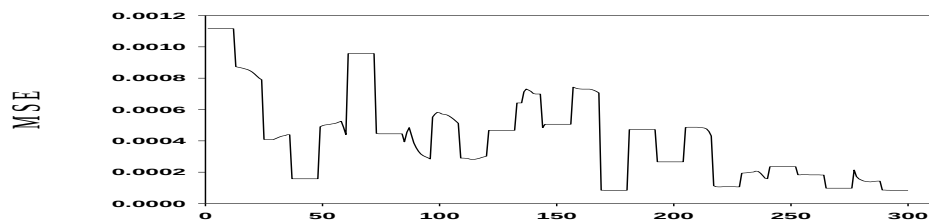
(ب)



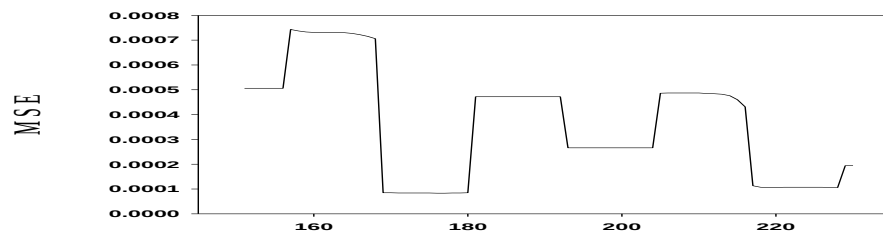
(ج)



(د)



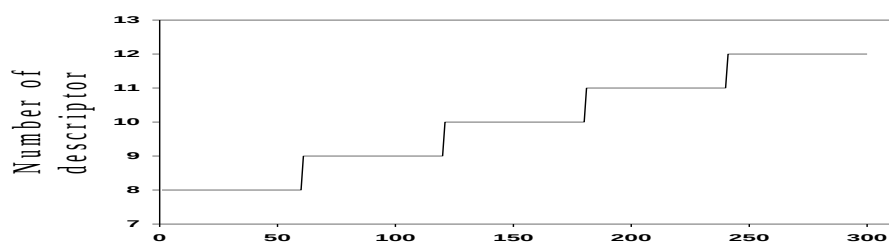
(ه)



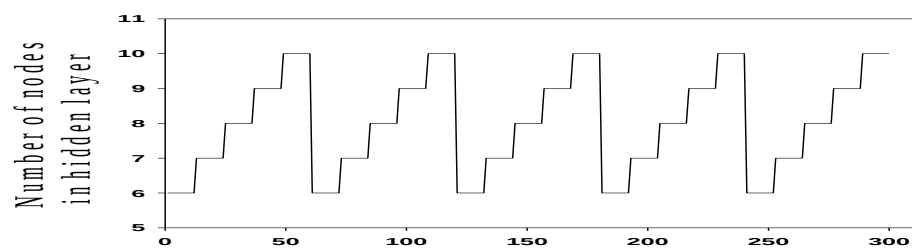
بردار مرجع

شکل (۳-۳) - 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 بایزین با استفاده از توصیف‌کننده‌های حاصل از SR

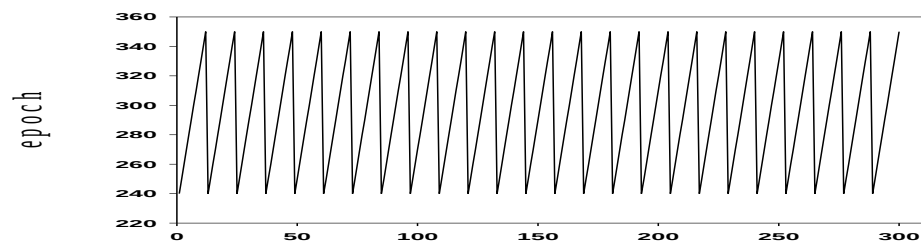
(الف)



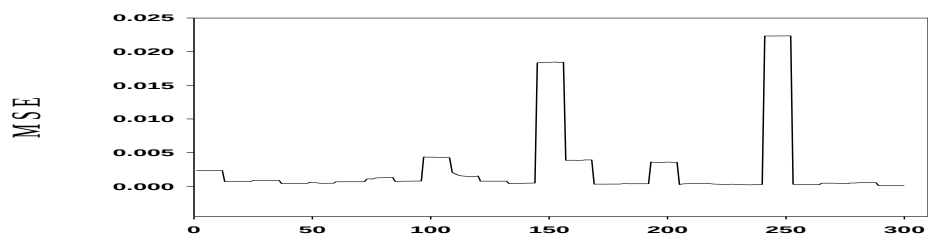
(ب)



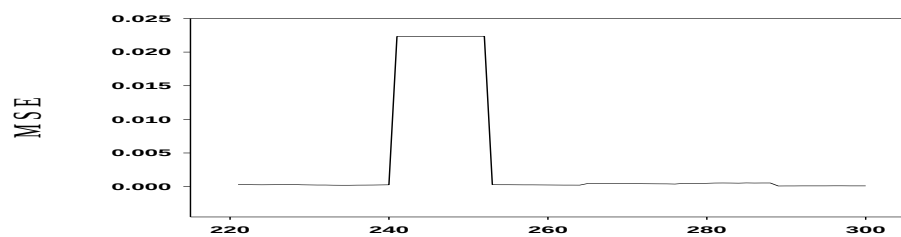
(ج)



(د)



(ه)



بردار مرجع

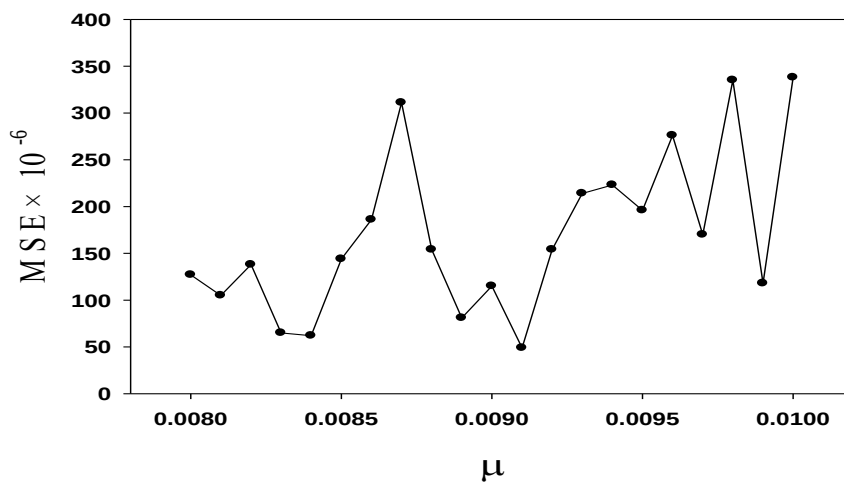
شکل (۳-۴) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده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

مقادیر بهینه‌ی به دست آمده برای هر یک از شبکه‌های طراحی شده در جدول (۹-۳) جمع‌آوری شده است. شبکه‌ای که تعداد توصیف‌کننده‌ی کمتر و میانگین مربع خطای کمتری دارد به عنوان شبکه‌ی بهینه انتخاب می‌شود. بنابراین طبق نتایج موجود در جدول (۹-۳) شبکه‌ای با تابع انتقال تانژانت سیگموئیدی و تابع آموزش تنظیم بایزین با ۱۰ توصیف‌کننده در لایه ورودی و ۱۰ گره در لایه پنهان و دور آموزشی ۳۱۰، به عنوان شبکه بهینه انتخاب گردید.

جدول (۹-۳) - توابع و پارامترهای بهینه‌ی شبکه با استفاده از توصیف‌کننده‌های حاصل از SR (SR-ANN)

تابع انتقال	تابع آموزش	تعداد توصیف‌کننده	تعداد نرون لایه پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۱۲	۱۰	۲۷۰	۰/۰۰۰۰۵۱
لگاریتم سیگموئید	لونبرگ-مارکوارت	۱۲	۱۰	۳۳۰	۰/۰۰۰۰۴۹
تانژانت سیگموئید	تنظیم بایزین	۱۰	۱۰	۳۱۰	۰/۰۰۰۰۸۳
تانژانت سیگموئید	لونبرگ-مارکوارت	۱۲	۱۰	۱۷۰	۰/۰۰۰۰۸۵

پس از مشخص شدن شبکه‌ی بهینه، باید پارامتر μ بهینه شود. برای این منظور در شبکه‌ی بهینه، تغییرات μ از ۰/۰۰۰۱ تا ۰/۱ با گام ۰/۰۰۰۱ اعمال شد و برای هر مقدار مومنتوم، مقدار میانگین مربعات خطا برای سری ارزیابی محاسبه گردید. نمودار تغییرات MSE برحسب μ حول نقطه‌ی بهینه در شکل (۵-۳) نشان داده شده است. مومنتوم ۰/۰۰۹۱ دارای کمترین مقدار میانگین مربعات خطا است و این مقدار به عنوان مقدار مومنتوم بهینه انتخاب شد.



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

توابع و مقادیر بهینه شده پارامترهای شبکه عصبی برای پیش بینی $\ln \gamma_{13}^{\infty}$ با استفاده از روش SR-ANN در جدول (۳-۱۰) ارائه شده است.

جدول (۳-۱۰)- توابع و پارامترهای بهینه شده ی شبکه ی عصبی (SR-ANN)

trainbr	الگوریتم آموزش
Tansig	تابع انتقال لایه ی پنهان
pureline	تابع انتقال لایه ی خروجی
۱۰	تعداد نرون لایه ی پنهان
۱۰	تعداد متغیرهای ورودی
۳۱۰	تعداد دوره های آموزش
۰/۰۰۹۱	پارامتر μ
۰/۰۰۰۰۴۹	مقدار MSE سری ارزیابی

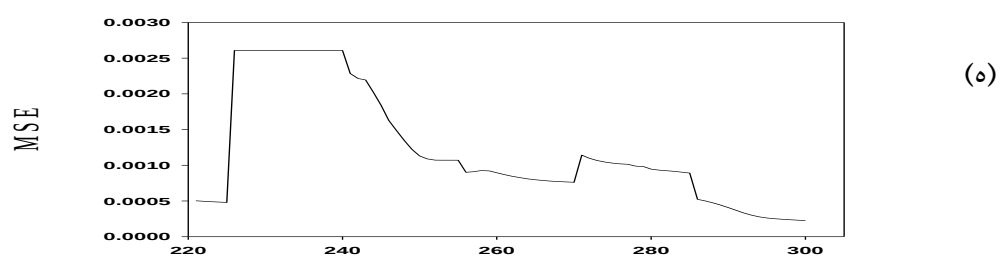
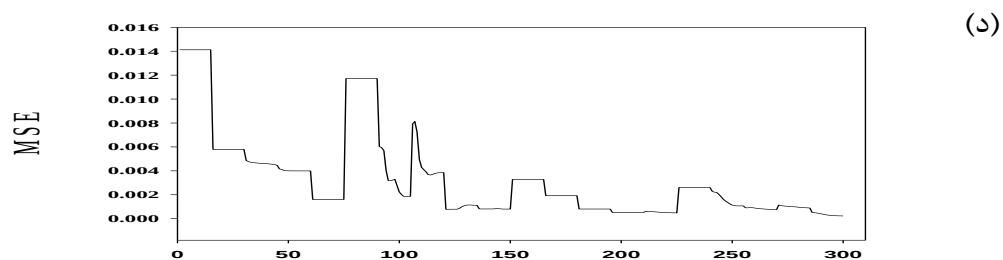
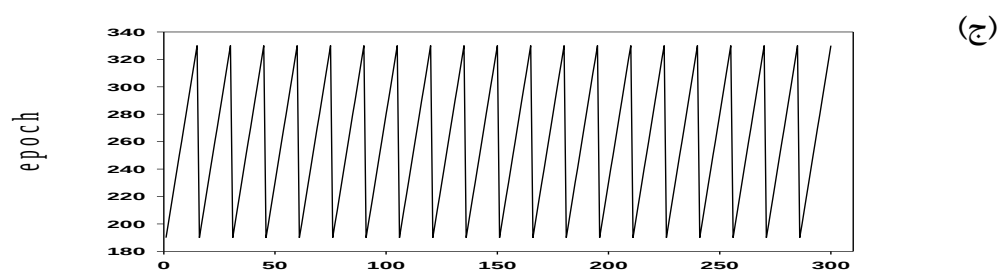
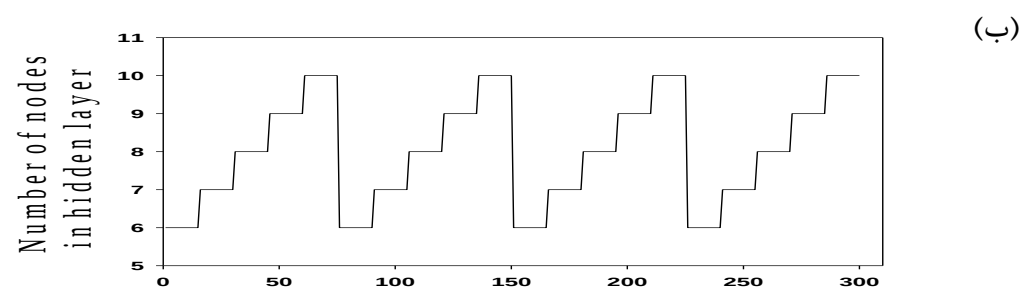
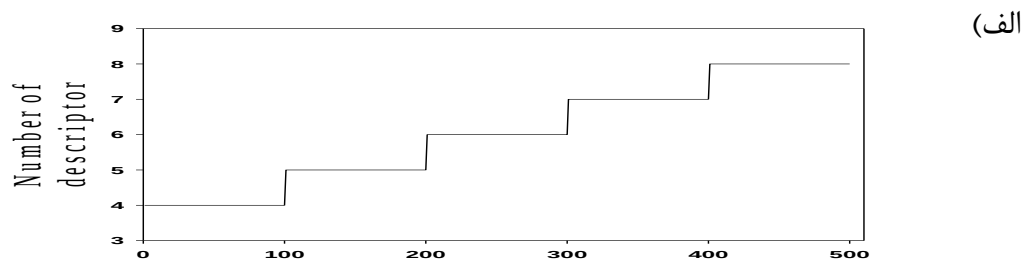
۳-۱-۶-۲- بهینه‌سازی شبکه عصبی مصنوعی با توصیف‌کننده‌های به‌دست آمده از روش CDFS

توصیف‌کننده‌های انتخاب شده‌ی نهایی با روش CDFS، برای مدل‌سازی غیرخطی به عنوان ورودی به شبکه‌ی عصبی مصنوعی ارائه شد. برای این منظور شبکه‌هایی با دو الگوریتم آموزشی لونبرگ-مارکوات و تنظیم بایزین، توابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید طراحی شدند. سپس در هر یک از شبکه‌های طراحی شده پارامترهای شبکه بهینه سازی گردیدند. برای این منظور شبکه‌هایی با مقدار μ ثابت و تعداد ورودی‌های از ۱ تا ۸ توصیف‌کننده با گام ۱، تعداد گره لایه‌ی پنهان از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی از ۱۰ تا ۴۰۰ با گام ۱۰ در نظر گرفته شدند و به ازای همه ترکیب-های ممکن از این سه پارامتر، چهار شبکه‌ی عصبی مصنوعی طراحی و بهینه سازی گردیدند. لازم به ذکر است که در اینجا نیز توصیف‌کننده‌های انتخاب شده توسط روش CDFS که در جدول (۳-۶) نمایش داده شده‌اند، به روش رتبه‌بندی همبستگی، CR، به شبکه عصبی وارد شدند. در جدول (۳-۱۱) مقادیر همبستگی توصیف‌کننده‌های انتخاب شده توسط این روش با متغیر وابسته نمایش داده شده است.

جدول (۳-۱۱)- مقادیر همبستگی توصیف‌کننده‌های انتخابی با روش CDFS با ضریب فعالیت در رقت بی‌نهایت

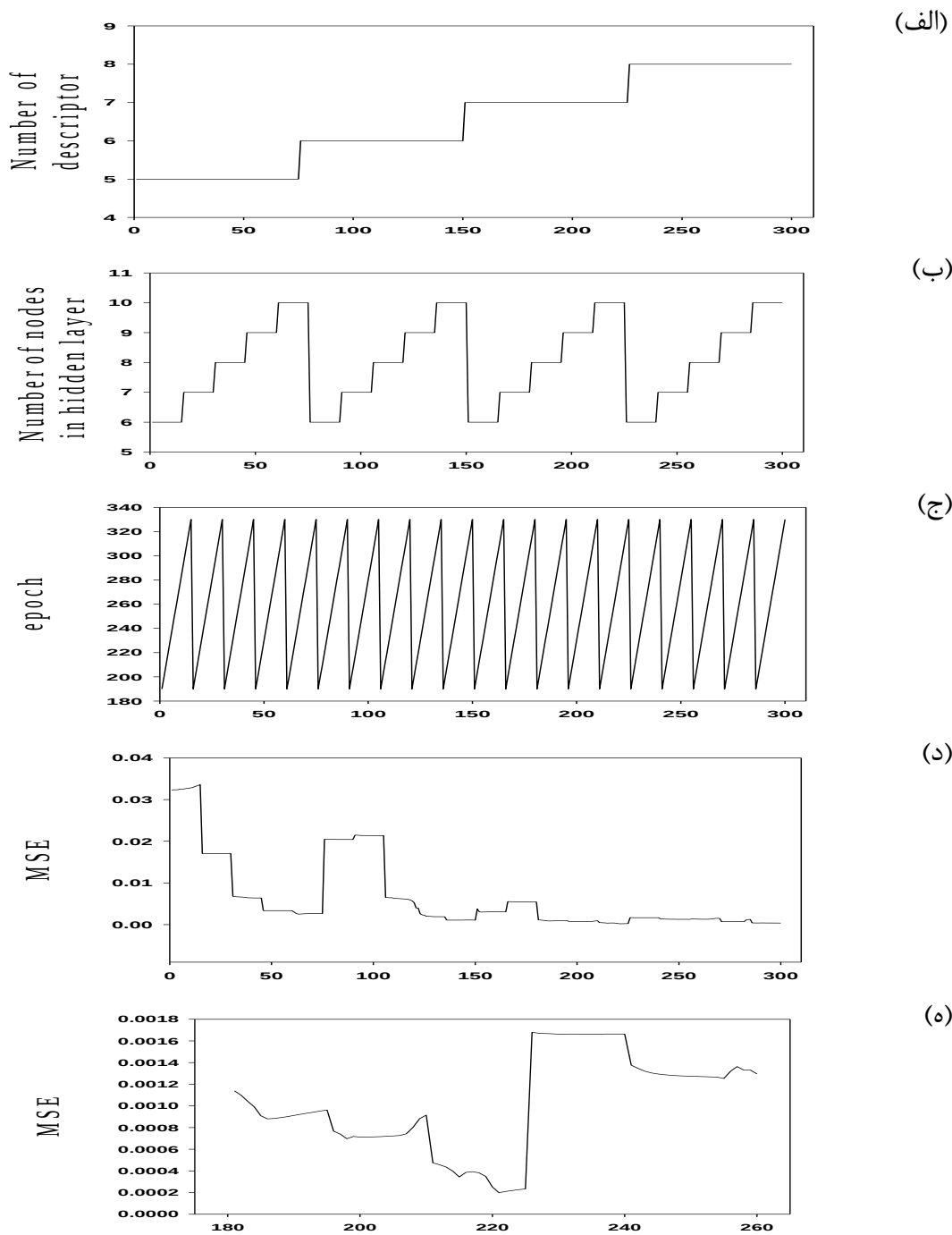
	nH	GATS1e	Mor11e	RDF020p	H050	GATS3p	Mor15v	T
همبستگی مقدار $\ln(\gamma_{13}^{\infty})$	۰/۹۲۸	۰/۷۷۶	۰/۶۷۹	۰/۴۶۲	۰/۳۷۸	۰/۳۲۰	۰/۳۵۲	۰/۲۰۱

در روند بهینه‌سازی در هر یک از شبکه‌های طراحی شده، 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 هر یک از سه کمیت تعداد توصیف‌کننده (K)، تعداد نرون لایه‌ی پنهان (n) و تعداد دور آموزش برحسب یک بردار مرجع در شکل‌های (۳-۶) تا (۳-۹) نشان داده شده است.



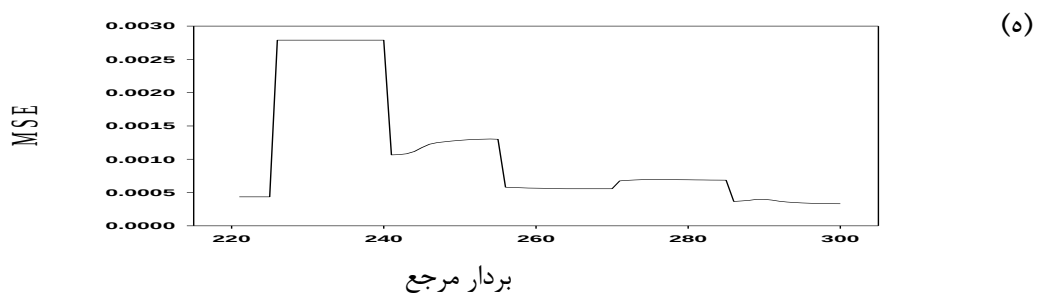
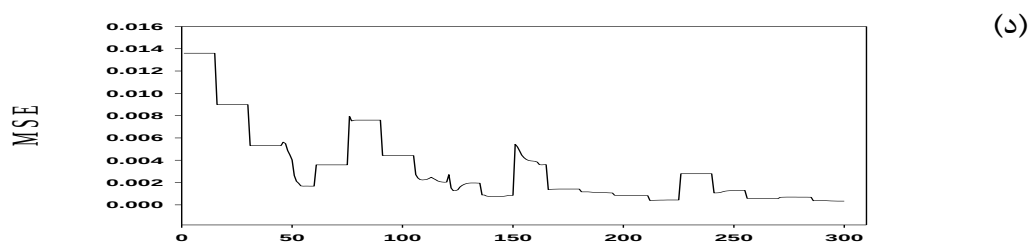
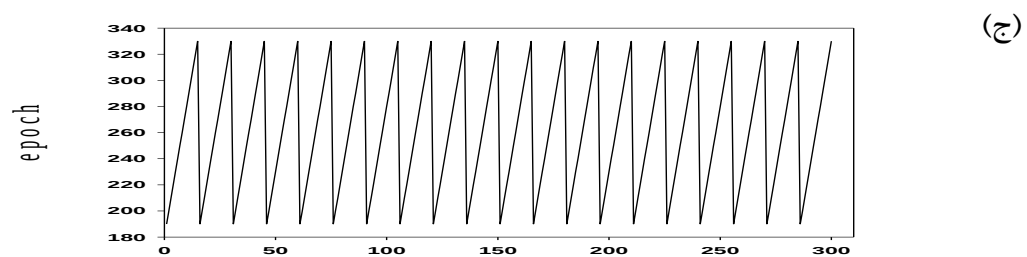
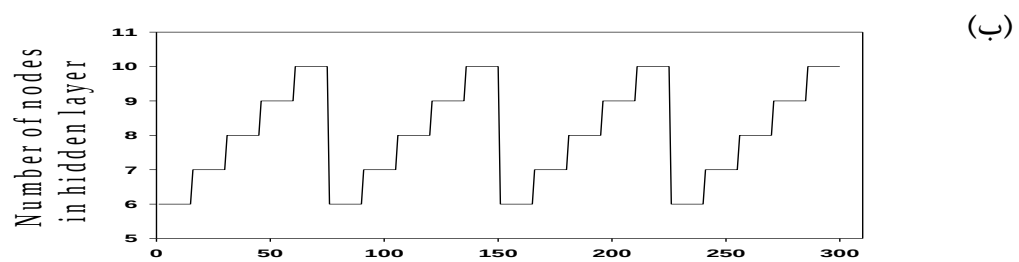
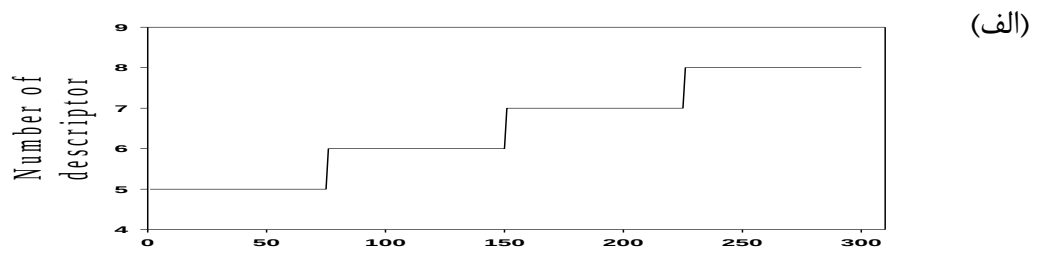
بردار مرجع

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 CDFS



بردار مرجع

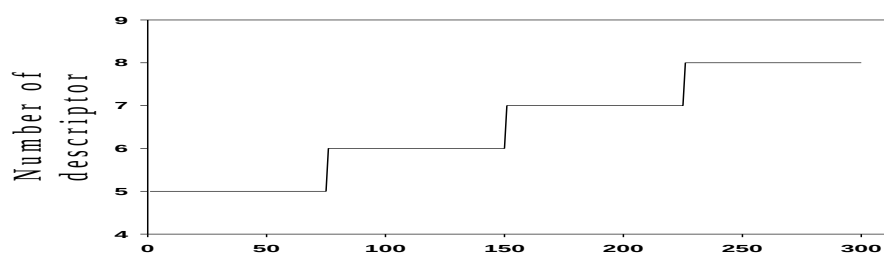
شکل (۷-۳) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوات با استفاده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



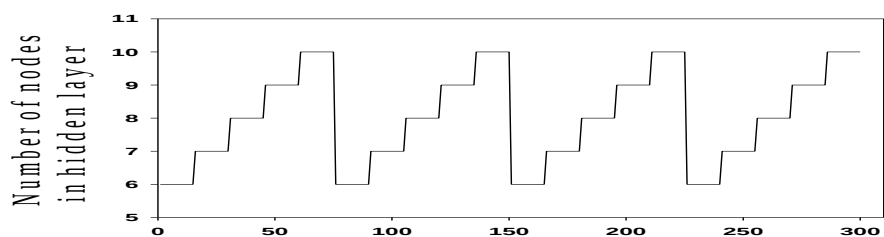
بردار مرجع

شکل (۳-۸) - نمودارهای (الف) تعداد توصیف‌کننده، (ب) تعداد نرون لایه 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 بایزین با استفاده از توصیف‌کننده‌های حاصل از CDFS

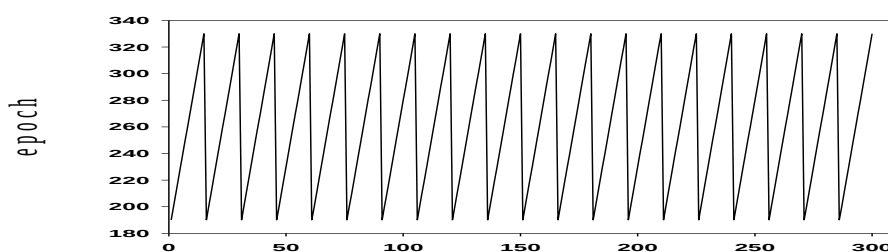
(الف)



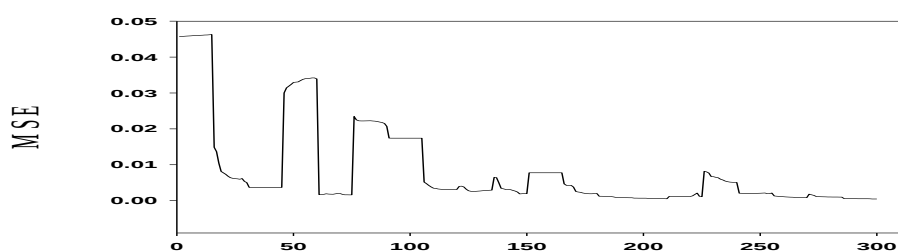
(ب)



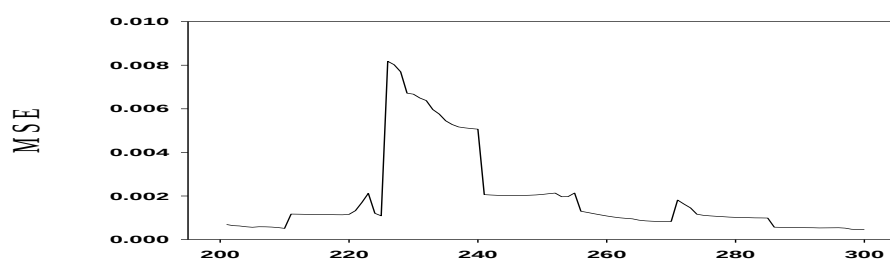
(ج)



(د)



(ه)



بردار مرجع

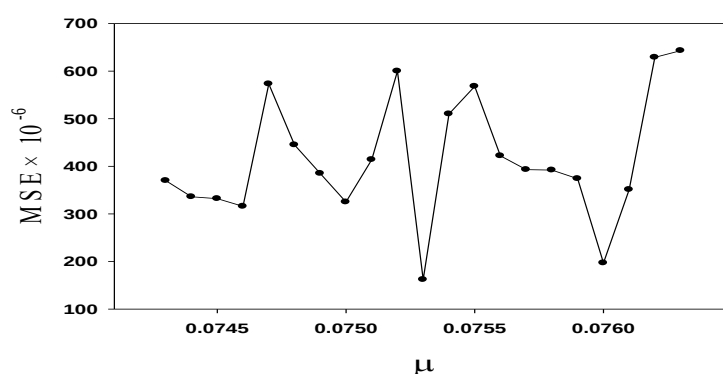
شکل (۳-۹) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، برحسب بردار مرجع، برای شبکه عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ - مارکوات با استفاده از توصیف کننده‌های حاصل از CDFS

مقدار و پارامترهای بهینه‌شده‌ی هر یک از شبکه‌های CDFS-ANN در جدول (۳-۱۲) گزارش شده است. نتایج موجود در این جدول نشان می‌دهد از میان چهار شبکه بهینه‌شده، شبکه بهینه‌ای که تعداد توصیف‌کننده کمتر و همچنین MSE کمتر داشته باشد به عنوان بهترین شبکه بهینه انتخاب می‌شود. به همین دلیل شبکه‌ای با الگوریتم آموزشی لونبرگ-مارکوات و تابع انتقال لگاریتم سیگموئیدی با ۷ توصیف‌کننده در لایه ورودی، ۱۰ گره در لایه پنهان و دور آموزشی بهینه‌ی ۲۹۰ کمترین میانگین مربع خطا را داشته و به عنوان شبکه‌ی بهینه انتخاب می‌شود.

جدول (۳-۱۲)-توابع و پارامترهای بهینه‌ی شبکه با استفاده از توصیف‌کننده‌های انتخابی به روش CDF (CDFS-ANN)

MSE	تعداد دور آموزش	تعداد نرون لایه‌ی پنهان	تعداد توصیف‌کننده	تابع آموزش	تابع انتقال
۰/۰۰۰۱۹۷	۴۰۰	۱۰	۸	تنظیم بایزین	لگاریتم سیگموئید
۰/۰۰۰۱۹۹	۲۹۰	۱۰	۷	لونبرگ-مارکوات	لگاریتم سیگموئید
۰/۰۰۰۳۰۵	۴۰۰	۱۰	۸	تنظیم بایزین	تانژانت سیگموئید
۰/۰۰۰۴۳۲	۴۰۰	۹	۷	لونبرگ-مارکوات	تانژانت سیگموئید

سپس باید مقدار پارامتر μ نیز بهینه گردد که بدین منظور، در ساختار شبکه‌ی بهینه با ۷ متغیر ورودی، ۱۰ گره در لایه‌ی پنهان، الگوریتم آموزشی لونبرگ-مارکوات، تابع انتقال لگاریتم سیگموئیدی و دور آموزشی ۲۹۰، مقدار μ از ۰/۰۰۰۱ تا ۰/۱ با گام‌های ۰/۰۰۰۱ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطا برای سری ارزیابی محاسبه و ثبت گردید که مقدار ۰/۰۷۵۳ به عنوان مقدار بهینه‌ی μ انتخاب شد. نمودار مقدار میانگین مربع خطا برحسب μ حول نقطه‌ی بهینه در شکل (۳-۱۰) رسم شده است.



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

با توجه به روند بهینه‌سازی ارائه شده، مقدار پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های به‌دست آمده از روش CDFS را در جدول (۳-۱۳) نمایش داده شده است.

جدول (۳-۱۳)- توابع و پارامترهای بهینه‌شده‌ی شبکه‌ی عصبی CDFS-ANN

trainlm	الگوریتم آموزش
Logsig	تابع انتقال لایه‌ی پنهان
pureline	تابع انتقال لایه‌ی خروجی
۱۰	تعداد نرون لایه‌ی پنهان
۷	تعداد متغیرهای ورودی
۲۹۰	تعداد دوره‌های آموزش
۰/۰۷۵۳	پارامتر μ
۰/۰۰۰۱۶۲	مقدار MSE سری ارزیابی

۳-۱-۶-۳- مدل سازی با استفاده از روش ماشین بردار پشتیبان (SVM)

روش غیرخطی دیگری که برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت مورد استفاده قرار گرفت، روش ماشین بردار پشتیبان است. برای انجام این روش، ابتدا داده‌های سری تست خارجی کنار گذاشته شدند و سپس باقیمانده‌ی داده‌ها به دو دسته‌ی آموزش و تست تقسیم گردیدند. داده‌های سری تست همان داده‌های سری تست داخلی در روش شبکه‌ی عصبی مصنوعی است. برای آموزش ماشین برداری پشتیبان باید پارامترهای آن شامل تابع حساسیت و پنایک، پارامتر موازنه‌ی C و پارامتر کرنل بهینه گردد. بنابراین تابع حساسیت و پنایک در بازه‌ی $0/00001$ تا $0/1$ ، پارامتر موازنه‌ی C در بازه‌ی 100 تا 1000000 و پارامتر کرنل در بازه‌ی $0/00001$ تا $0/1$ انتخاب گردید. لازم به ذکر است که در ماشین برداری پشتیبانی که در این تحقیق استفاده شد نوع کرنل، کرنل گوسین می‌باشد. در این تحقیق از کد برنامه‌ی جدیدی که در محیط MATLAB نوشته شد، برای بهینه کردن پارامترهای ذکرشده به طور همزمان استفاده شد. در پایان نامه‌های قبلی، از کد برنامه‌ای استفاده می‌شد که این سه پارامتر را به طور مجزا بهینه سازی می‌کرد.

بهینه‌سازی سه پارامتر ماشین برداری پشتیبان یک بار با استفاده از توصیف‌کننده‌های انتخاب شده به روش SR انجام گرفت که نتایج آن در جدول (۳-۱۴) گزارش شده است. بار دوم همین کار با استفاده از توصیف‌کننده‌های انتخاب شده به روش CDFS انجام گرفت که نتایج آن در جدول (۳-۱۵) ارائه شده است.

نتایج جدول (۳-۱۴) نشان می‌دهد که استفاده از کد برنامه جدید که سه پارامتر را به صورت همزمان بهینه می‌کند، نتایج به مراتب مطلوب‌تری می‌دهد. در نتیجه از مقادیر بهینه به دست آمده با این برنامه برای پیش‌بینی خاصیت مورد نظر استفاده گردید. به این ترتیب ماشین بردار پشتیبانی با مقادیر $0/05634$ برای تابع حساسیت و پنایک، 549649 برای پارامتر موازنه C و $0/00065$ برای پارامتر کرنل به عنوان ماشین بهینه برای پیش‌بینی ضریب فعالیت متوسط در رقت بی‌نهایت ترکیبات آلی و آب در مایع یونی [TCM][BMMOR] با استفاده از توصیف‌کننده‌های انتخاب‌شده به روش SR

می‌باشد.

جدول (۳-۱۴) - مقادیر پارامترهای بهینه و نتایج به دست آمده برای سری تست داخلی با استفاده از توصیف‌کننده‌های

انتخاب شده با روش SR

تابع حساسیت و پنیاک ε	با کد برنامه قبلی	با کد برنامه جدید
پارامتر موازنه C	۰/۰۰۵۲	۰/۰۵۶۳۴
پارامتر کرنل	۰/۰۰۱۵	۰/۰۰۰۶۵
MSE	۰/۰۳۰۴۴	۰/۰۲۲۲۰
R^2	۰/۹۸۲۰۱	۰/۹۸۶۱۶

جدول (۳-۱۵) - مقادیر پارامترهای بهینه و نتایج به دست آمده برای سری تست داخلی با استفاده از توصیف‌کننده‌های

انتخاب شده با روش CDFS

تابع حساسیت و پنیاک ε	با کد برنامه قبلی	با کد برنامه جدید
پارامتر موازنه C	۰/۰۰۹	۰/۰۳۲۷۸
پارامتر کرنل	۰/۰۰۰۸	۰/۰۰۰۳۵
MSE	۰/۰۳۹۵۴	۰/۰۳۱۸۹
R^2	۰/۹۷۸۰۳	۰/۹۸۰۲۷

نتایج جدول (۳-۱۵) نیز نشان می‌دهد که استفاده از کد برنامه‌ی جدید که سه پارامتر را به‌صورت همزمان بهینه می‌کند، نتایج به مراتب مطلوب‌تری می‌دهد. در نتیجه از مقادیر بهینه به‌دست آمده با این برنامه برای پیش‌بینی خاصیت موردنظر استفاده گردید. به این ترتیب ماشین بردار پشتیبانی با مقادیر ۰/۰۳۲۷۸ برای تابع حساسیت و پنیاک، ۰/۰۰۰۳۵ برای پارامتر موازنه C و ۰/۰۰۰۳۵ برای پارامتر کرنل به عنوان ماشین بهینه برای پیش‌بینی ضریب فعالیت متوسط در رقت بی‌نهایت ترکیبات آلی و آب در مایع یونی [TCM][BMMOR] با استفاده از توصیف‌کننده‌های انتخاب‌شده با

روش CDFS می‌باشد.

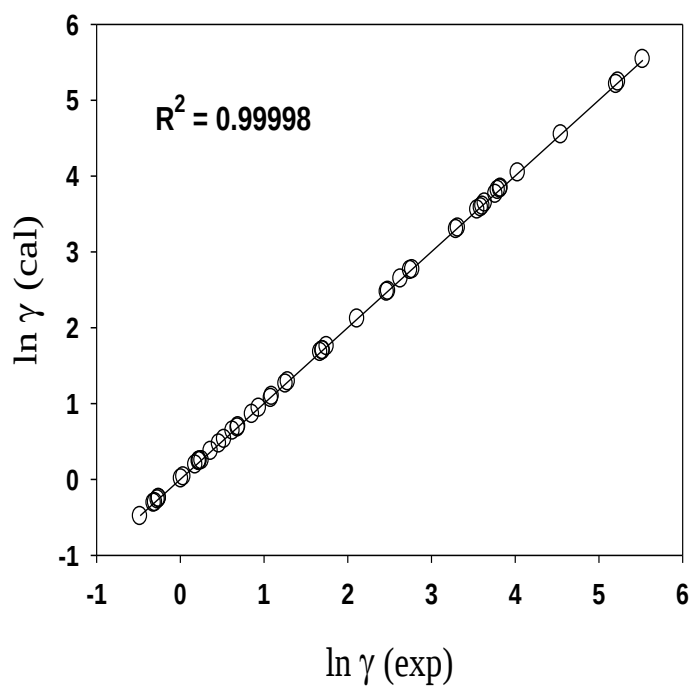
۳-۲- ارزیابی قدرت پیش‌بینی مدل‌های غیرخطی

۳-۲-۱- با استفاده از نمودار برگشتی

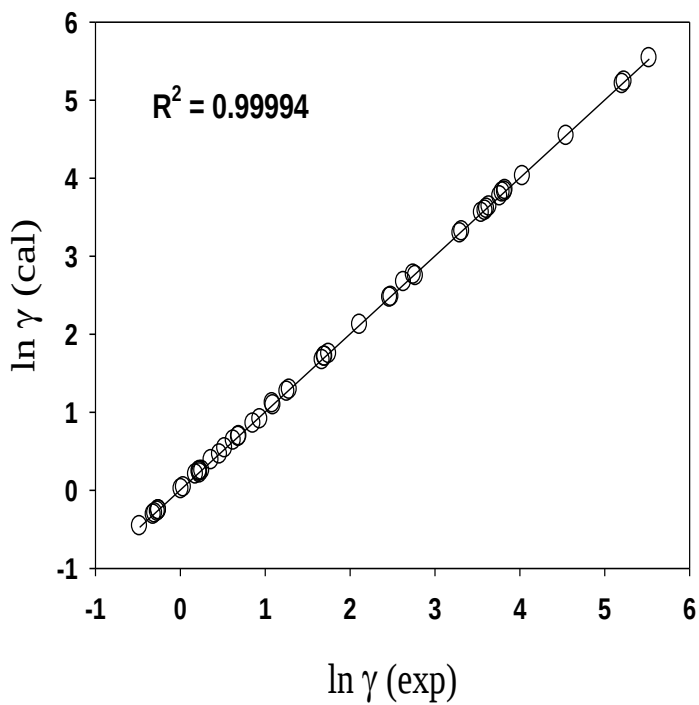
در نمودار برگشتی مقادیر پیش‌بینی شده برحسب مقادیر تجربی رسم می‌گردد و با توجه به مقدار ضریب تعیین (R^2) به دست آمده از نمودار، پراکندگی نقاط در اطراف خط برگشت تعیین می‌شود. هر چه مقدار ضریب تعیین به یک نزدیک‌تر باشد، مدل ساخته شده، مدل بهتری است. نتایج حاصل از نمودارهای برگشتی مربوط به چهار روش انجام شده در این کار برای سری‌های ارزیابی، تست داخلی و تست خارجی در شکل‌های (۳-۱۱) تا (۳-۱۵) نشان داده شده است. لازم به ذکر است که مقادیر پیش‌بینی شده γ_{13}^{∞} سری تست خارجی توسط چهار روش SR-ANN، CDFS-ANN، SR-SVM و CDFS-SVM به همراه مقادیر واقعی مربوطه در جدول (پ-۲) در بخش پیوست گزارش شده است. همان طور که در این شکل‌ها مشاهده می‌شود، در پیش‌بینی داده‌های سری تست داخلی دو روش SR-ANN و CDFS-ANN چندان تفاوتی ندارند و توافق بسیار خوبی بین داده‌های تجربی و محاسباتی وجود دارد اما برای داده‌های سری تست خارجی روش SR-ANN خیلی بهتر از مدل CDFS-ANN پیش‌بینی نموده است که این موضوع برتری روش SR-ANN نسبت به CDFS-ANN را اثبات می‌کند.

در مقایسه بین دو روش مدلسازی ANN و SVM مشاهده می‌شود که روش SVM در سری تست داخلی قدرت پیش‌بینی ضعیف‌تری نسبت به ANN دارد. اما در سری تست خارجی قدرت پیش‌بینی CDFS-ANN از روش‌های دیگر ضعیف‌تر است.

(الف)

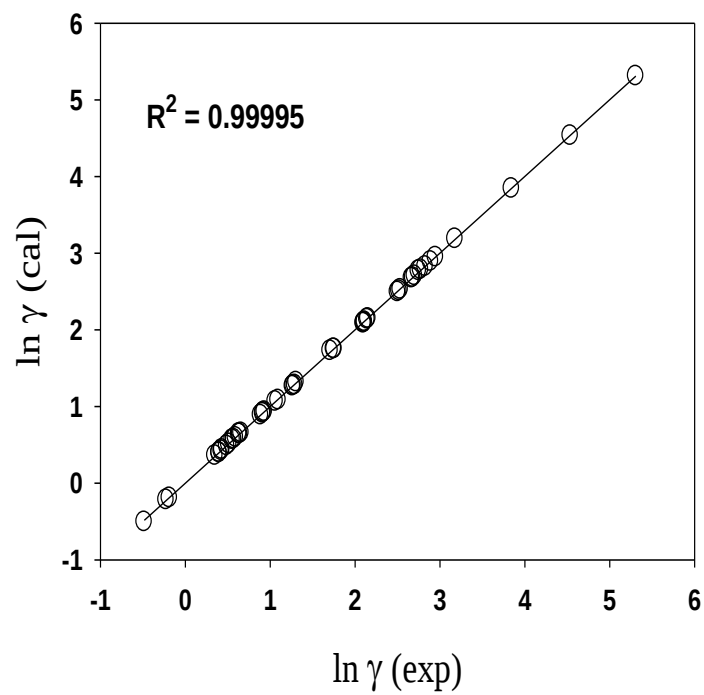


(ب)

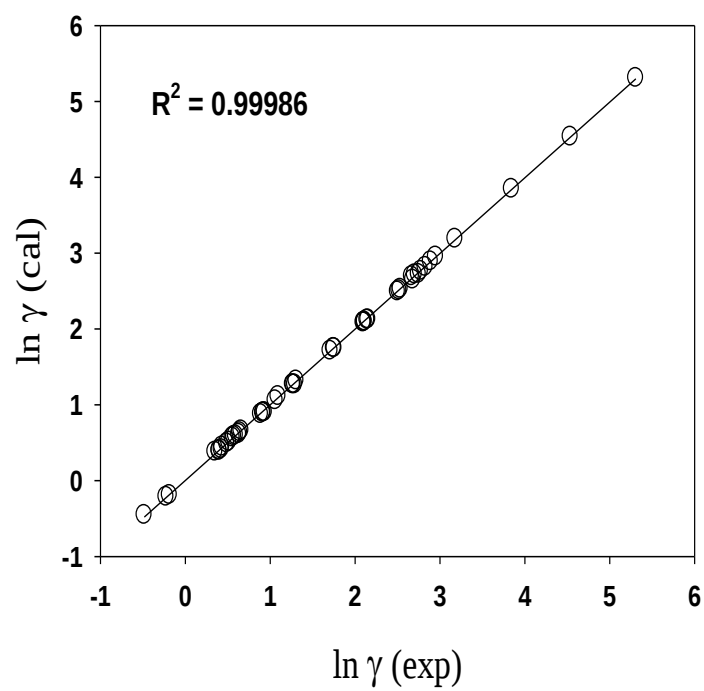


شکل (۳-۱۱) - نمودار مقادیر پیش‌بینی شده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN و (ب) مدل CDFS-ANN

(الف)

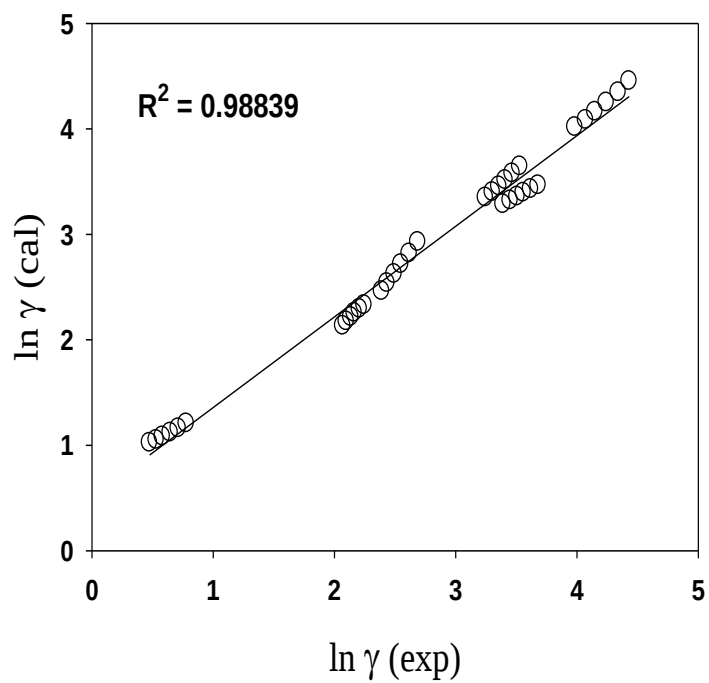


(ب)

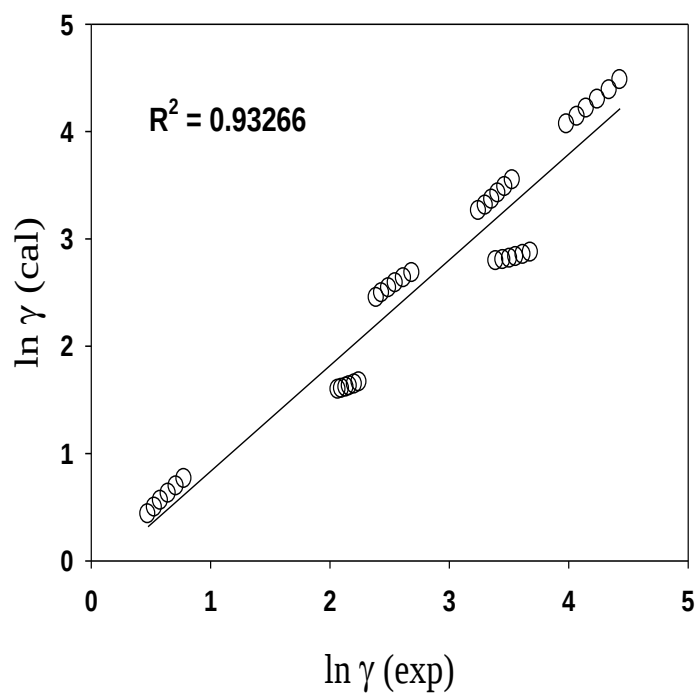


شکل (۳-۱۲) - نمودار مقادیر پیش‌بینی شده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری تست داخلی
(الف) مدل SR-ANN و (ب) مدل CDFS-ANN

(الف)

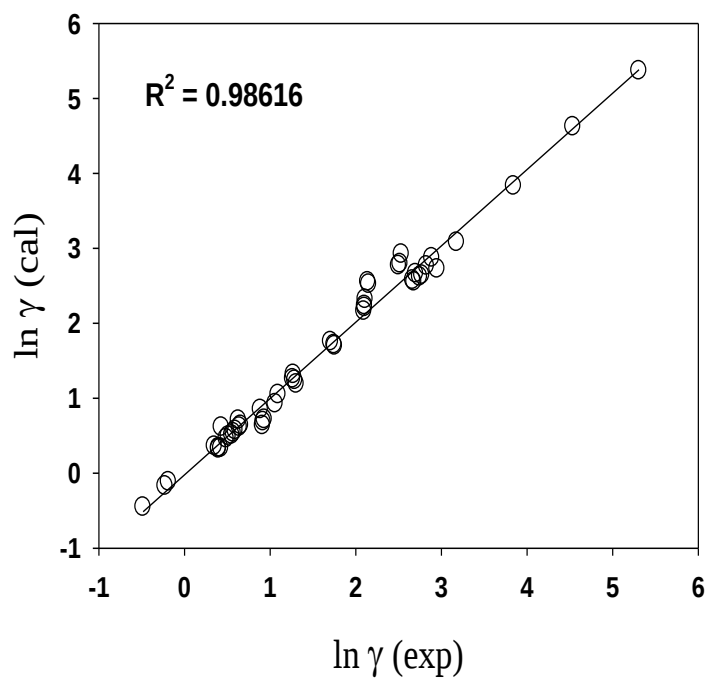


(ب)

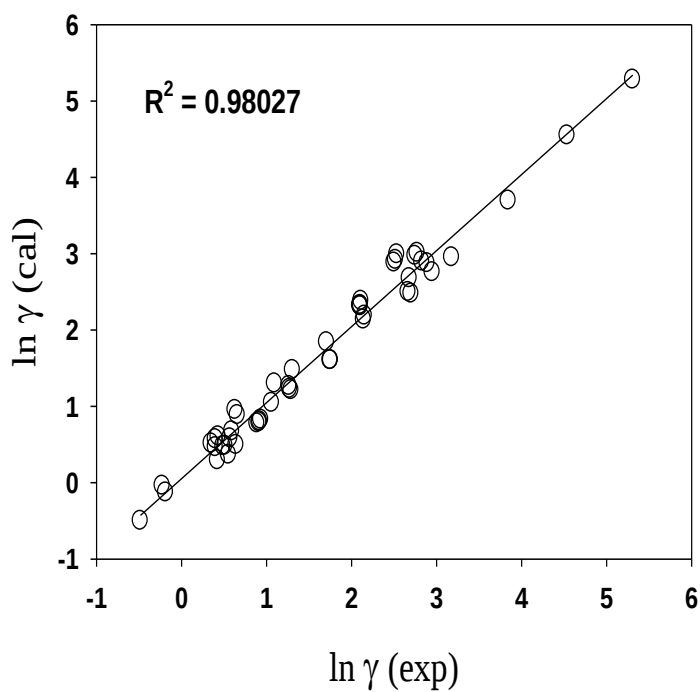


شکل (۳-۱۳) - نمودار مقادیر پیش‌بینی شده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری تست خارجی
(الف) مدل SR-ANN و (ب) مدل CDFS-ANN

(الف)

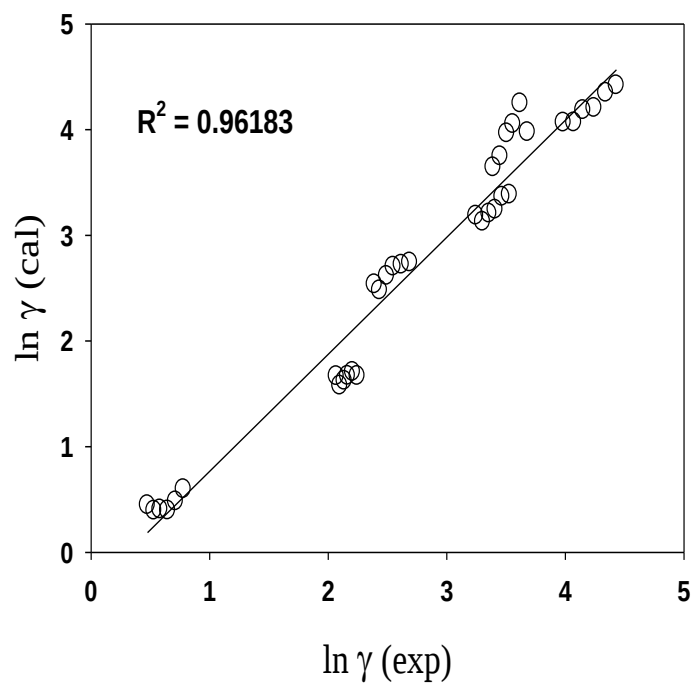


(ب)

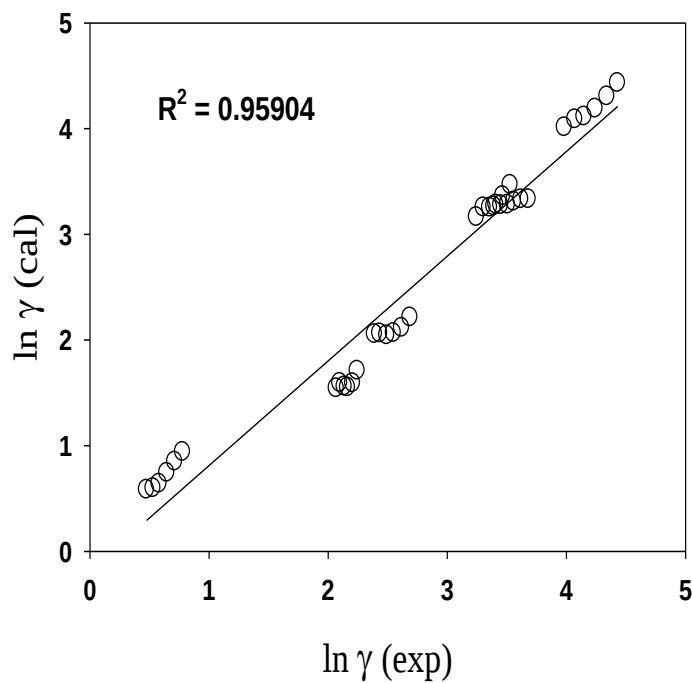


شکل (۳-۱۴)- نمودار مقادیر پیش‌بینی شده $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست داخلی
(الف) مدل SR-SVM و (ب) مدل CDFS-SVM

(الف)



(ب)



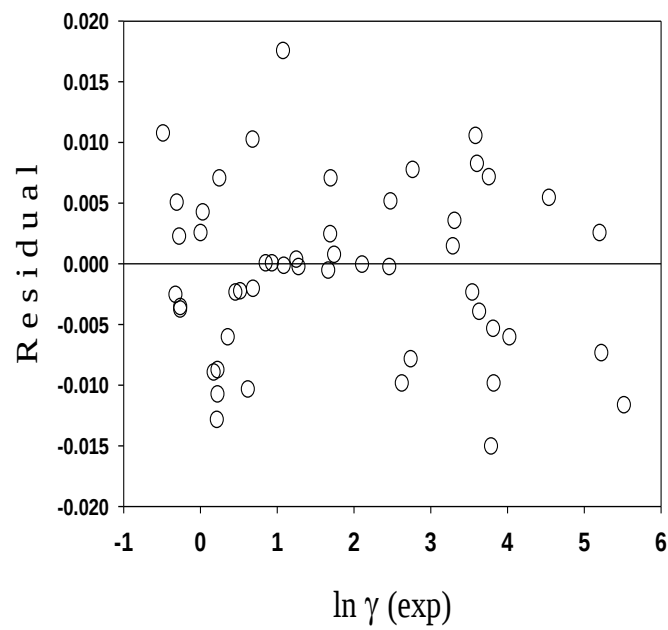
شکل (۳-۱۵)- نمودار مقادیر پیش‌بینی شده $\ln \gamma_{13}^{\infty}$ بر حسب مقادیر تجربی برای سری تست خارجی
(الف) مدل SR-SVM و (ب) مدل CDFS-SVM

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

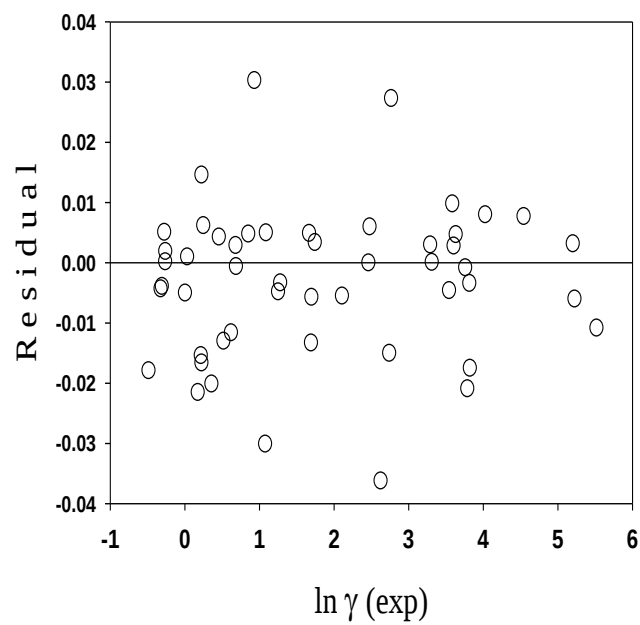
اختلاف مقادیر پیش‌بینی‌شده و مقادیر تجربی، خطای باقیمانده نامیده می‌شود. پراکندگی یکنواخت نقاط، حول محور افقی که بیانگر خطای باقیمانده‌ی صفر است، نشان‌دهنده‌ی آن است که خطای سیستماتیک در مدل‌سازی وجود ندارد. نمودار خطای باقیمانده برحسب مقادیر تجربی، برای مدل‌های ذکر شده در شکل‌های (۳-۱۶) تا (۳-۱۸)، نشان داده شده است.

مقایسه نمودارهای خطای باقیمانده نشان می‌دهد که برای سری ارزیابی و همچنین سری تست داخلی مدل SR-ANN مقدار خطای باقیمانده بین مقادیر تجربی و محاسباتی از روش‌های دیگر کمتر است. به همین دلیل روش ذکرشده برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت کارآمدتر می‌باشد.

(الف)

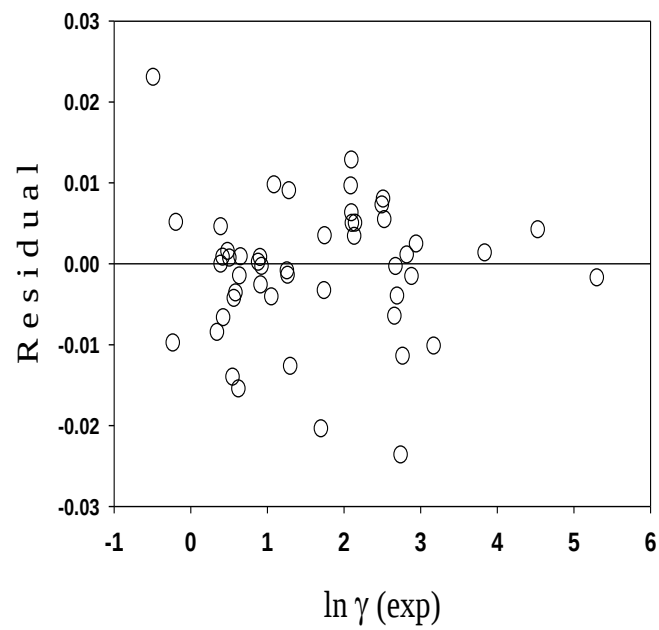


(ب)

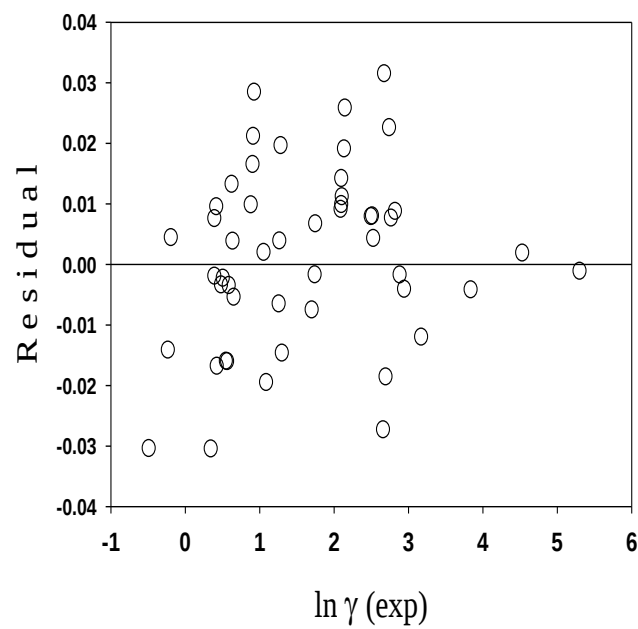


شکل (۳-۱۶) - نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری ارزیابی
(الف) مدل SR-ANN (ب) مدل CDFS-ANN

(الف)

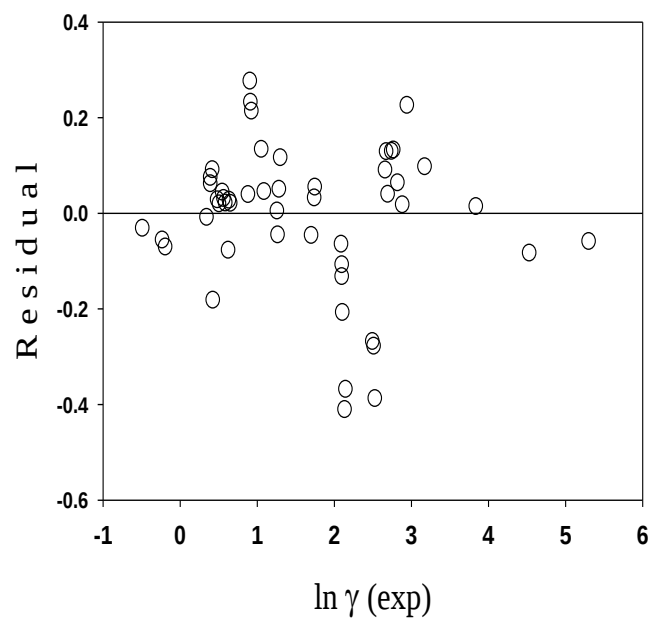


(ب)

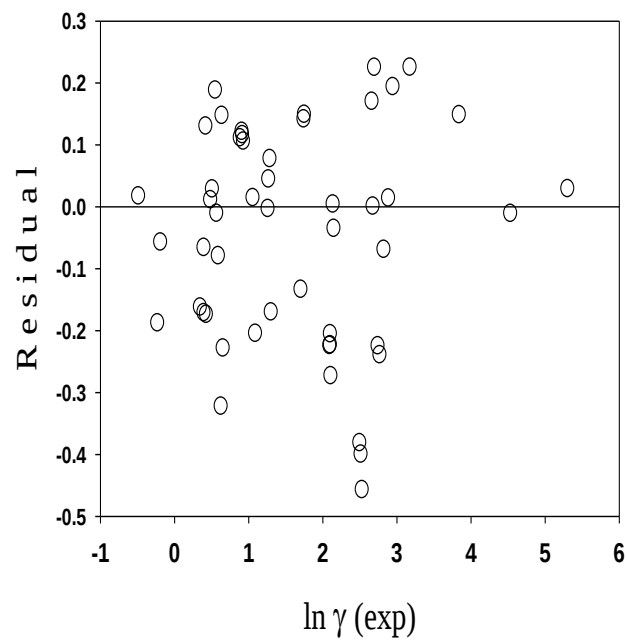


شکل (۳-۱۷) - نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری تست داخلی
(الف) مدل SR-ANN (ب) مدل CDFS-ANN

(الف)



(ب)



شکل (۳-۱۸) - نمودار مقادیر خطای باقیمانده $\ln \gamma_{13}^{\infty}$ برحسب مقادیر تجربی برای سری تست داخلی
(الف) مدل SR-SVM (ب) مدل CDFS-SVM

۳-۲-۳- ارزیابی مدل‌های غیرخطی بهینه شده با استفاده از پارامترهای آماری

همان‌طور که قبلاً بیان شد، پارامترهای آماری مقادیر تعیین‌کننده‌ای برای ارزیابی مدل‌ها هستند. لذا در این مرحله از این مقادیر برای ارزیابی مدل‌های ساخته شده با روش‌های غیرخطی SR-ANN، CDFS-ANN، SR-SVM و CDFS-ANN به‌کار گرفته شد که نتایج به‌دست آمده در جدول (۳-۱۶) نشان داده شده است.

با توجه به اینکه نتایج سری تست خارجی از اهمیت بیشتری برخوردار است، نتایج جدول نشان از برتری روش SR-ANN نسبت به سایر روش‌ها دارد.

جدول (۳-۱۶) - پارامترهای آماری روش‌های SR-ANN ، CDFS-ANN ، SR-SVM و CDFS-SVM

		SR-ANN	CDFS-ANN	SR-SVM	CDFS-SVM
R^2	سری ارزیابی	۰/۹۹۹۹۸	۰/۹۹۹۹۴	—	—
	سری تست داخلی	۰/۹۹۹۹۵	۰/۹۹۹۸۶	۰/۹۸۶۱۶	۰/۹۸۰۲۷
	سری تست خارجی	۰/۹۸۸۳۹	۰/۹۳۲۶۶	۰/۹۶۱۸۳	۰/۹۵۹۰۴
MSE	سری ارزیابی	۰/۰۰۰۰۴۹	۰/۰۰۰۱۶۲	—	—
	سری تست داخلی	۰/۰۰۰۰۶۹	۰/۰۰۰۲۱۳	۰/۰۲۲۲۰۳	۰/۰۳۱۸۹۷
	سری تست خارجی	۰/۰۵۱۵۵۳	۰/۱۳۲۲۹۶	۰/۰۸۳۶۵۱	۰/۰۹۸۸۱۱
PRESS	سری ارزیابی	۰/۰۰۲۴	۰/۰۰۸۰	—	—
	سری تست داخلی	۰/۰۰۳۴	۰/۰۱۰۶	۱/۱۱۰۱	۱/۵۹۴۸
	سری تست خارجی	۱/۸۵۵۸	۴/۷۶۲۶	۳/۰۱۱۴	۳/۵۵۷۱
MAE	سری ارزیابی	۰/۰۰۵۵	۰/۰۰۹۳	—	—
	سری تست داخلی	۰/۰۰۶۰	۰/۰۱۱۷	۰/۱۰۸۶	۰/۱۴۲۵
	سری تست خارجی	۰/۱۶۶۶	۰/۲۲۸۸	۰/۲۲۳۴	۰/۲۴۲۱
MPE	سری ارزیابی	۰/۳۳۳۷	-۱/۴۴۶۷	—	—
	سری تست داخلی	-۰/۲۴۷۳	۰/۱۴۷۵	۲/۸۷۳۸	-۱/۷۰۹۱
	سری تست خارجی	-۱۵/۰۷۲۷	۷/۶۴۱۰	۶/۲۲۶۳	۶/۰۱۰۹
$D_{\max}$	سری ارزیابی	۱۰/۷۶۹۳	۵۱/۰۰۰۰	—	—
	سری تست داخلی	۴/۷۶۰۹	۸/۷۰۴۸	۴۲/۲۸۵۴	۸۳/۲۱۲۹
	سری تست خارجی	۱۱/۵۲۳۳	۲۵/۹۴۷۶	۳۸/۷۷۳۴	۲۸/۳۰۵۲
MAPE	سری ارزیابی	۱/۵۶۴۴	۲/۴۷۳۲	—	—
	سری تست داخلی	۰/۶۹۲۹	۱/۳۵۵۷	۹/۴۶۱۱	۱۴/۵۰۴۵
	سری تست خارجی	۱۱/۹۲۱۱	۸/۷۳۷۱	۱۱/۵۹۵۴	۱۱/۶۲۹۸

۳-۳- بررسی ارتباط توصیف‌کننده‌های واردشده در مدل با ضریب فعالیت در رقت بی‌نهایت

در این قسمت به‌طور خلاصه ارتباط بین توصیف‌کننده‌های انتخاب شده در مدل SR-ANN و ضریب فعالیت در رقت بی‌نهایت مولکول‌ها مورد بررسی قرار خواهد گرفت. این توصیف‌کننده‌ها شامل: nH , T , $GATS1e$, $HATS0p$, $RDF020p$, $H-050$, De , $H4m$, $ATS3m$ و $ATS3p$ می‌باشد.

۳-۳-۱- توصیف‌کننده‌ی گروه Constitutional

این گروه از توصیف‌کننده‌ها ساده‌ترین و معمول‌ترین توصیف‌کننده‌های استفاده شده می‌باشند، که خصوصیات مولکول را بدون داشتن اطلاعات زیادی در مورد آن منعکس می‌کنند. تعداد اتم‌ها، تعداد پیوندها، تعداد مطلق و نسبی یک نوع اتم خاص در مولکول، تعداد پیوندهای ساده، دوگانه، سه‌گانه، و آروماتیک، تعداد حلقه‌ها، تعداد حلقه‌ها بر اساس تعداد اتم‌های آن‌ها و وزن مولکولی متداول‌ترین توصیف‌کننده‌های ساختاری می‌باشد.

یکی از این توصیف‌کننده‌ها، توصیف‌کننده‌ی nH می‌باشد که بیانگر تعداد اتم‌های هیدروژن می‌باشد. جدول (۳-۱۷) چگونگی ارتباط این توصیف‌کننده را با ضریب فعالیت در رقت بی‌نهایت برای چند مولکول از سری داده‌ها نشان می‌دهد.

جدول (۳-۱۷) - مثال‌هایی از مقدار توصیف‌کننده‌ی nH

	nH	γ^∞	$\ln \gamma^\infty$
Pentane	۱۰	۳۹/۷	۳/۶۸
Heptane	۱۶	۸۴	۴/۴۳۱
Decane	۲۲	۲۸۹	۵/۶۶

. افزایش مقدار این توصیف‌کننده به معنای افزایش تعداد اتم‌های هیدروژن، همچنین افزایش طول زنجیر مولکول شده که موجب کمتر شدن برهم‌کنش حلال و حل شونده می‌شود، در نتیجه ضریب فعالیت در رقت بی‌نهایت افزایش می‌یابد که مثبت بودن اثر متوسط توصیف‌کننده ذکر شده، این مطلب را تأیید می‌کند.

۳-۲-۳- توصیف‌کننده ی ACFC^۱

کدهای اجزای با مرکزیت اتمی (ACFC)، یک کد با مرکزیت اتمی با دامنه‌ی کوتاه است که هر اتم را با نوع اتم، انواع پیوند و انواع اتم‌های همسایه توصیف می‌کند. توصیف‌کننده ی H-050 جزء این دسته از توصیف‌کننده‌ها می‌باشد و نام کامل آن H attached to hetroatom است. در سری داده‌ها، فقط در الکل‌ها و آب، اتم هیدروژن به اتم اکسیژن متصل است و در سایر مولکول‌ها اتم‌های هیدروژن فقط به اتم کربن متصل می‌باشند. جدول (۳-۱۸) چگونگی ارتباط این توصیف‌کننده را با ضریب فعالیت در رقت بی‌نهایت برای چند مولکول از سری داده‌ها نشان می‌دهد. افزایش مقدار توصیف‌کننده موجب بیشتر شدن برهم‌کنش حلال و حل شونده می‌شود و به این ترتیب ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد. که منفی بودن اثر متوسط این توصیف‌کننده مطلب فوق را به‌درستی نشان می‌دهد.

جدول (۳-۱۸)- مثال‌هایی از مقدار توصیف‌کننده ی H-050

	H-050	γ^∞	$\ln \gamma^\infty$
Pentane	۰	۳۹/۷	۳/۶۸
Methanol	۱	۲/۱۸	۰/۷۷۹
Water	۲	۱/۰۳	۰/۰۳

^۱-Atom-Centred Fragment Code

۳-۳-۳- توصیف کننده های RDF^۱

این دسته از توصیف کننده ها، بر اساس فاصله ی توزیع در نمایش هندسی یک مولکول، شکل گرفته اند. تابع توزیع شعاعی مجموعه ای از اتم های A، به صورت احتمال توزیع اتم در اطراف کره ای به شعاع R تعریف می شوند. انواع این توصیف کننده به صورت RDF_{sw} مشخص می شوند که مقادیر S در محدوده ی ۱۰ تا ۱۵۵ با گام ۵ و w خاصیت اتمی (الکترون گاتیویته، جرم اتمی، قطبش پذیری اتمی و غیره) می باشد و به صورت زیر نمایش داده می شود:

$$RDF_{sw} = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A w_i \cdot w_j \cdot e^{-\beta(R-r_{ij})^2} \quad (۲-۳)$$

در این رابطه، f فاکتور اندازه گیری، w خواص اتمی اتم های i و j، r_{ij} فاصله ی بین اتم های i و j، A تعداد اتم ها و پارامتر تنظیم پذیر β به عنوان فاکتور دما که به جنبش اتم ها بستگی دارد، تعریف می شود.

RDF020p یکی از این توصیف کننده هاست که با قطبش پذیری اتمی وزن دار شده است. با توجه به رابطه ی فوق هر چه قطبش پذیری اتمی بیشتر شود، مقدار این توصیف کننده نیز افزایش می یابد. جدول (۳-۱۹) چگونگی ارتباط این توصیف کننده را با ضریب فعالیت در رقت بی نهایت برای چند مولکول از سری داده ها نشان می دهد. با افزایش این توصیف کننده برهمکنش حلال با حل شونده افزایش یافته و مقدار ضریب فعالیت کاهش می یابد، یا اثر متوسط این توصیف کننده منفی می باشد.

جدول (۳-۱۹)- مثال هایی از مقدار توصیف کننده ی RDF020p

	RDF020p	γ^∞	$\ln \gamma^\infty$
Benzene	۰/۳۵۳	۱/۵۳	۰/۴۲۵
Pyridine	۰/۴۳	۰/۷۳۳	-۰/۳۱۱

^۱-Radial Distribution Function

۳-۳-۴- توصیف‌کننده‌ی گروه 2D-Autocorrelation^۱

توصیف‌کننده‌های دو بعدی خود ارتباطی از گراف مولکولی از طریق محاسبه‌ی مجموع اوزان اتم‌های انتهایی کل مسیرها با طول مسیر مورد نظر، به دست می‌آیند. نوعی از این توصیف‌کننده‌ها گروه همبستگی گری^۲ در سال ۱۹۵۴ توسط گری^۳ پیشنهاد شدند که ضریب گری نام دارند و طبق رابطه‌ی (۳-۳) محاسبه می‌شوند.

$$C(d) = \frac{\frac{1}{2A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (\omega_i - \omega_j)^2}{\frac{1}{(A-1)} \cdot \sum_{i=1}^A (\omega_i - \bar{\omega})^2} \quad (3-3)$$

که ω یک ویژگی‌اتم، $\bar{\omega}$ میانگین مقدار آن خاصیت روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} ، دلتای کرونکر است (اگر $d_i = d_j$ باشد $\delta_{ij} = 1$ است و در غیر این صورت صفر است). Δ هم مجموع دلتاهای کرونکر است. مقدار این توصیف‌کننده‌ها از جنس فاصله است بنابراین از یک تا بی‌نهایت متغیر است. معنی عبارت ۱e در توصیف‌کننده‌ی GATS1e بیانگر فاصله‌ی توپولوژیکی ۱ و وزن دار شدن این توصیف‌کننده توسط الکترونگاتیوی اتمی ساندerson^۴ است. با توجه به این‌که علامت اثر متوسط این توصیف‌کننده مثبت می‌باشد، با ضریب فعالیت در رقت بی‌نهایت رابطه‌ی مستقیم دارد.

توصیف‌کننده‌های ATS3p و ATS3m نیز جزء این دسته از توصیف‌کننده‌ها می‌باشند که ATS3p دارای اثر متوسط منفی بوده که نشان می‌دهد با افزایش مقدار این توصیف‌کننده‌ها ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد و ATS3m دارای اثر متوسط مثبت است که نشان می‌دهد با افزایش مقدار این توصیف‌کننده‌ها ضریب فعالیت در رقت بی‌نهایت افزایش می‌یابد. جدول (۳-۲۰)

۱-Reciprocal distance wiener-type index

۲-Ceary Autocorrelation

۳-Ceary

۴-Atomic Sanderson electronegativities

چگونگی ارتباط توصیف کننده ATS3p را با ضریب فعالیت در رقت بی‌نهایت برای چند مولکول از سری داده‌ها نشان می‌دهد.

جدول (۳-۲۰) - مثال‌هایی از مقدار توصیف کننده ی ATS3p

	ATS3p	γ^∞	$\ln \gamma^\infty$
Butan-1-ol	۰/۲۶۳	۲/۱۸	۰/۷۷۹
Diethyl ether	۰/۲۶۱	۵/۳۹	۱/۶۸۵

۳-۵-۳- توصیف کننده‌های GETAWAY^۱

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

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

که M ماتریس مختصات اتمی، T به معنای ماتریس ترانپوز 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.

یک سری از این توصیف کننده‌ها، توصیف کننده‌های H-GETAWAY هستند که به اندازه و شکل مولکول وابسته بوده به گونه‌ای که با افزایش اندازه اتم و فاصله ی اتم از مرکز مولکول، مقدار آنها

۱-Geometry, Topology and Atom-Weights Assembly

۵- Leverage Matrix

۱- Molecular Infliuence Matrix

افزایش می‌یابد.

توصیف کننده‌های H4m و HATS0p جزء این دسته از توصیف کننده‌ها می‌باشند که H4m دارای اثر متوسط منفی بوده که نشان می‌دهد با افزایش مقدار این توصیف کننده ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد. HATS0p دارای اثر متوسط مثبت می‌باشند و با افزایش مقدار این توصیف کننده‌ها مقدار ضریب فعالیت در رقت بی‌نهایت نیز افزایش می‌یابد.

۳-۳-۶-توصیف کننده‌های WHIM^۱

این شاخص از مختصات کارتزین سه بعدی مولکول، با استفاده از صورتبندی با حداقل انرژی محاسبه می‌شود و شامل اطلاعاتی درباره اندازه، شکل، تقارن و توزیع اتمی ساختار سه بعدی مولکول می‌باشد. این توصیف کننده‌ها از رابطه‌ی (۳-۵) به دست می‌آیند.

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (۳-۵)$$

که S_{jk} کوواریانس وزن دار شده بین کئوردینه j ام و k ام، A تعداد اتم‌ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب j امین و k امین کئوردینه‌های اتم i ام و $\bar{q}$ مقدار میانگین متناظر است.

توصیف کننده De جزء این دسته از توصیف کننده‌هاست که دارای اثر متوسط منفی بوده و نشان می‌دهد با افزایش مقدار این توصیف کننده، ضریب فعالیت در رقت بی‌نهایت کاهش می‌یابد.

۳-۴-بررسی میزان مشارکت توصیف کننده‌های منتخب شبکه عصبی

میزان مشارکت توصیف کننده‌های انتخاب شده به صورت زیر تعیین می‌شود:

۱. ابتدا وزن مربوط به توصیف کننده‌ی موردنظر از شبکه بهینه حذف می‌شود.

۱- Weighted Holistic Invariant Molecular descriptors

۲. مقدار متغیر وابسته با استفاده از توصیف‌کننده‌های باقی‌مانده برای هر یک از ترکیبات سری

ارزیابی پیش‌بینی می‌شود.

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

۴. مراحل ۱ تا ۳ برای همه‌ی توصیف‌کننده‌ها تکرار می‌شود.

۵. در نهایت درصد مشارکت هر توصیف‌کننده توسط رابطه‌ی (۳-۶) محاسبه گردید.

$$C_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (۳-۶)$$

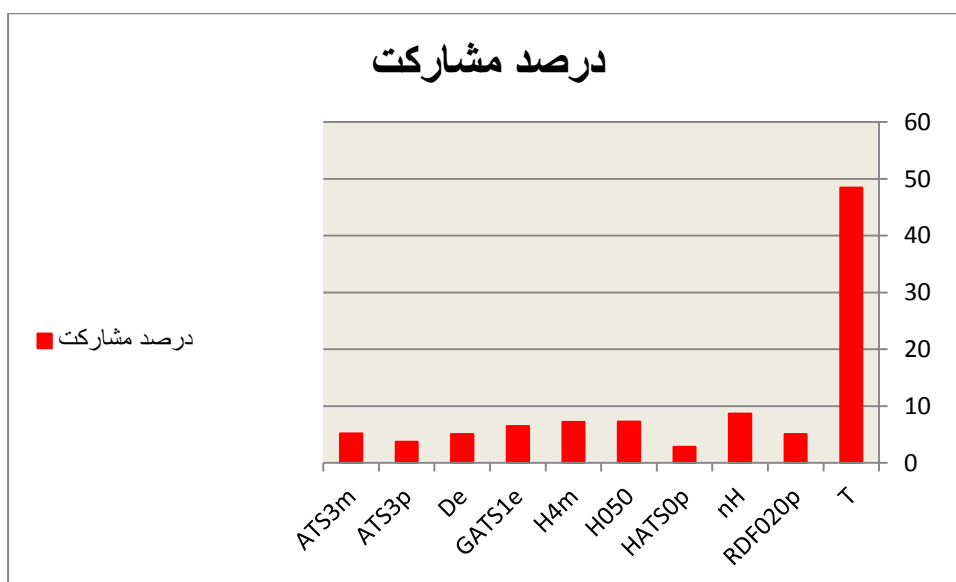
در این رابطه C_i درصد مشارکت توصیف‌کننده‌ی حذف شده i ام، N تعداد توصیف‌کننده‌ها در

مدل و Δm_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده‌ی i ام را نشان

می‌دهد. بر این اساس میزان مشارکت توصیف‌کننده‌های منتخب در پیش‌بینی ضریب فعالیت در رقت

بی‌نهایت ترکیبات سری ارزیابی برای مدل منتخب SR-ANN برآورد شد که نتایج آن در شکل

(۳-۱۹) ارائه شده است.



شکل (۳-۱۹) - درصد مشارکت توصیف‌کننده‌ها در مدل بهینه SR-ANN

بر اساس شکل (۳-۱۹)، توصیف‌کننده دما بیشترین اثر مشارکت را دارا می‌باشد.

۳-۵- 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 مورفولینیوم تری سیانو متانید [BMMOR][TCM]، در ۶ دمای مختلف استفاده شده است. از آنجایی که اندازه‌گیری ضریب فعالیت در رقت بی‌نهایت به‌صورت 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 روش‌های کمومتریکس از اهمیت بالایی برخوردار است.

در این تحقیق برای انتخاب توصیف‌کننده‌های مناسب از دو روش SR و CDFS و برای مدل‌سازی از دو روش غیرخطی ANN و SVM استفاده شد. نمودارهای (۳-۱۱) تا (۳-۱۵) و همچنین نتایج موجود در جدول (۳-۱۶) نشان می‌دهد روش SR-ANN روش مناسب‌تری برای پیش‌بینی ضریب فعالیت در رقت بی‌نهایت می‌باشد. همچنین مقایسه نمودارهای خطای باقیمانده نشان می‌دهد که مقدار خطای مطلق باقیمانده بین مقادیر تجربی و مقادیر پیش‌بینی برای سری تست داخلی، در روش SR-ANN گستره‌ی محدودتری دارد که این نشان‌دهنده‌ی خطای کمتر این روش در پیش‌بینی γ_{13}^{∞} است. در نتیجه روش SR-ANN به عنوان بهترین روش برای پیش‌بینی γ_{13}^{∞} ترکیبات آلی و آب در مایع یونی [BMMOR][TCM] می‌باشد.

۳-۶- آینده‌نگری

- ❖ برای انتخاب توصیف‌کننده‌های مناسب، می‌توان از سایر روش‌ها نظیر روش‌های الگوریتم جستجوی ممنوع^۱، الگوریتم اجتماع مورچگان^۲ و الگوریتم بهینه‌سازی جمعیت ذره‌ها^۳ استفاده نمود و نتایج را با نتایج به دست آمده از این تحقیق مقایسه نمود.
- ❖ برای مدل‌سازی می‌توان از حداقل مربعات ماشین‌های بردار پشتیبان (LS- SVM)^۴، سیستم استنتاج فازی-عصبی سازگار (ANFIS)^۵ و روش‌های تابع پایه‌ی شعاعی (RBF)^۶ استفاده نمود.
- ❖ مدلی را ارائه نمود که علاوه بر اثر حل‌شونده‌ها، اثر کاتیون و آنیون مایع یونی (حلال) را نیز در مدل‌های بهینه شامل شود
- ❖ در مدل‌سازی با روش SVM از دیگر پارامترهای کرنل استفاده نمود و نتایج را با نتایج به دست آمده از این تحقیق مقایسه نمود

۱-Tabu Search

۲-Ant Colony Optimization

۳-Honey-bee Matikg Optimization Algorithm

۴-Least Square Support Vectore Machines

۵-Adaptive neuron-fuzzy inference system

۶-Radial basis function

پیوست

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

شماره	نام	دما	$\ln \gamma^\infty$	سری
۱	Pentane	۳۱۸/۱۵	۳/۶۸۱	آموزش
۲	Hexane	۳۱۸/۱۵	۴/۰۳۴	ارزیابی
۳	3-Methylpentane	۳۱۸/۱۵	۳/۹۳۸	آموزش
۴	2,2-Dimethylbutane	۳۱۸/۱۵	۴/۰۳۸	آموزش
۵	Heptane	۳۱۸/۱۵	۴/۴۳۱	تست خارجی
۶	Octane	۳۱۸/۱۵	۴/۸۲۸	آموزش
۷	2,2,4-Trimethylpentane	۳۱۸/۱۵	۴/۶۳۵	آموزش
۸	Nonane	۳۱۸/۱۵	۵/۲۳۱	ارزیابی
۹	Decane	۳۱۸/۱۵	۵/۶۶۶	آموزش
۱۰	Cyclopentane	۳۱۸/۱۵	۲/۶۸۸	تست خارجی
۱۱	Cyclohexane	۳۱۸/۱۵	۳/۱۱۸	آموزش
۱۲	Methylcyclohexane	۳۱۸/۱۵	۳/۵۶۱	آموزش
۱۳	Cycloheptane	۳۱۸/۱۵	۳/۳۳۶	آموزش
۱۴	Cyclooctane	۳۱۸/۱۵	۳/۶۱۱	ارزیابی
۱۵	Hex-1-ene	۳۱۸/۱۵	۳/۱۱۸	آموزش
۱۶	Cyclohexene	۳۱۸/۱۵	۲/۲۴۵	تست خارجی
۱۷	Hept-1-ene	۳۱۸/۱۵	۳/۵۲۹	تست خارجی
۱۸	Oct-1-ene	۳۱۸/۱۵	۳/۹۶۵	آموزش
۱۹	Dec-1-ene	۳۱۸/۱۵	۴/۷۷۱	آموزش
۲۰	Pent-1-yne	۳۱۸/۱۵	۱/۳۰۶	تست داخلی
۲۱	Hex-1-yne	۳۱۸/۱۵	۱/۶۹۹	ارزیابی
۲۲	Hept-1-yne	۳۱۸/۱۵	۲/۱۱۴	آموزش
۲۳	Oct-1-yne	۳۱۸/۱۵	۲/۵۴۹	آموزش
۲۴	Benzene	۳۱۸/۱۵	۰/۴۲۵	تست داخلی
۲۵	Toluene	۳۱۸/۱۵	۰/۸۳۳	آموزش
۲۶	Ethylbenzene	۳۱۸/۱۵	۱/۳۱۹	آموزش
۲۷	o-Xylene	۳۱۸/۱۵	۱/۰۸۵	ارزیابی
۲۸	m-Xylene	۳۱۸/۱۵	۱/۲۶۴	تست داخلی
۲۹	p-Xylene	۳۱۸/۱۵	۱/۲۵	آموزش
۳۰	Styrene	۳۱۸/۱۵	۰/۵۹۹	آموزش
۳۱	α -Methylstyrene	۳۱۸/۱۵	۱/۰۳۰	آموزش
۳۲	Methanol	۳۱۸/۱۵	-۰/۱۹۲	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۳۳	Ethanol	۳۱۸/۱۵	۰/۲۳۱	ارزیابی
۳۴	Propan-1-ol	۳۱۸/۱۵	۰/۴۸۲	آموزش
۳۵	Propan-2-ol	۳۱۸/۱۵	۰/۵۳۶	آموزش
۳۶	Butan-1-ol	۳۱۸/۱۵	۰/۷۷۹	تست خارجی
۳۷	Butan-2-ol	۳۱۸/۱۵	۰/۷۱۸	آموزش
۳۸	2-Methyl-propan-1-ol	۳۱۸/۱۵	۰/۷۴۸	آموزش
۳۹	tert-Butanol	۳۱۸/۱۵	۰/۶۹۳	ارزیابی
۴۰	Water	۳۱۸/۱۵	۰/۰۳۰	آموزش
۴۱	Thiophene	۳۱۸/۱۵	۰/۰۱۰	آموزش
۴۲	Tetrahydrofuran	۳۱۸/۱۵	۰/۱۱۳	آموزش
۴۳	1,4-Dioxane	۳۱۸/۱۵	-۰/۳۴۷	آموزش
۴۴	tert-Butyl methyl ether	۳۱۸/۱۵	۱/۷۴۷	آموزش
۴۵	Ethyl tert-butyl ether	۳۱۸/۱۵	۲/۸۲۱	آموزش
۴۶	Methyl tert-pentyl ether	۳۱۸/۱۵	۲/۱۵۳	تست داخلی
۴۷	Diethyl ether	۳۱۸/۱۵	۱/۶۸۵	آموزش
۴۸	Di-n-propyl ether	۳۱۸/۱۵	۲/۷۹۷	آموزش
۴۹	Di-iso-propyl ether	۳۱۸/۱۵	۲/۸۷۹	آموزش
۵۰	Di-n-butyl ether	۳۱۸/۱۵	۳/۶۸۱	تست خارجی
۵۱	Acetone	۳۱۸/۱۵	-۰/۳	ارزیابی
۵۲	Pentan-2-one	۳۱۸/۱۵	۰/۳۶۵	آموزش
۵۳	Pentan-3-one	۳۱۸/۱۵	۰/۳۵۱	تست داخلی
۵۴	Methyl acetate	۳۱۸/۱۵	۰/۱۸۲	آموزش
۵۵	Ethyl acetate	۳۱۸/۱۵	۰/۶۳۱	تست داخلی
۵۶	Methyl propionate	۳۱۸/۱۵	۰/۴۹۵	آموزش
۵۷	Methyl butyrate	۳۱۸/۱۵	۰/۸۸۴	آموزش
۵۸	Butyraldehyde	۳۱۸/۱۵	۰/۱۸۲	ارزیابی
۵۹	Acetonitrile	۳۱۸/۱۵	-۰/۴۸۳	تست داخلی
۶۰	Pyridine	۳۱۸/۱۵	-۰/۳۱۱	آموزش
۶۱	1-Nitropropane	۳۱۸/۱۵	-۰/۰۱۸	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^\infty$	سری
۶۲	Pentane	۳۲۸/۱۵	۳/۵۷۰	آموزش
۶۳	Hexane	۳۲۸/۱۵	۳/۹۳۴	آموزش
۶۴	3-Methylpentane	۳۲۸/۱۵	۳/۸۲۹	ارزیابی
۶۵	2,2-Dimethylbutane	۳۲۸/۱۵	۳/۹۰۸	آموزش
۶۶	Heptane	۳۲۸/۱۵	۴/۳۴۱	تست خارجی
۶۷	Octane	۳۲۸/۱۵	۴/۷۲۷	آموزش
۶۸	2,2,4-Trimethylpentane	۳۲۸/۱۵	۴/۵۳۷	آموزش
۶۹	Nonane	۳۲۸/۱۵	۵/۱۲۴	آموزش
۷۰	Decane	۳۲۸/۱۵	۵/۵۲۵	ارزیابی
۷۱	Cyclopentane	۳۲۸/۱۵	۲/۶۱۷	تست خارجی
۷۲	Cyclohexane	۳۲۸/۱۵	۳/۰۳۵	آموزش
۷۳	Methylcyclohexane	۳۲۸/۱۵	۳/۴۷۵	آموزش
۷۴	Cycloheptane	۳۲۸/۱۵	۳/۲۵۸	آموزش
۷۵	Cyclooctane	۳۲۸/۱۵	۳/۵۲۹	آموزش
۷۶	Hex-1-ene	۳۲۸/۱۵	۳/۰۵۹	آموزش
۷۷	Cyclohexene	۳۲۸/۱۵	۲/۲۰۶	تست خارجی
۷۸	Hept-1-ene	۳۲۸/۱۵	۳/۴۶۶	تست خارجی
۷۹	Oct-1-ene	۳۲۸/۱۵	۳/۸۹۴	آموزش
۸۰	Dec-1-ene	۳۲۸/۱۵	۴/۷	آموزش
۸۱	Pent-1-yne	۳۲۸/۱۵	۱/۳۱۶	آموزش
۸۲	Hex-1-yne	۳۲۸/۱۵	۱/۷۰۵	ارزیابی
۸۳	Hept-1-yne	۳۲۸/۱۵	۲/۱۱۱	تست داخلی
۸۴	Oct-1-yne	۳۲۸/۱۵	۲/۵۳۴	تست داخلی
۸۵	Benzene	۳۲۸/۱۵	۰/۴۳۸	آموزش
۸۶	Toluene	۳۲۸/۱۵	۰/۸۴۶	آموزش
۸۷	Ethylbenzene	۳۲۸/۱۵	۱/۳۱۹	آموزش
۸۸	o-Xylene	۳۲۸/۱۵	۱/۰۹۵	تست داخلی
۸۹	m-Xylene	۳۲۸/۱۵	۱/۲۷۰	آموزش
۹۰	p-Xylene	۳۲۸/۱۵	۱/۲۵۸	ارزیابی
۹۱	Styrene	۳۲۸/۱۵	۰/۶۲۱	آموزش
۹۲	α -Methylstyrene	۳۲۸/۱۵	۱/۰۶۱	تست داخلی
۹۳	Methanol	۳۲۸/۱۵	-۰/۲۳۶	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۹۴	Ethanol	۳۲۸/۱۵	۰/۱۷۴	آموزش
۹۵	Propan-1-ol	۳۲۸/۱۵	۰/۴۱۹	آموزش
۹۶	Propan-2-ol	۳۲۸/۱۵	۰/۴۶۴	ارزیابی
۹۷	Butan-1-ol	۳۲۸/۱۵	۰/۷۱۳	تست خارجی
۹۸	Butan-2-ol	۳۲۸/۱۵	۰/۶۵۸	آموزش
۹۹	2-Methyl-propan-1-ol	۳۲۸/۱۵	۰/۷۰۸	آموزش
۱۰۰	tert-Butanol	۳۲۸/۱۵	۰/۶۳۱	آموزش
۱۰۱	Water	۳۲۸/۱۵	-۰/۰۲۹	آموزش
۱۰۲	Thiophene	۳۲۸/۱۵	۰/۰۳۹	ارزیابی
۱۰۳	Tetrahydrofuran	۳۲۸/۱۵	۰/۱۴۰	آموزش
۱۰۴	1,4-Dioxane	۳۲۸/۱۵	-۰/۳۰۱	آموزش
۱۰۵	tert-Butyl methyl ether	۳۲۸/۱۵	۱/۷۵۳	آموزش
۱۰۶	Ethyl tert-butyl ether	۳۲۸/۱۵	۲/۷۸۵	آموزش
۱۰۷	Methyl tert-pentyl ether	۳۲۸/۱۵	۲/۱۴۲	تست داخلی
۱۰۸	Diethyl ether	۳۲۸/۱۵	۱/۶۷۳	ارزیابی
۱۰۹	Di-n-propyl ether	۳۲۸/۱۵	۲/۷۶۶	آموزش
۱۱۰	Di-iso-propyl ether	۳۲۸/۱۵	۲/۸۵۶	آموزش
۱۱۱	Di-n-butyl ether	۳۲۸/۱۵	۳/۶۱۹	تست خارجی
۱۱۲	Acetone	۳۲۸/۱۵	-۰/۲۸۸	آموزش
۱۱۳	Pentan-2-one	۳۲۸/۱۵	۰/۳۸۵	آموزش
۱۱۴	Pentan-3-one	۳۲۸/۱۵	۰/۳۶۵	ارزیابی
۱۱۵	Methyl acetate	۳۲۸/۱۵	۰/۲۰۷	آموزش
۱۱۶	Ethyl acetate	۳۲۸/۱۵	۰/۶۴۲	آموزش
۱۱۷	Methyl propionate	۳۲۸/۱۵	۰/۵۱۳	تست داخلی
۱۱۸	Methyl butyrate	۳۲۸/۱۵	۰/۹۰	آموزش
۱۱۹	Butyraldehyde	۳۲۸/۱۵	۰/۲۱۵	آموزش
۱۲۰	Acetonitrile	۳۲۸/۱۵	-۰/۴۷۸	ارزیابی
۱۲۱	Pyridine	۳۲۸/۱۵	-۰/۲۹۰	آموزش
۱۲۲	1-Nitropropane	۳۲۸/۱۵	-۰/۰۱۰	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۱۲۳	Pentane	۳۳۸/۱۵	۳/۴۷۲	آموزش
۱۲۴	Hexane	۳۳۸/۱۵	۳/۸۴۴	تست داخلی
۱۲۵	3-Methylpentane	۳۳۸/۱۵	۳/۷۴۵	آموزش
۱۲۶	2,2-Dimethylbutane	۳۳۸/۱۵	۳/۷۹۵	ارزیابی
۱۲۷	Heptane	۳۳۸/۱۵	۴/۲۴۳	تست خارجی
۱۲۸	Octane	۳۳۸/۱۵	۴/۶۲۵	آموزش
۱۲۹	2,2,4-Trimethylpentane	۳۳۸/۱۵	۴/۴۳۳	آموزش
۱۳۰	Nonane	۳۳۸/۱۵	۵/۰۱۷	آموزش
۱۳۱	Decane	۳۳۸/۱۵	۵/۴۱۲	آموزش
۱۳۲	Cyclopentane	۳۳۸/۱۵	۲/۵۴۹	تست خارجی
۱۳۳	Cyclohexane	۳۳۸/۱۵	۲/۹۶۰	آموزش
۱۳۴	Methylcyclohexane	۳۳۸/۱۵	۳/۳۹۱	آموزش
۱۳۵	Cycloheptane	۳۳۸/۱۵	۳/۱۷۸	تست داخلی
۱۳۶	Cyclooctane	۳۳۸/۱۵	۳/۴۴۰	آموزش
۱۳۷	Hex-1-ene	۳۳۸/۱۵	۲/۹۹۶	آموزش
۱۳۸	Cyclohexene	۳۳۸/۱۵	۲/۱۶۴	تست خارجی
۱۳۹	Hept-1-ene	۳۳۸/۱۵	۳/۴۰۸	تست خارجی
۱۴۰	Oct-1-ene	۳۳۸/۱۵	۳/۸۲۲	ارزیابی
۱۴۱	Dec-1-ene	۳۳۸/۱۵	۴/۶۱۵	آموزش
۱۴۲	Pent-1-yne	۳۳۸/۱۵	۱/۳۲۲	آموزش
۱۴۳	Hex-1-yne	۳۳۸/۱۵	۱/۷۰۵	آموزش
۱۴۴	Hept-1-yne	۳۳۸/۱۵	۲/۱۰۴	تست داخلی
۱۴۵	Oct-1-yne	۳۳۸/۱۵	۲/۵۱۸	تست داخلی
۱۴۶	Benzene	۳۳۸/۱۵	۰/۴۵۱	آموزش
۱۴۷	Toluene	۳۳۸/۱۵	۰/۸۵۹	ارزیابی
۱۴۸	Ethylbenzene	۳۳۸/۱۵	۱/۳۲۲	آموزش
۱۴۹	o-Xylene	۳۳۸/۱۵	۱/۱۰۵	آموزش
۱۵۰	m-Xylene	۳۳۸/۱۵	۱/۲۸۴	آموزش
۱۵۱	p-Xylene	۳۳۸/۱۵	۱/۲۶۴	آموزش
۱۵۲	Styrene	۳۳۸/۱۵	۰/۶۴۲	تست داخلی
۱۵۳	α -Methylstyrene	۳۳۸/۱۵	۱/۰۹۵	ارزیابی
۱۵۴	Methanol	۳۳۸/۱۵	-۰/۲۷۲	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۱۵۵	Ethanol	۳۳۸/۱۵	۰/۱۱۳	آموزش
۱۵۶	Propan-1-ol	۳۳۸/۱۵	۰/۳۵۸	آموزش
۱۵۷	Propan-2-ol	۳۳۸/۱۵	۰/۴۰۵	آموزش
۱۵۸	Butan-1-ol	۳۳۸/۱۵	۰/۶۴۷	تست خارجی
۱۵۹	Butan-2-ol	۳۳۸/۱۵	۰/۵۹۳	تست داخلی
۱۶۰	2-Methyl-propan-1-ol	۳۳۸/۱۵	۰/۶۲۶	ارزیابی
۱۶۱	tert-Butanol	۳۳۸/۱۵	۰/۵۷۱	تست داخلی
۱۶۲	Water	۳۳۸/۱۵	-۰/۰۸۸	آموزش
۱۶۳	Thiophene	۳۳۸/۱۵	۰/۰۵۸	آموزش
۱۶۴	Tetrahydrofuran	۳۳۸/۱۵	۰/۱۵۷	آموزش
۱۶۵	1,4-Dioxane	۳۳۸/۱۵	-۰/۲۵۹	آموزش
۱۶۶	tert-Butyl methyl ether	۳۳۸/۱۵	۱/۷۴۷	تست داخلی
۱۶۷	Ethyl tert-butyl ether	۳۳۸/۱۵	۲/۷۴۷	ارزیابی
۱۶۸	Methyl tert-pentyl ether	۳۳۸/۱۵	۲/۱۳۱	آموزش
۱۶۹	Diethyl ether	۳۳۸/۱۵	۱/۶۶۲	آموزش
۱۷۰	Di-n-propyl ether	۳۳۸/۱۵	۲/۷۳۴	آموزش
۱۷۱	Di-iso-propyl ether	۳۳۸/۱۵	۲/۸۲۱	آموزش
۱۷۲	Di-n-butyl ether	۳۳۸/۱۵	۳/۵۵۸	تست خارجی
۱۷۳	Acetone	۳۳۸/۱۵	-۰/۲۶۸	ارزیابی
۱۷۴	Pentan-2-one	۳۳۸/۱۵	۰/۳۹۹	تست داخلی
۱۷۵	Pentan-3-one	۳۳۸/۱۵	۰/۳۸۵	آموزش
۱۷۶	Methyl acetate	۳۳۸/۱۵	۰/۲۱۵	آموزش
۱۷۷	Ethyl acetate	۳۳۸/۱۵	۰/۶۵۲	آموزش
۱۷۸	Methyl propionate	۳۳۸/۱۵	۰/۵۳۱	آموزش
۱۷۹	Methyl butyrate	۳۳۸/۱۵	۰/۹۱۲	تست داخلی
۱۸۰	Butyraldehyde	۳۳۸/۱۵	۰/۲۳۱	ارزیابی
۱۸۱	Acetonitrile	۳۳۸/۱۵	-۰/۴۷۲	آموزش
۱۸۲	Pyridine	۳۳۸/۱۵	-۰/۲۷۷	آموزش
۱۸۳	1-Nitropropane	۳۳۸/۱۵	-۰/۰۰۲	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۱۸۴	Pentane	۳۴۸/۱۵	۳/۳۹۵	آموزش
۱۸۵	Hexane	۳۴۸/۱۵	۳/۷۶۶	ارزیابی
۱۸۶	3-Methylpentane	۳۴۸/۱۵	۳/۶۶۱	آموزش
۱۸۷	2,2-Dimethylbutane	۳۴۸/۱۵	۳/۷۱۱	آموزش
۱۸۸	Heptane	۳۴۸/۱۵	۴/۱۴۸	تست خارجی
۱۸۹	Octane	۳۴۸/۱۵	۴/۵۳۷	تست داخلی
۱۹۰	2,2,4-Trimethylpentane	۳۴۸/۱۵	۴/۳۴۳	آموزش
۱۹۱	Nonane	۳۴۸/۱۵	۴/۹۲۰	آموزش
۱۹۲	Decane	۳۴۸/۱۵	۵/۳۰۸	تست داخلی
۱۹۳	Cyclopentane	۳۴۸/۱۵	۲/۴۹۳	تست خارجی
۱۹۴	Cyclohexane	۳۴۸/۱۵	۲/۸۹۰	تست داخلی
۱۹۵	Methylcyclohexane	۳۴۸/۱۵	۳/۳۱۸	ارزیابی
۱۹۶	Cycloheptane	۳۴۸/۱۵	۳/۱۱۴	آموزش
۱۹۷	Cyclooctane	۳۴۸/۱۵	۳/۳۷۱	آموزش
۱۹۸	Hex-1-ene	۳۴۸/۱۵	۲/۹۵۰	تست داخلی
۱۹۹	Cyclohexene	۳۴۸/۱۵	۲/۱۳۵	تست خارجی
۲۰۰	Hept-1-ene	۳۴۸/۱۵	۳/۳۵۷	تست خارجی
۲۰۱	Oct-1-ene	۳۴۸/۱۵	۳/۷۶۱	آموزش
۲۰۲	Dec-1-ene	۳۴۸/۱۵	۴/۵۴۹	ارزیابی
۲۰۳	Pent-1-yne	۳۴۸/۱۵	۱/۳۳۵	آموزش
۲۰۴	Hex-1-yne	۳۴۸/۱۵	۱/۷۱۲	آموزش
۲۰۵	Hept-1-yne	۳۴۸/۱۵	۲/۱۰۴	تست داخلی
۲۰۶	Oct-1-yne	۳۴۸/۱۵	۲/۵۰۱	تست داخلی
۲۰۷	Benzene	۳۴۸/۱۵	۰/۴۵۷	آموزش
۲۰۸	Toluene	۳۴۸/۱۵	۰/۸۶۷	آموزش
۲۰۹	Ethylbenzene	۳۴۸/۱۵	۱/۳۲۲	آموزش
۲۱۰	o-Xylene	۳۴۸/۱۵	۱/۱۱۲	آموزش
۲۱۱	m-Xylene	۳۴۸/۱۵	۱/۲۸۴	آموزش
۲۱۲	p-Xylene	۳۴۸/۱۵	۱/۲۷۳	تست داخلی
۲۱۳	Styrene	۳۴۸/۱۵	۰/۶۵۸	آموزش
۲۱۴	α -Methylstyrene	۳۴۸/۱۵	۱/۱۲۲	آموزش
۲۱۵	Methanol	۳۴۸/۱۵	-۰/۳۱۶	ارزیابی

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۲۱۶	Ethanol	۳۴۸/۱۵	۰/۰۵۸	آموزش
۲۱۷	Propan-1-ol	۳۴۸/۱۵	۰/۳	آموزش
۲۱۸	Propan-2-ol	۳۴۸/۱۵	۰/۳۳۶	آموزش
۲۱۹	Butan-1-ol	۳۴۸/۱۵	۰/۵۸۲	تست خارجی
۲۲۰	Butan-2-ol	۳۴۸/۱۵	۰/۵۳۶	آموزش
۲۲۱	2-Methyl-propan-1-ol	۳۴۸/۱۵	۰/۵۵۴	تست داخلی
۲۲۲	tert-Butanol	۳۴۸/۱۵	۰/۵۲۵	ارزیابی
۲۲۳	Water	۳۴۸/۱۵	-۰/۱۵۱	آموزش
۲۲۴	Thiophene	۳۴۸/۱۵	۰/۰۷۷	آموزش
۲۲۵	Tetrahydrofuran	۳۴۸/۱۵	۰/۱۶۶	آموزش
۲۲۶	1,4-Dioxane	۳۴۸/۱۵	-۰/۲۲۶	آموزش
۲۲۷	tert-Butyl methyl ether	۳۴۸/۱۵	۱/۷۵۱	ارزیابی
۲۲۸	Ethyl tert-butyl ether	۳۴۸/۱۵	۲/۷۲۸	آموزش
۲۲۹	Methyl tert-pentyl ether	۳۴۸/۱۵	۲/۱۲۵	آموزش
۲۳۰	Diethyl ether	۳۴۸/۱۵	۱/۶۵۶	آموزش
۲۳۱	Di-n-propyl ether	۳۴۸/۱۵	۲/۷۰۱	تست داخلی
۲۳۲	Di-iso-propyl ether	۳۴۸/۱۵	۲/۸۰۳	آموزش
۲۳۳	Di-n-butyl ether	۳۴۸/۱۵	۳/۵۰۶	تست خارجی
۲۳۴	Acetone	۳۴۸/۱۵	-۰/۲۵۶	ارزیابی
۲۳۵	Pentan-2-one	۳۴۸/۱۵	۰/۴۰۵	آموزش
۲۳۶	Pentan-3-one	۳۴۸/۱۵	۰/۳۹۹	تست داخلی
۲۳۷	Methyl acetate	۳۴۸/۱۵	۰/۲۳۱	آموزش
۲۳۸	Ethyl acetate	۳۴۸/۱۵	۰/۶۵۸	تست داخلی
۲۳۹	Methyl propionate	۳۴۸/۱۵	۰/۵۴۸	آموزش
۲۴۰	Methyl butyrate	۳۴۸/۱۵	۰/۹۲۰	تست داخلی
۲۴۱	Butyraldehyde	۳۴۸/۱۵	۰/۲۶۲	آموزش
۲۴۲	Acetonitrile	۳۴۸/۱۵	-۰/۴۷۲	آموزش
۲۴۳	Pyridine	۳۴۸/۱۵	-۰/۲۶۷	آموزش
۲۴۴	1-Nitropropane	۳۴۸/۱۵	۰	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۲۴۵	Pentane	۳۵۸/۱۵	۳/۳۲۱	آموزش
۲۴۶	Hexane	۳۵۸/۱۵	۳/۶۹۶	آموزش
۲۴۷	3-Methylpentane	۳۵۸/۱۵	۳/۵۹۲	ارزیابی
۲۴۸	2,2-Dimethylbutane	۳۵۸/۱۵	۳/۶۳۰	آموزش
۲۴۹	Heptane	۳۵۸/۱۵	۴/۰۷۲	تست خارجی
۲۵۰	Octane	۳۵۸/۱۵	۴/۴۶۰	آموزش
۲۵۱	2,2,4-Trimethylpentane	۳۵۸/۱۵	۴/۲۶۰	آموزش
۲۵۲	Nonane	۳۵۸/۱۵	۴/۸۳۶	آموزش
۲۵۳	Decane	۳۵۸/۱۵	۵/۲۰۹	ارزیابی
۲۵۴	Cyclopentane	۳۵۸/۱۵	۲/۴۳۴	تست خارجی
۲۵۵	Cyclohexane	۳۵۸/۱۵	۲/۸۲۷	تست داخلی
۲۵۶	Methylcyclohexane	۳۵۸/۱۵	۳/۲۴۶	آموزش
۲۵۷	Cycloheptane	۳۵۸/۱۵	۳/۰۴۵	آموزش
۲۵۸	Cyclooctane	۳۵۸/۱۵	۳/۲۹۶	ارزیابی
۲۵۹	Hex-1-ene	۳۵۸/۱۵	۲/۹۰۱	آموزش
۲۶۰	Cyclohexene	۳۵۸/۱۵	۲/۰۹۹	تست خارجی
۲۶۱	Hept-1-ene	۳۵۸/۱۵	۳/۳۰۳	تست خارجی
۲۶۲	Oct-1-ene	۳۵۸/۱۵	۳/۷۰۴	آموزش
۲۶۳	Dec-1-ene	۳۵۸/۱۵	۴/۴۷۵	آموزش
۲۶۴	Pent-1-yne	۳۵۸/۱۵	۱/۳۳۸	آموزش
۲۶۵	Hex-1-yne	۳۵۸/۱۵	۱/۷۱۰	آموزش
۲۶۶	Hept-1-yne	۳۵۸/۱۵	۲/۰۹۶	تست داخلی
۲۶۷	Oct-1-yne	۳۵۸/۱۵	۲/۴۸۵	ارزیابی
۲۶۸	Benzene	۳۵۸/۱۵	۰/۴۷۶	آموزش
۲۶۹	Toluene	۳۵۸/۱۵	۰/۸۸۰	آموزش
۲۷۰	Ethylbenzene	۳۵۸/۱۵	۱/۳۲۲	آموزش
۲۷۱	o-Xylene	۳۵۸/۱۵	۱/۱۲۲	آموزش
۲۷۲	m-Xylene	۳۵۸/۱۵	۱/۲۸۶	ارزیابی
۲۷۳	p-Xylene	۳۵۸/۱۵	۱/۲۷۸	آموزش
۲۷۴	Styrene	۳۵۸/۱۵	۰/۶۷۳	آموزش
۲۷۵	α -Methylstyrene	۳۵۸/۱۵	۱/۱۴۷	آموزش
۲۷۶	Methanol	۳۵۸/۱۵	-۰/۳۵۱	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۲۷۷	Ethanol	۳۵۸/۱۵	۰/۰۱۰	ارزیابی
۲۷۸	Propan-1-ol	۳۵۸/۱۵	۰/۲۵۵	آموزش
۲۷۹	Propan-2-ol	۳۵۸/۱۵	۰/۲۸۵	آموزش
۲۸۰	Butan-1-ol	۳۵۸/۱۵	۰/۵۳۱	تست خارجی
۲۸۱	Butan-2-ol	۳۵۸/۱۵	۰/۴۹۵	آموزش
۲۸۲	2-Methyl-propan-1-ol	۳۵۸/۱۵	۰/۵۰۱	آموزش
۲۸۳	tert-Butanol	۳۵۸/۱۵	۰/۴۸۹	تست داخلی
۲۸۴	Water	۳۵۸/۱۵	-۰/۲۰۰	آموزش
۲۸۵	Thiophene	۳۵۸/۱۵	۰/۰۹۵	آموزش
۲۸۶	Tetrahydrofuran	۳۵۸/۱۵	۰/۱۹۱	آموزش
۲۸۷	1,4-Dioxane	۳۵۸/۱۵	-۰/۱۸۵	تست داخلی
۲۸۸	tert-Butyl methyl ether	۳۵۸/۱۵	۱/۷۵۳	آموزش
۲۸۹	Ethyl tert-butyl ether	۳۵۸/۱۵	۲/۷۰۱	آموزش
۲۹۰	Methyl tert-pentyl ether	۳۵۸/۱۵	۲/۱۱۵	ارزیابی
۲۹۱	Diethyl ether	۳۵۸/۱۵	۱/۶۵۱	آموزش
۲۹۲	Di-n-propyl ether	۳۵۸/۱۵	۲/۶۶۷	تست داخلی
۲۹۳	Di-iso-propyl ether	۳۵۸/۱۵	۲/۷۷۳	تست داخلی
۲۹۴	Di-n-butyl ether	۳۵۸/۱۵	۳/۴۵۰	تست خارجی
۲۹۵	Acetone	۳۵۸/۱۵	-۰/۲۳۴	آموزش
۲۹۶	Pentan-2-one	۳۵۸/۱۵	۰/۴۲۵	آموزش
۲۹۷	Pentan-3-one	۳۵۸/۱۵	۰/۴۱۲	آموزش
۲۹۸	Methyl acetate	۳۵۸/۱۵	۰/۲۵۵	ارزیابی
۲۹۹	Ethyl acetate	۳۵۸/۱۵	۰/۶۷۳	آموزش
۳۰۰	Methyl propionate	۳۵۸/۱۵	۰/۵۶۵	آموزش
۳۰۱	Methyl butyrate	۳۵۸/۱۵	۰/۹۳۲	تست داخلی
۳۰۲	Butyraldehyde	۳۵۸/۱۵	۰/۲۸۵	آموزش
۳۰۳	Acetonitrile	۳۵۸/۱۵	-۰/۴۶۴	آموزش
۳۰۴	Pyridine	۳۵۸/۱۵	-۰/۲۵۴	ارزیابی
۳۰۵	1-Nitropropane	۳۵۸/۱۵	۰/۰۱۰	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۳۰۶	Pentane	۳۶۸/۱۵	۳/۲۶۶	آموزش
۳۰۷	Hexane	۳۶۸/۱۵	۳/۶۱۶	آموزش
۳۰۸	3-Methylpentane	۳۶۸/۱۵	۳/۵۱۲	آموزش
۳۰۹	2,2-Dimethylbutane	۳۶۸/۱۵	۳/۵۵۰	ارزیابی
۳۱۰	Heptane	۳۶۸/۱۵	۳/۹۸۳	تست خارجی
۳۱۱	Octane	۳۶۸/۱۵	۴/۳۶۴	آموزش
۳۱۲	2,2,4-Trimethylpentane	۳۶۸/۱۵	۴/۱۶۷	آموزش
۳۱۳	Nonane	۳۶۸/۱۵	۴/۷۴۵	آموزش
۳۱۴	Decane	۳۶۸/۱۵	۵/۱۱۲	آموزش
۳۱۵	Cyclopentane	۳۶۸/۱۵	۲/۳۸۹	تست خارجی
۳۱۶	Cyclohexane	۳۶۸/۱۵	۲/۷۷۳	ارزیابی ارزیابی
۳۱۷	Methylcyclohexane	۳۶۸/۱۵	۳/۱۷۸	آموزش
۳۱۸	Cycloheptane	۳۶۸/۱۵	۲/۹۸۱	آموزش
۳۱۹	Cyclooctane	۳۶۸/۱۵	۳/۲۲۳	آموزش
۳۲۰	Hex-1-ene	۳۶۸/۱۵	۲/۸۵۶	آموزش
۳۲۱	Cyclohexene	۳۶۸/۱۵	۲/۰۶۹	تست خارجی
۳۲۲	Hept-1-ene	۳۶۸/۱۵	۳/۲۴۶	تست خارجی
۳۲۳	Oct-1-ene	۳۶۸/۱۵	۳/۶۴۰	ارزیابی
۳۲۴	Dec-1-ene	۳۶۸/۱۵	۴/۳۹۹	آموزش
۳۲۵	Pent-1-yne	۳۶۸/۱۵	۱/۳۴۳	آموزش
۳۲۶	Hex-1-yne	۳۶۸/۱۵	۱/۷۰۷	تست داخلی
۳۲۷	Hept-1-yne	۳۶۸/۱۵	۲/۰۸۶	آموزش
۳۲۸	Oct-1-yne	۳۶۸/۱۵	۲/۴۶۸	ارزیابی
۳۲۹	Benzene	۳۶۸/۱۵	۰/۴۸۲	آموزش
۳۳۰	Toluene	۳۶۸/۱۵	۰/۸۸۸	تست داخلی
۳۳۱	Ethylbenzene	۳۶۸/۱۵	۱/۳۲۲	آموزش
۳۳۲	o-Xylene	۳۶۸/۱۵	۱/۱۲۵	آموزش
۳۳۳	m-Xylene	۳۶۸/۱۵	۱/۲۸۶	آموزش
۳۳۴	p-Xylene	۳۶۸/۱۵	۱/۲۸۹	تست داخلی
۳۳۵	Styrene	۳۶۸/۱۵	۰/۶۸۸	ارزیابی
۳۳۶	α -Methylstyrene	۳۶۸/۱۵	۱/۱۶۹	آموزش
۳۳۷	Methanol	۳۶۸/۱۵	-۰/۳۹۰	آموزش

ادامه جدول (پ-۱)

شماره	نام	دما	$\ln \gamma^{\infty}$	سری
۳۳۸	Ethanol	۳۶۸/۱۵	-۰/۰۳۶	آموزش
۳۳۹	Propan-1-ol	۳۶۸/۱۵	۰/۲۰۷	آموزش
۳۴۰	Propan-2-ol	۳۶۸/۱۵	۰/۲۲۳	ارزیابی
۳۴۱	Butan-1-ol	۳۶۸/۱۵	۰/۴۷۶	تست خارجی
۳۴۲	Butan-2-ol	۳۶۸/۱۵	۰/۴۵۷	آموزش
۳۴۳	2-Methyl-propan-1-ol	۳۶۸/۱۵	۰/۴۴۵	آموزش
۳۴۴	tert-Butanol	۳۶۸/۱۵	۰/۴۵۷	آموزش
۳۴۵	Water	۳۶۸/۱۵	-۰/۲۴۸	آموزش
۳۴۶	Thiophene	۳۶۸/۱۵	۰/۱۱۳	آموزش
۳۴۷	Tetrahydrofuran	۳۶۸/۱۵	۰/۱۹۹	آموزش
۳۴۸	1,4-Dioxane	۳۶۸/۱۵	-۰/۱۵۹	آموزش
۳۴۹	tert-Butyl methyl ether	۳۶۸/۱۵	۱/۷۵۴	تست داخلی
۳۵۰	Ethyl tert-butyl ether	۳۶۸/۱۵	۲/۶۸۱	تست داخلی
۳۵۱	Methyl tert-pentyl ether	۳۶۸/۱۵	۲/۱۰۵	آموزش
۳۵۲	Diethyl ether	۳۶۸/۱۵	۱/۶۴۱	آموزش
۳۵۳	Di-n-propyl ether	۳۶۸/۱۵	۲/۶۳۲	ارزیابی
۳۵۴	Di-iso-propyl ether	۳۶۸/۱۵	۲/۷۴۷	تست داخلی
۳۵۵	Di-n-butyl ether	۳۶۸/۱۵	۳/۳۹۱	تست خارجی
۳۵۶	Acetone	۳۶۸/۱۵	-۰/۲۲۶	تست داخلی
۳۵۷	Pentan-2-one	۳۶۸/۱۵	۰/۴۳۲	تست داخلی
۳۵۸	Pentan-3-one	۳۶۸/۱۵	۰/۴۳۲	آموزش
۳۵۹	Methyl acetate	۳۶۸/۱۵	۰/۲۶۲	آموزش
۳۶۰	Ethyl acetate	۳۶۸/۱۵	۰/۶۸۳	آموزش
۳۶۱	Methyl propionate	۳۶۸/۱۵	۰/۵۸۲	آموزش
۳۶۲	Methyl butyrate	۳۶۸/۱۵	۰/۹۴۰	ارزیابی
۳۶۳	Butyraldehyde	۳۶۸/۱۵	۰/۳	آموزش
۳۶۴	Acetonitrile	۳۶۸/۱۵	-۰/۴۶۲	آموزش
۳۶۵	Pyridine	۳۶۸/۱۵	-۰/۲۳۸	آموزش
۳۶۶	1-Nitropropane	۳۶۸/۱۵	۰/۰۱۰	آموزش

جدول (پ-۲) - نتایج حاصل از ارزیابی مدل‌های SR-ANN, CDFS-ANN, SR-SVM, CDFS-SVM با استفاده از

سری تست خارجی

شماره ترکیب	مقدار تجربی	مقادیر پیش‌بینی شده			
		SR-ANN	CDFS-ANN	SR-SVM	CDFS-SVM
۵	۴/۴۳۱	۴/۴۵۳	۴/۴۷۸	۴/۴۱۹	۴/۴۳۱
۱۰	۲/۶۸۸	۲/۹۲۷	۲/۶۸۰	۲/۷۴۳	۲/۲۱۳
۱۶	۲/۲۴۵	۲/۳۲۹	۱/۶۶۲	۱/۶۶۸	۱/۷۰۹
۱۷	۳/۵۲۹	۳/۶۴۴	۳/۵۴۵	۳/۳۸۵	۳/۴۶۸
۳۶	۰/۷۷۹	۱/۲۰۶	۰/۷۶۳	۰/۵۹۸	۰/۹۴۰
۵۰	۳/۶۸۱	۳/۴۶۵	۲/۸۷۱	۳/۹۷۶	۳/۳۳۳
۶۶	۴/۳۴۱	۴/۳۴۷	۴/۳۸۴	۴/۳۴۸	۴/۳۰۴
۷۱	۲/۶۱۷	۲/۸۱۸	۲/۶۳۳	۲/۷۲۱	۲/۱۱۴
۷۷	۲/۲۰۶	۲/۲۹۳	۱/۶۴۳	۱/۷۰۶	۱/۵۹۱
۷۸	۳/۴۶۶	۳/۵۷۸	۳/۴۸۲	۳/۳۶۵	۳/۳۶۱
۹۷	۰/۷۱۳	۱/۱۶۰	۰/۶۹۳	۰/۴۸۲	۰/۸۴۸
۱۱۱	۳/۶۱۹	۳/۴۳۰	۲/۸۴۹	۴/۲۴۸	۳/۳۳۰
۱۲۷	۴/۲۴۳	۴/۲۵۰	۴/۲۹۶	۴/۲۰۴	۴/۱۸۸
۱۳۲	۲/۵۴۹	۲/۷۱۷	۲/۵۸۶	۲/۷۰۳	۲/۰۶۵
۱۳۸	۲/۱۶۴	۲/۲۵۶	۱/۶۲۷	۱/۶۷۰	۱/۵۵۱
۱۳۹	۳/۴۰۸	۳/۵۱۴	۳/۴۲۲	۳/۲۴۱	۳/۲۸۲
۱۵۸	۰/۶۴۷	۱/۱۱۹	۰/۶۲۶	۰/۳۹۶	۰/۷۴۳
۱۷۲	۳/۵۵۸	۳/۳۹۴	۲/۸۳۰	۴/۰۵۲	۳/۳۰۹
۱۸۸	۴/۱۴۸	۴/۱۶۲	۴/۲۱۳	۴/۱۸۴	۴/۱۱۴
۱۹۳	۲/۴۹۳	۲/۶۲۳	۲/۵۳۹	۲/۶۱۵	۲/۰۴۴
۱۹۹	۲/۱۳۵	۲/۲۱۶	۱/۶۱۳	۱/۶۲۳	۱/۵۶۰
۲۰۰	۳/۳۵۷	۳/۴۵۵	۳/۳۶۴	۳/۲۰۵	۳/۲۵۲
۲۱۹	۰/۵۸۲	۱/۰۸۲	۰/۵۵۹	۰/۴۰۵	۰/۶۴۰
۲۳۳	۳/۵۰۶	۳/۳۵۸	۲/۸۱۴	۳/۹۶۴	۳/۲۸۰

ادامه جدول (پ-۲)

شماره ترکیب	مقدار تجربی	مقادیر پیش‌بینی شده			
		SR-ANN	CDFS-ANN	SR-SVM	CDFS-SVM
۲۴۹	۴/۰۷۲	۴/۰۸۴	۴/۱۳۷	۴/۰۶۷	۴/۰۸۷
۲۵۴	۲/۴۳۴	۲/۵۳۸	۲/۴۹۳	۲/۴۷۸	۲/۰۶۲
۲۶۰	۲/۰۹۹	۲/۱۷۵	۱/۶۰۲	۱/۵۷۸	۱/۵۹۴
۲۶۱	۳/۳۰۳	۳/۴۰۰	۳/۳۱	۳/۱۲۸	۳/۲۵۵
۲۸۰	۰/۵۳۱	۱/۰۵۰	۰/۴۹۵	۰/۳۹۳	۰/۵۹۸
۲۹۴	۳/۴۵	۳/۳۲۱	۲/۸۰۱	۳/۷۴۷	۳/۲۷۶
۳۱۰	۳/۹۸۳	۴/۰۱۷	۴/۰۶۶	۴/۰۶۵	۴/۰۱۳
۳۱۵	۲/۳۸۹	۲/۴۶۱	۲/۴۴۸	۲/۵۳۶	۲/۰۵۵
۳۲۱	۲/۰۶۹	۲/۱۳۲	۱/۵۹۳	۱/۶۶۷	۱/۵۴۳
۳۲۲	۳/۲۴۶	۳/۳۴۹	۳/۲۵۸	۳/۱۸۵	۳/۱۶۳
۳۴۱	۰/۴۷۶	۱/۰۲۴	۰/۴۳۳۹	۰/۴۴۶	۰/۵۸۳
۳۵۵	۳/۳۹۱	۳/۲۸۵	۲/۷۹۱۷	۳/۶۴۳	۳/۲۶۳
*	*	*	*	*	*

منابع

- [1] Domańska U.& Lukoshko E. V. (2013) "Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-methyl pyrrolidinium tricyanomethanid" *J. Chem. Thermodyn.*, 66, pp 144-150.
- [۲] نظیفی م، (۱۳۹۰)، پایان‌نامه کارشناسی ارشد: "مایعات یونی دو کاتیونی جدید بر اساس پیرولیدینیوم و مورفولین به عنوان فاز ساکن در کروماتوگرافی گازی و محاسبه پارامترهای ترمودینامیکی مواد با روش کروماتوگرافی گازی معکوس"، پژوهشکده محیط زیست و فن آوری‌های پاک، پژوهشگاه شیمی و مهندسی شیمی ایران
- [3] Nami F.& Deyhimi F. (2011) "Prediction of activity coefficients at infinite dilution for organic solutes in ionic liquids by artificial neural network" *J. Chem. Thermodyn.*, 43, pp 22-27.
- [4] Domańska U.& Lukoshko E. V. (2014) "Thermodynamic and activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-methyl morpholinium tricyanomethanid" *J. Chem. Thermodyn.*, 68, pp 53-59.
- [۵] دهنوی م، (۱۳۸۸)، پایان‌نامه کارشناسی ارشد: "مطالعه و بررسی خواص مایعات یونی و بررسی ضریب گزینش پذیری مایعات یونی در استخراج مایع-مایع"، دانشکده شیمی، دانشگاه صنعتی شریف
- [۶] صدق کردار م، (۱۳۸۸)، پایان‌نامه کارشناسی ارشد: "پیش بینی حلالیت گازها در مایعات یونی با استفاده از مدل UNIFAC"، دانشکده مهندسی شیمی و نفت، دانشگاه صنعتی شریف.
- [۷] ابراهیمی م، (۱۳۹۱)، پایان‌نامه کارشناسی ارشد: "جداسازی ترکیبات حلقوی (آروماتیک) از برش‌های نفتی با استفاده از مایعات یونی ویژه"، دانشکده مهندسی شیمی و نفت، دانشگاه صنعتی شریف.
- [۸] خانمحمدی خرمی م، بدر ح، (۱۳۸۶) "شیمی تجزیه دستگاهی" چاپ اول، مرکز نشر دانشگاه بین المللی امام خمینی، قزوین.

[۹] سالمی ا، (۱۳۷۹)، پایان نامه کارشناسی ارشد: "گونه شناسی ترکیبات آلی قلع بوسیله تلفیق

کروماتوگرافی گازی و طیف سنج اتمی"، دانشکده شیمی، دانشگاه صنعتی شریف.

[10] 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, pp 1039-1048.

[11] 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.

[12] 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, pp 195-201.

[۱۳] طاهرزاده م، (۱۳۹۳)، پایان نامه کارشناسی ارشد: "مطالعه کمی ساختار- خاصیت ضریب فعالیت

در رقت بی نهایت ترکیبات آلی و آب در محیط مایع یونی ۱-بوتیل-۱-متیل پیرولیدینیوم تریسیانومتانید"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[۱۴] پرستار شهری ه، (۱۳۹۰)، پایان نامه دکتری: "توسعه و کاربرد روش های تفکیک منحنی چند

متغیره جهت آنالیز داده های پیچیده کروماتوگرافی"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۱۵] برهانی ا، (۱۳۸۷)، پایان نامه کارشناسی ارشد: "کاربرد الگوریتم مورچه جهت انتخاب

توصیف کننده ها در مطالعات QSPR مربوط به طول موج ماکزیمم مشتقات آنتراکینون و ضریب فعالیت هیدروکربن ها در رقت بی نهایت"، دانشکده شیمی، دانشگاه دامغان.

[۱۶] ژبانی م، (۱۳۹۲)، پایان نامه کارشناسی ارشد: "پیش بینی گرمای جذب آلکان ها بر روی

ژئولیت های مختلف بر پایه شبکه های عصبی مصنوعی"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۱۷] ابراهیمی ز، (۱۳۹۲)، پایان نامه کارشناسی ارشد: "مدل سازی خطی و غیر خطی برای پیش بینی

انبساط پذیری هم فشار سیالات آلی با استفاده از استراتژی CDFS"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

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

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

[۲۰] کیانی آ، (۱۳۸۲)، پایان نامه کارشناسی ارشد: "مدل‌سازی کامپیوتری زمان بازداری ترکیبات آلی فرار با استفاده از روش‌های کمومتریکس"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۲۱] آزمون فر ن، (۱۳۹۲)، پایان نامه کارشناسی ارشد: "مدل‌سازی شبکه‌ی عصبی مصنوعی برای پیش‌بینی دانسیته‌ی سیالات تجمعی با استفاده از توصیف‌کننده‌های مولکولی"، دانشکده‌ی شیمی، دانشگاه صنعتی شاهرود.

[22] Hemmateenejad B., Javadnia K. & Elyasi M. (2007) "Quantitative structure–retention relationship for the Kovats retention indices of a large set of terpenes: A combined data splitting-feature selection strategy", *Anal. Chim. Acta*, 592 , pp 72–81

[23] Hemmateenejad B., Shamsipur M., Miri R., Elyasi M., Foroghnia F. & Sharghi H. (2008) "Linear and nonlinear quantitative structure–property relationship models for solubility of some anthraquinone, anthrone and xanthone derivatives in supercritical carbon dioxide", *Anal. Chim. Acta.*, 610, pp 25–34

[۲۴] شهابی ز، (۱۳۹۲)، پایان‌نامه کارشناسی ارشد: "پیش‌بینی ارتباط کمی ساختار-خاصیت ضریب خودنفوذی آلکان‌ها"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

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

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

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

[۲۸] نکویی س، (۱۳۹۴)، پایان نامه کارشناسی ارشد: "کاربرد روش‌های مختلف کمومتری کس جهت

پیش‌بینی فعالیت برخی از ترکیبات دارویی"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[29] Xu J., Zhang H., Wang L., Ye W., Xu W. & Li Z.(2010) "QSPR analysis of infinite dilution activity coefficients of chlorinated organic compounds in water" *Fluid Phase Equilibria.*, 291, pp 111-116

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

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

[۳۲] یزدی‌زاده راوری ا، (۱۳۹۲)، پایان‌نامه کارشناسی ارشد: "پیش‌بینی فعالیت ضد فشار خون

مهارکننده‌های غیر اوره‌ای هیدرولاز اپوکسید محلول"، دانشکده شیمی، دانشگاه ولی‌عصر (عج) رفسنجان.

[۳۳] شاکری س، (۱۳۸۹)، پایان‌نامه کارشناسی ارشد: "پیش‌بینی بلند مدت جریان رودخانه کارون

با روش ماشین بردار پشتیبان و استفاده از سیگنال‌های اقلیمی"، دانشکده مهندسی عمران، دانشگاه صنعتی شریف.

[۳۴] تحویلدارزاده ع، (۱۳۷۱)، پایان‌نامه کارشناسی ارشد: "مکانیک آماری شبکه‌های عصبی

مصنوعی در حضور چشمه خارجی"، دانشکده علوم فیزیک، دانشگاه صنعتی شریف.

[۳۵] رضایی م، (۱۳۸۹)، پایان‌نامه کارشناسی ارشد: "مطالعه 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.

Abstract

Activity coefficient at infinite dilution (γ^∞) is a key parameter which can be used for the selection of effective solvent in the separation processes. In this work, a quantitative structure-property relationship (QSPR) approach was employed to predict activity coefficients at infinite dilution for 60 organic compounds and water in the ionic liquid 1-butyl-1-methylmorpholinium tricyanomethanide ([BMMOR][TCM]) at six different temperature. A large number of descriptors were calculated using HyperChem and Dragon software and the best calculated descriptors was selected by stepwise regression (SR) and combined data splitting-feature selection (CDFS) strategy. A total of 12 molecular descriptors, as the most feasible ones, were selected using SR method. Also, 7 significant molecular descriptors and one experimental variable (temperature) were selected using CDFS method. The selected descriptors by two methods were used as input for artificial neural network (ANN) and support vector machine (SVM). After training and optimization of the ANN and SVM parameters, the performance of each model was evaluated using the test set. The mean square error (MSE) value for the test data set from SR-ANN, CDFS-ANN, SR-SVM and CDFS-SVM models were 0.0515, 0.1322, 0.0836 and 0.0988, respectively. The obtained results showed the superiority of the SR-ANN than other methods.

Keywords: activity coefficients at infinite dilution, Stepwise Regression (SR), combined data splitting-feature selection (CDFS), artificial neural network (ANN), support vector machine (SVM)



Shahrood University of Technology
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-butyl-1-methyl
morpholinium tricyanomethanide**

Hossein Nemati

Supervisor:

Dr. Z. Kalantar

Advisor:

Dr. N. Goudarzi

February 2016