



# بهینه‌سازی الگوریتم‌های یادگیری ماشین جهت تخمین تراوایی مخازن زیرزمینی نفت و گاز

مصیب کمبری<sup>۱</sup>، شرکت ملی مناطق نفت خیز جنوب

عباس سالخور<sup>۲</sup>، فریید بهیلی نجف آبادی<sup>۳</sup>، دانشگاه آزاد اسلامی

چکیده

خواص سنگ و سیال نقش بسیار مهمی در موفقیت مدیریت و توسعه‌ی مخازن نفت و گاز دارند. تراوایی<sup>۲</sup> (نفوذپذیری) از مهم‌ترین ویژگی‌های سازندهای هیدروکربنی است. تراوایی را از طریق اندازه‌گیری‌های مستقیم آزمایشگاهی روی مغزه‌های سنگ یا آزمایش چاه به‌دست می‌آورند. اما به‌دلیل وقت‌گیر و هزینه‌بر بودن و همچنین مشکلات عملیاتی، فقط تعداد محدودی از چاه‌ها مغزه‌گیری شده یا آزمایش چاه فقط به تعداد محدودی در میدان نفتی صورت می‌گیرد. بنابراین استفاده از روش‌های غیرمستقیم محاسباتی، اجتناب‌ناپذیر است. یکی از اهداف مطالعات پتروفیزیکی، تخمین دقیق و صحیح تراوایی در چاه‌هایی است که اندازه‌گیری تراوایی در آنها به‌هر دلیلی (نبود مغزه، وجود شکستگی در نمونه‌ها و ...) ممکن نیست.

در این مطالعه از الگوریتم رگرسیون ماشین بردار پشتیبان<sup>۴</sup>، کاربرد حالت‌های مختلف آن و الگوریتم رگرسیون ماشین بردار پشتیبان یکنواخت<sup>۵</sup> جهت تخمین تراوایی مخازن نفتی و همچنین از اطلاعات نگارهای چاه‌ها به‌عنوان ورودی الگوریتم‌ها استفاده شده است. نتایج این تحقیق نشان می‌دهد که اگر اطلاعات کمی و کیفی نسبتاً خوبی از داده‌های نگارهای چاه در اختیار باشد، می‌توان تراوایی را به‌خوبی و با ضریب خطای نسبتاً کمتری برای داده‌های جدید یا داده‌های دچار اشکال شده محاسبه کرد؛ علاوه بر این در بهترین حالت با آموزش ۸۰ درصد کل داده‌های موجود می‌توان با استفاده از رگرسیون ماشین بردار پشتیبان یکنواخت، تراوایی را با ضریب همبستگی ۹۶/۱ درصد و خطای میانگین مربعات ۰/۰۰۰۸ تخمین زد.

واژگان کلیدی تراوایی، الگوریتم یادگیری ماشین، رگرسیون ماشین بردار پشتیبان، تابع هسته، نمودار پتروفیزیکی، مغزه، چاه‌پیمایی

مقدمه

پشتیبان یکنواخت با تابع هسته‌ی پایه‌ی شعاعی) رویکرد جدیدی را برای تخمین تراوایی ارائه کرد.

یکی از متغیرهای مهم در مطالعات مخازن زیرزمینی نفت و گاز، تراوایی (نفوذپذیری) است که جهت مدیریت مخزن و فرآیندهای توسعه‌ی میادین نفت و گاز، اطلاع از مقدار آن اجتناب‌ناپذیر است. از لحاظ ماهیت و مشکلات عملیاتی، تخمین این متغیر نسبت به برخی متغیرهای دیگر سنگ و سیال، دشوارتر است. شیوه‌ای که در حال حاضر معمول است تعیین تراوایی و تخلخل مغزه در آزمایشگاه، به‌دست آوردن یک رگرسیون خطی بین این دو متغیر و تعمیم آن به کل چاه است که چندان مناسب نیست. تلاش‌های مختلفی جهت تخمین تراوایی صورت گرفته که می‌توان آن‌ها را به دو دسته‌ی روش‌های تجربی و آماری تقسیم کرد. روش‌های آماری، روش‌های نوینی مبتنی بر هوش مصنوعی، شبکه‌های عصبی، منطق فازی و... هستند که بسیار کارآمد ظاهر شده‌اند. در این مقایسه باید تعداد متغیرهای مخزن مدنظر، کیفیت اطلاعات آزمایشی انتخاب شده، تعداد نقاط نمونه‌برداری شده و همگن یا ناهمگن بودن مخزن که همگی بر دقت شبکه مؤثرند را در نظر گرفت. در این مقاله سعی شده با استفاده از حداکثر متغیرهای مرتبط و در دسترس (شامل نمودارهای پتروفیزیکی، طیف پرتو گاما، تخلخل، نوترون، چگالی سازند، صوتی، عامل فتوالکتریک، داده‌های تخلخل و درصد اشباع آب مغزه) به‌عنوان ورودی و همچنین استفاده از الگوریتم‌های یادگیرنده (رگرسیون بردار پشتیبان با کاربرد چهار نوع تابع هسته و رگرسیون بردار

## ۱- رگرسیون بردار پشتیبان

روش ماشین بردار پشتیبان از روش‌های یادگیری ماشینی است که بر مبنای تئوری یادگیری آماری vapnik در دهه‌ی ۹۰ میلادی توسط vapnik و همکاران ارائه شد. در ماشین بردار پشتیبان (SVM) از اصول کمینه‌کردن ریسک ساختاری (SRM) استفاده شده؛ در حالی که سایر روش‌ها از اصول کمینه‌کردن تجربی (ERM) بهره می‌برند [۱]. به‌طور کلی از ماشین بردار پشتیبان در مسائل طبقه‌بندی دو یا چند گروه و رگرسیون استفاده می‌شود. مانند بسیاری از روش‌های یادگیری ماشینی، در روش ماشین بردار پشتیبان نیز فرآیند ساخت مدل، شامل دو مرحله‌ی آموزش و آزمایش است [۲]. در انتهای فاز آموزش، قابلیت تعمیم مدل آموزش داده شده با استفاده از داده‌های آزمایشی ارزیابی می‌شود. به‌صورت خلاصه سازوکار اصلی SVM در حل مسأله‌ی رگرسیون به‌صورت زیر بیان می‌شود:

الف) ماشین بردار پشتیبان، تابع رگرسیون را با به‌کارگیری یک دسته تابع خطی تخمین می‌زند.

ب) ماشین بردار پشتیبان، عملیات رگرسیون را با تابعی که انحراف از مقدار واقعی در آن، کمتر از تابع تلفات تعیین شده است، انجام می‌دهد.

\*نویسنده‌ی عهده‌دار مکاتبات (salahshor.a@gmail.com)

ج) ماشین بردار پشتیبان با کمیته کردن ریسک ساختاری، بهترین جواب را ارائه می‌دهد [۲۳].

پیشنهاد داد. سپس یوان تابع چندجمله‌ای و تابع spline را پیشنهاد داد [۶]. در اینجا با یک مجموعه داده‌ی s شامل m نقطه در فضای حقیقی n بُعدی  $R^n$  که توسط ماتریس  $A \in R^{m \times n}$  ارائه شده و m مشاهده (پاسخ) از مقدار واقعی برای هر نقطه سروکار داریم که به شکل رابطه‌ی ۱- است [۶ و ۷]:

$$s = \{(A_i, y_i) | A_i \in R^n, y_i \in R, \text{ for } i = 1 \dots m\} \quad (1)$$

مابه‌دنبال پیدا کردن یک تابع رگرسیون خطی یا غیرخطی  $f(x)$  با تحمل خطایی کوچک در جفت‌سازی (ارتباط بین ورودی و خروجی) این مجموعه داده‌ها هستیم. این تابع را می‌توان با استفاده از تابع تلفات  $\epsilon$ -insensitive به دست آورد که به‌عنوان یک لوله<sup>۱</sup> اطراف داده‌ها قرار می‌گیرد و در آن اشتباهات دور انداخته شده است. همچنین با استفاده از ایده‌ی ماشین بردارهای پشتیبان، تابع  $f(x)$  به‌همان شکل مسطح یا یکنواخت ممکن در جفت کردن (تخمین) مجموعه داده‌های آموزشی ساخته شده است [۸ و ۹].

## ۲- روش کار

در این مطالعه ابتدا اطلاعات نمودارگیری و مغزه‌های مخزن آسماری در سه چاه A، B و C در یکی از میادین جنوب غربی ایران جمع‌آوری شد. سپس این داده‌ها و اطلاعات بر اساس متغیرهای ورودی (SW, POR, RHOB, DT, NPHI, GR, PEF, RT) و متغیر خروجی PERM پالایش و داده‌های اضافه حذف شد. سپس با استفاده از فرمول‌های مربوط به نرمالیزه کردن، داده‌های خارج از محدوده نیز حذف و در نهایت اطلاعات هر سه چاه در یک پایگاه داده (شامل ۱۲۲ نقطه) جمع‌آوری گردید.

همچنین برای تخمین تراوایی از نرم‌افزار متلب<sup>۲</sup> و جعبه‌ابزار<sup>۱</sup> مربوط به SVM استفاده شده که تمامی حالات و مقادیر ممکن و قابل آزمایش توسط برنامه‌نویسی و ایجاد حلقه‌های متناسب مورد بررسی و تحلیل قرار گرفته است؛ به‌نحوی که درصدهای مختلفی از اطلاعات را به‌صورت تصادفی انتخاب کرده (مثلاً ۲۰، ۴۰، ۶۰ و ۸۰ درصد)، آنها را توسط توابع هسته‌ی مختلف آموزش داده و سپس با درصد دیگری از اطلاعات

## ۱-۱- انواع توابع هسته

در مسائل دسته‌بندی، از یک تابع هسته در فضای ورودی استفاده می‌شود. مزیت اصلی استفاده از تابع هسته‌ی  $k(x_i, x_j)$  پرهیز از انجام نگاشت  $\Phi(x)$  به‌جای حاصل ضرب‌های عددی در فضای ویژگی  $\Phi^T(x_i)\Phi(x_j)$  و محاسبه توسط توابع هسته‌ی  $K(x_i, x_j)$  برای به‌دست آوردن بردارهای داده‌ی آموزشی در فضای ورودی است. به‌منظور کاهش حجم محاسبات ناشی از ازدیاد ابعاد، از ترفند هسته استفاده می‌شود. به کمک هسته، معادل ضرب داخلی دو بردار نگاشت شده به فضای ویژگی، بدون نیاز به نگاشت تک‌تک آن دو محاسبه می‌شود [۴ و ۵].

## ۱-۲- رگرسیون بردار پشتیبان یکنواخت (SSVM)

رگرسیون بردار پشتیبان یکنواخت، توسعه‌ای از SVM است که نخستین بار در ۲۰۰۱ توسط لی معرفی شد [۵]. ایده‌ی اولیه‌ی SSVM تبدیل فرمول اولیه‌ی SVM به یک مسأله‌ی حداقل کردن مشکل غیریکنواختی است. از آنجا که تابع هدف این مسأله‌ی بهینه‌سازی نمی‌تواند دو بار تشخیص‌پذیر باشد، تابع یکنواخت‌کننده می‌تواند جهت یکنواخت کردن این مشکل استفاده شود. لی انتگرال تابع سیگموئید را به‌عنوان تقریبی از تابع پلاس

۱   انواع توابع هسته‌ی پرکاربرد و معروف [۴]	
نوع تابع	عبارت ریاضی
خطی	$(u \cdot v)^n$
چندجمله‌ای	$(\gamma u \cdot v + C_{\text{coef}})^{\text{degree}}^n$
سیگموئیدی	$\tanh(\gamma u \cdot v + C_{\text{coef}})^n$
پایه‌ی شعاعی	$\exp(-\gamma  u - v ^2)^n$

۳   نتایج حاصل برای ضریب همبستگی و خطای میانگین مربعات به‌روش SVR با تابع هسته‌ی خطی					
R	Train set (%)	Test set (%)	C	Mse	ScC (%)
1	Tr40%	Te60%	6	۰/۰۲۹۷	۲۱/۰۶
2	Tr60%	Te40%	1	۰/۰۱۵۶	۲۸/۲۱
3	Tr60%	Te80%	9	۰/۰۲۶۷	۲۲/۲۲
4	Tr40%	Te20%	1	۰/۰۱۳۱	۳۶/۲۵
5	Tr40%	Te80%	6	۰/۰۲۰۵	۲۴/۶۳
6	Tr80%	Te60%	80	۰/۰۳۳۴	۱۷/۲۶
7	Tr80%	Te40%	1	۰/۰۱۵۶	۲۵/۷۱
8	Tr80%	Te20%	65	۰/۰۱۱۵	۴۹/۴

۲   محدوده‌ی مقادیر داده‌های استفاده شده				
ردیف	نوع داده	نام داده	واحد داده	محدوده‌ی داده‌ها
۱	ورودی	DT	Us/F	۴۰-۱۴۰
۲	ورودی	CGR	GPAI	۰-۱۰۰
۳	ورودی	PEF	--	۰-۱۰۰
۴	ورودی	RHOB	G/C3	۱/۹۵-۲/۹۵
۵	ورودی	NPHI	V/V	(-۰/۱۵)-۰/۴۵
۶	ورودی	RT	OHMM	۰/۲-۲۰۰۰۰۰
۷	ورودی	Por	%	۰-۱۰۰
۸	ورودی	SW	%	۰-۱۰۰
۹	خروجی	Perm	Md	۰-۱۰۰۰۰



کل داده‌ها و  $x_{max}$  حداکثر مقدار داده‌هاست.

۲-۲- روش ارزیابی استفاده شده

نتایج حاصل از الگوریتم‌های مورد استفاده از لحاظ دقت، صحت، مقدار و کیفیت بررسی شدند و بهترین حالت برای مدل‌سازی انتخاب گردید. برای این منظور، الگوریتم‌های مختلف با ملاک‌های رابطه‌ی ۳- ارزیابی شدند:

$$MSE = \frac{\sum(\text{Actual}-\text{Forecast})^2}{n} = \frac{\sum(\text{Error})^2}{n} \quad (3)$$

رابطه‌ی ۳-، خطای میانگین مربعات است که در آن  $n$ ، تعداد نمونه‌های تخمینی، Actual، مقدار واقعی هر برجسب و Forecast، مقدار تخمینی برای هر یک است. هر چه مقدار این کمیت کمتر باشد، ضریب خطا نیز کمتر خواهد بود. ضریب همبستگی، شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) را نشان می‌دهد. این ضریب بین ۱ تا -۱ است و در حالت عدم وجود رابطه بین دو متغیر، برابر صفر می‌باشد.

همبستگی بین دو متغیر تصادفی  $X$  و  $Y$  به صورت رابطه‌ی ۴- تعریف می‌شود:

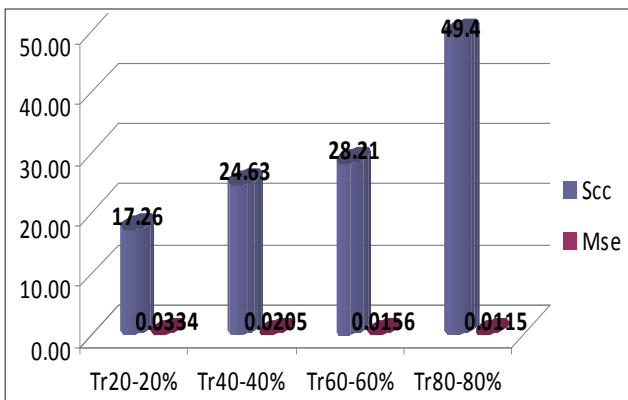
که آنها نیز به صورت تصادفی انتخاب شده‌اند به عنوان داده‌های آزمایشی ارزیابی و مقایسه گردید. در انتخاب تصادفی اطلاعات از نرم‌افزار MS SQLSERVER 2008 و زبان برنامه‌نویسی TSQL استفاده شد. خروجی تابع پیش‌بینی دارای دو معیار محاسبه‌ی خطا به نام‌های ضریب همبستگی مربعات<sup>۱۱</sup> و خطای میانگین مربعات<sup>۱۲</sup> است که در تمامی مراحل بعدی از آنها برای مقایسه و رسیدن به جواب بهتر استفاده می‌شود.

۲-۱- انتخاب روش استانداردسازی داده‌ها

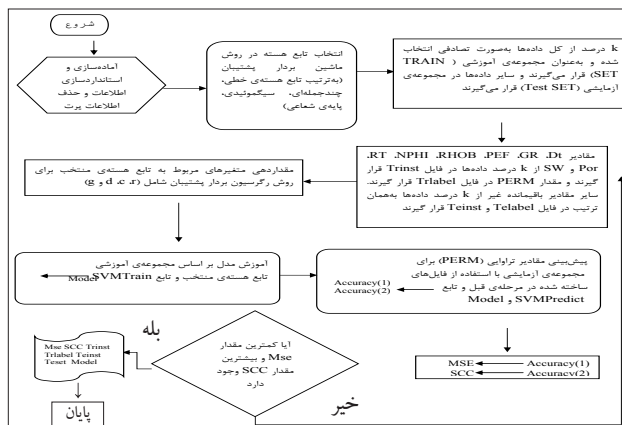
روش‌های مختلفی برای استانداردسازی وجود دارد که در تمامی آنها داده‌ها طوری تبدیل می‌شوند که شرایط مدنظر تأمین گردد. یکی از مهم‌ترین روش‌های استانداردسازی، تبدیل داده‌ها به مجموعه‌ای جدید است که در آن، همه‌ی مقادیر بین صفر و یک باشند:

$$Z_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

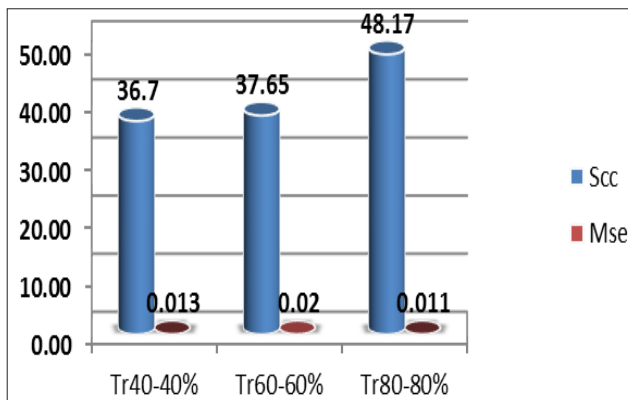
که در آن  $Z_i$  مقدار استاندارد شده،  $X_i$  مقدار هر یک از داده‌ها،  $x_{min}$  حداقل



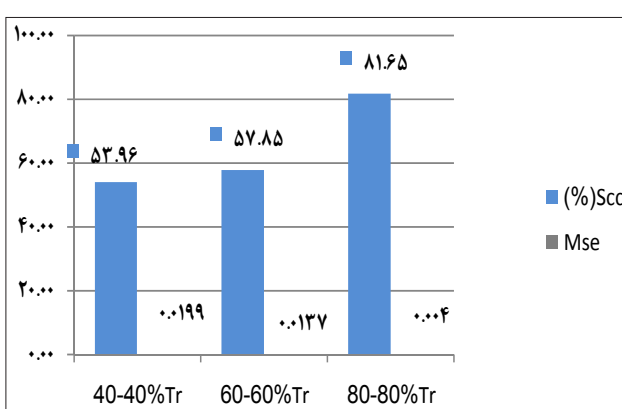
شکل ۲ | مقایسه‌ی ضریب همبستگی و خطای میانگین مربعات با استفاده از مجموعه داده‌های مختلف پس از اجرای رگرسیون بردار پشتیبان با تابع هسته‌ی خطی



شکل ۱ | نمودار گردش کار تخمین تراوایی توسط SVR



شکل ۳ | مقایسه‌ی ضریب همبستگی و خطای میانگین مربعات با استفاده از مجموعه داده‌های مختلف پس از اجرای رگرسیون بردار پشتیبان با تابع هسته‌ی سیگموئیدی



شکل ۳ | مقایسه‌ی ضریب همبستگی با استفاده از مجموعه داده‌های انتخابی مختلف پس از اجرای رگرسیون بردار پشتیبان با تابع هسته‌ی چندجمله‌ای

مخزن، منجر به تنوع کیفی در آنها شده که می‌توان بعداً با اطمینان بیشتری مدل را برای سایر چاه‌های مخزن استفاده کرد. در هر نمونه‌ی آزمایشی و آموزشی، دو مجموعه داده وجود دارد که در نهایت به صورت رابطه‌ی ۵- دسته‌بندی خواهد شد:

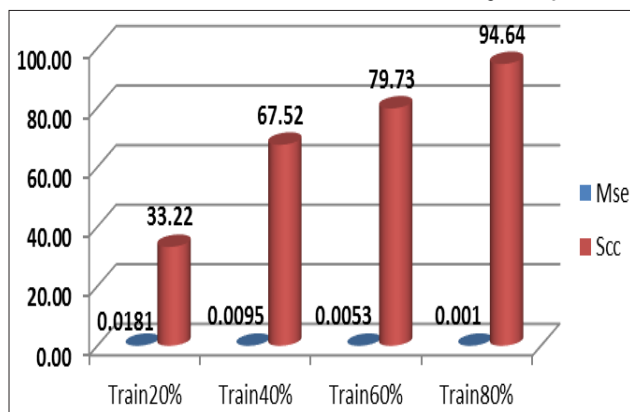
$$\{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{R}, y \in \mathbb{R} \quad (5)$$

که در آن:

$$X_i = (DT, CGR, PEF, RHOB, NPHI, RT, POR)$$

$$Y_i = Perm$$

بدین ترتیب که تراوایی به‌عنوان متغیر خروجی و مقادیر داده‌های پتروفیزیکی به‌عنوان بردار ورودی تعریف شده‌اند. در ابتدای روند انتخاب مدل، جهت آموزش و ساخت مدل در یک مرحله، داده‌های آموزشی به صورت کاملاً تصادفی انتخاب شدند و آزمایش مدل یک‌بار روی داده‌های آزمایشی بدون انتخاب تصادفی آنها انجام گردید. در مرحله‌ی دیگر، آزمایش مجدد روی داده‌ها در حالت انتخاب تصادفی مجموعه‌ی آزمایشی انجام شد که در نهایت حالت انتخاب تصادفی نتایج بهتری در ساخت مدل داشت. چراکه دامنه‌ی بیشتری از انطباق مقدار تراوایی تخمینی و مقدار واقعی تراوایی را در برمی‌گرفت. به‌طور کلی این مجموعه عملیات در چهار گام اصلی انجام شد که از گام اول تا چهارم به ترتیب ۲۰، ۴۰، ۶۰ و ۸۰ درصد از کل داده‌ها برای آموزش انتخاب شدند. نکته‌ی قابل توجه این‌که در اکثر تحقیقات انجام شده در خصوص الگوریتم‌های یادگیری ماشین، پس از ساخت مدل آموزش دیده ابتدا همان داده‌هایی که منجر به ساخت مدل شده را برای آزمایش روی مدل و گرفتن خروجی آزمایش می‌کنند تا کارایی مدل را در تحلیل خودش بررسی نمایند. این کار در مطالعه‌ی جاری نیز انجام شد که کاملاً موفقیت‌آمیز بود. بنابراین حتی اگر در انتخاب مجموعه داده‌های آزمایشی، داده‌های تکرارپذیری وجود داشته باشند از لحاظ صحت و استفاده‌ی مجدد، هیچ اشکالی در تحلیل و کارایی الگوریتم استفاده شده ایجاد نخواهد کرد.



شکل ۵ | مقایسه‌ی ضریب همبستگی و خطای میانگین مربعات با استفاده از مجموعه داده‌های مختلف پس از اجرای رگرسیون بردار پشتیبان با تابع هسته‌ی پایمی شعاعی

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (4)$$

که در آن E عملگر امید ریاضی، cov به معنای کوواریانس، corr نماد معمول برای همبستگی پیرسون و  $\sigma$  نماد انحراف معیار است.

### ۳-۲- تقسیم‌بندی و انتخاب داده‌ها

با توجه به آنکه قرار است از الگوریتم‌های یادگیری ماشین استفاده گردد و در تمامی الگوریتم‌های یادگیرنده، ابتدا باید از روی یک مجموعه داده‌ی اولیه‌ی آموزش دیده، مدلی ساخته شود، بنابراین تمامی داده‌ها به دو دسته‌ی آموزش و آزمایش تقسیم می‌شوند. این مجموعه‌ها (آموزشی و آزمایشی) به صورت تصادفی از روی داده‌های نمودارهای مربوط به سه چاه انتخاب شدند. انتخاب داده‌های تصادفی از سه چاه و به تبع آن سه محدوده‌ی مختلف

نتایج حاصل برای ضریب همبستگی و خطای میانگین مربعات به روش SVR با تابع هسته‌ی چندجمله‌ای

R	Train set(%)	Test set(%)	R	d	C	Mse	Sc (%)
1	Tr40%	Te40%	۱	۵	۱۰۰	۰/۰۱۹۹	۵۳/۹۶
2	Tr60%	Te60%	۰/۵	۵	۹۹	۰/۰۱۱۱	۴۹/۸۷
3	Tr60%	Te20%	۱	۵	۱۰۰	۰/۰۱۳۷	۵۷/۸۵
4	Tr60%	Te80%	۱	۵	۱۰۰	۰/۰۲۱۸	۳۹/۷۶
5	Tr40%	Te20%	۱	۵	۵۷	۰/۰۰۶۴	۷۰/۳۴
6	Tr40%	Te80%	۱	۵	۱۰۰	۰/۰۱۵۹	۴۹/۴۸
7	Tr80%	Te60%	۰/۴	۵	۶۸	۰/۰۱۱۹	۴۴/۶۸
8	Tr80%	Te20%	۱	۵	۱۰۰	۰/۰۰۰۴	۸۱/۶۵
9	Tr80%	Te40%	۱	۵	۱۰۰	۰/۰۱۲۷	۶۹/۲۹

نتایج حاصل برای ضریب همبستگی و خطای میانگین مربعات به روش SVR با تابع هسته‌ی سیگموئیدی

R	Train set%	Test set%	C	G	R	Mse	Sc (%)
1	Tr40	Te60	۱۰۰	۰/۱۲۲	۱	۰/۰۱۵	۲۰/۱۶
2	Tr60	Te40	۱۰۰	۰/۰۱۱	۱	۰/۰۱۶	۲۶/۱۵
3	Tr60	Te20	۱	۰/۰۱۳	۲	۰/۰۰۲	۳۷/۶۵
4	Tr40	Te20	۱۰۰	۰/۰۰۸	۰/۰۰۰۱	۰/۰۱۳	۳۶/۷
5	Tr40	Te80	۱۰۰	۰/۰۰۸	۰/۰۰۰۱	۰/۰۰۲	۲۴/۴۵
6	Tr80	Te60	۱	۰/۰۰۱	۰/۰۰۰۱	۰/۰۱۷	۲۶/۳۳
7	Tr80	Te20	۱۸۲	۰/۰۰۱	۰/۰۰۰۱	۰/۰۱۲	۴۸/۱۷
8	Tr80	Te40	۱۸۲	۰/۰۰۸	۰/۱	۰/۰۳۳	۲۰/۴۲
9	Tr60	Te60	۵۰۰	۰/۰۰۸	۱	۰/۰۰۲۸	۲۴/۹۵
10	Tr40	Te60	۱	۰/۰۱۳	۰/۰۰۰۱	۰/۰۱۸	۲۱/۰۶



۴-۲- آموزش و ساخت مدل

در این مرحله، در آموزش مدل از ماشین بردار پشتیبان، الگوریتم‌های وابسته و داده‌های چاه‌نگاری و رخساره‌های مشخص شده برای هر نقطه از عمق مخزن توسط گزارش توصیف مغزه استفاده شد. با توجه به اینکه در مطالعه‌ی حاضر از دو نوع الگوریتم مذکور استفاده شده و خود رگرسیون بردار پشتیبان از چهار نوع تابع هسته (خطی، چند جمله‌ای، سیگموئیدی، پایه‌ی شعاعی) استفاده می‌کند (که منجر به استفاده از متغیرهای مختلف برای توابع مختلف شده)، در این قسمت هدف یافتن متغیرهای بهینه‌ی مربوط به هر الگوریتم است که بتواند در حین مواجهه با داده‌هایی که مدل بر اساس آنها ساخته نشده (داده‌های آزمایشی)، بهترین عملکرد را داشته باشند. به همین دلیل برای تشخیص عملکرد مدل از شاخص‌های بیشترین ضریب همبستگی و کمترین مقدار خطای میانگین مربعات در هر مرحله استفاده شده است (شکل-۱).

۳- بررسی نتایج

در این بخش، نتایج و نمودارهای حاصل از اجرای الگوریتم SVR با استفاده از توابع هسته‌ی مختلف و همچنین الگوریتم SSVR، با تابع هسته‌ی پایه‌ی شعاعی و اجراهای متنوع، جهت مقایسه‌ی عملکرد آنها ارائه می‌شود.

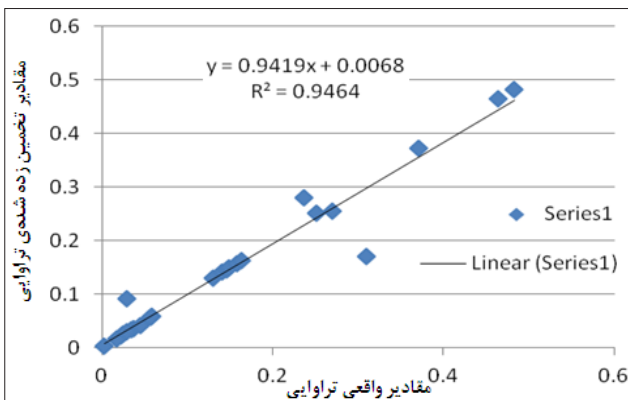
۳-۱- نتایج حاصل از اجرای رگرسیون بردار پشتیبان با استفاده از تابع هسته‌ی خطی (SVRL)

طبق جدول ۳- در ستون Trainset، مجموعه داده‌های آموزشی در ستون Testset، داده‌های آزمایش و اعتبارسنجی (که درصدی از کل داده‌هاست) انتخاب شده که با تغییر مقدار C (متغیر هزینه (وزن) که مقدار پیش فرض آن برای تمامی روش‌ها یک است) [۸] تابع هسته، نتایج متفاوتی برای تخمین تراوایی به دست آمده است. در ستون MSE، کمترین مقدار خطا با بهترین ضریب همبستگی در ستون SCC قرار دارد. این نتایج نشان می‌دهد همان‌طور که انتظار می‌رفت بهترین مقادیر تخمینی در استفاده از الگوریتم SVR با تابع هسته‌ی خطی در حالت آموزش مجموعه‌ی ۸۰ درصدی از داده‌هاست (مرحله‌ی ۸-) که دارای بهترین ضریب همبستگی (۴۹/۴ درصد) و کمترین خطای میانگین مربعات ۰/۱۱۵ می‌باشد. متغیر C تعیین شده برای این مجموعه‌ی آموزشی ۶۵ می‌باشد؛ لازم به ذکر است که با انتخاب مقدار مذکور برای C، همیشه این اتفاق نمی‌افتد؛ بلکه این مقدار نیازمند انتخاب نمونه‌های آموزشی نیز هست و نمونه‌های آموزشی به همراه تغییرات دامنه‌ی متغیر C با هم تعیین کننده‌ی مقدار تخمینی بهینه می‌شود.

۳-۲- نتایج حاصل از اجرای رگرسیون بردار پشتیبان با استفاده از تابع هسته‌ی چندجمله‌ای (SVRP)

در جدول ۴-، مقادیر ستون d نمایانگر آن است که تابع چندجمله‌ای در درجه‌ی پنج تخمین بهتری داشته است. همچنین بهترین مقادیر برای متغیر C (هزینه)، محدودی ۵۰ تا ۱۰۰ و بهترین مقادیر برای متغیر r (ضریب صفر) محدودی ۰/۴ تا ۱ می‌باشد.

تخمین مربوط به مجموعه‌ی آموزشی ۸۰ درصدی با مجموعه‌ی آزمایش



شکل ۶ ضریب همبستگی تراوایی تخمینی با مقدار واقعی ۸۰ درصد آموزش و ۲۰ درصد آزمایش در روش SVR

۶ نتایج حاصل برای ضریب همبستگی و خطای میانگین مربعات به روش SVR با تابع هسته‌ی پایه‌ی شعاعی

R	Train set%	Test set%	Gamma	C	Mse	Scce (%)
1	Tr60	Te60	۳۶/۰۳۳۴	۱	۰/۰۰۶۱	۷۱/۲۸
2	Tr40	Te40	۳۶/۰۳۳۴	۱	۰/۰۰۷۱	۸۰/۶۹
3	Tr60	Te80	۳۶/۰۳۳۴	۱	۰/۰۱۴۲	۵۱/۵۵
4	Tr40	Te20	۳۶/۰۳۳۴	۱	۰/۰۰۹۳	۵۵/۴
5	Tr40	Te80	۳۶/۰۳۳۴	۱	۰/۰۰۹۵	۶۷/۵۲
6	Tr20	Te60	۳۶/۰۳۳۴	۱	۰/۰۱۹۳	۳۲/۸۵
7	Tr20	Te80	۳۶/۰۳۳۴	۱	۰/۰۱۸۱	۳۲/۲۲
8	Tr60	Te20	۳۶/۰۳۳۴	۱	۰/۰۰۵۳	۷۹/۷۳
9	Tr80	Te40	۳۶/۰۳۳۴	۱	۰/۰۰۱۹	۹۴/۵۷
10	Tr80	Te60	۳۶/۰۳۳۴	۱	۰/۰۰۷۱	۶۵/۱۵
11	Tr80	Te80	۳۶/۰۳۳۴	۱	۰/۰۰۶۳	۷۸/۲۳
12	Tr80	Te20	۳۶/۰۳۳۴	۱	۰/۰۰۱	۹۴/۶۴

۷ نتایج حاصل برای ضریب همبستگی و خطای میانگین مربعات به روش SSVR با تابع هسته‌ی پایه‌ی شعاعی

R	Train set (%)	Testset(%)	C	g	r	E	Mse	Scce (%)
1	Tr60%	Te40%	۱۰۰	۱۲/۴	۱	۰/۰۱	۰/۰۲۲	۴۱/۸
2	Tr60%	Te20%	۷/۶۷	۵/۷۳	۱	۰/۰۰۰۱	۰/۰۸۲	۵۹
3	Tr40%	Te20%	۱۹۰۰	۳/۵۱	۱	۰/۰۰۰۱	۰/۰۰۲	۸۸
4	Tr40%	Te80%	۲۱	۲۸	۱	۰/۰۰۰۱	۰/۰۰۶	۷۶/۶
5	Tr80%	Te40%	۱۰۰۰	۲۲/۸۶	۱	۰/۰۰۰۰۱	۰/۰۰۱	۹۵/۸
6	Tr80%	Te20%	۳۰۰	۳/۶۶	۱	۰/۰۰۰۰۱	۰/۰۰۰۸	۹۶/۱
7	Tr40%	Te60%	۵	۹۶/۴	۱	۰/۰۰۰۰۱	۰/۰۱۰۶	۴۸/۴

(معادل ۰/۱ درصد) است. (شکل-۶).

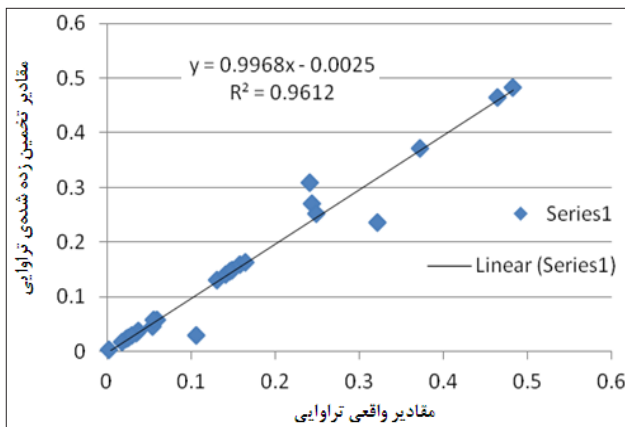
همچنین بدترین حالت در نمونه‌ی آموزشی ۲۰ درصدی و آزمایش ۶۰ درصد داده که ضریب همبستگی ۳۲/۸۵ درصد بوده، باز هم نسبت به روش‌های قبلی بهتر به نظر می‌رسد. علاوه بر این، با رشد مجموعه داده‌ی آموزشی نتایج بهبود یافته است. لازم به ذکر است که در همه‌ی حالات، مقدار متغیرهای  $C=1$  و  $g=36/0334$  ثابت در نظر گرفته شده که البته این مقادیر در حالت بهینه حاصل شده و برای سایر حالات و مجموعه‌های آموزشی و ساخت مدل‌ها توسعه پیدا کرده است.

### ۳-۵- نتایج حاصل از اجرای الگوریتم رگرسیون بردار پشتیبان یکنواخت (SSVR)

مقادیر حاصل برای ضریب همبستگی و خطای میانگین مربعات در جدول ۷- نشان می‌دهد که روش SSVR عملکرد مناسبی برای پیش‌بینی تراوایی دارد. بهترین مقدار برای ضریب همبستگی ۹۶/۱ درصد به دست آمده که حالت بهینه است (شکل-۷). در این روش نیز بهترین حالات وقتی رخ می‌دهد که مجموعه‌ی آموزش تعداد بیشتر و متنوع‌تری از داده‌ها را در برداشته باشد. مقادیر متغیر  $C$  دامنه‌ی تغییرات زیادی داشته و با تغییر محدوده‌ی داده‌ها در مجموعه‌ی آموزش، نوسانات آن نیز غیرقابل پیش‌بینی است. بنابراین در تعیین مقدار متغیر مذکور اغلب دامنه‌ها آزمایش و بررسی شد. دامنه‌ی نوسانات متغیر  $g$  بین ۱ تا ۲۰۰ بهترین مقدار را می‌پذیرد و متغیرهای  $T$  و  $C$  نیز تقریباً با مقادیر ۱ و ۰/۰۰۰۰۱ بهترین حالات را دارند.

### ۳-۶- مقایسه‌ی نتایج حاصل از اجرای حالات مختلف روش‌های SVR و SVR جهت پیش‌بینی تراوایی

با مقایسه و ارزیابی الگوریتم‌های مورد بحث در این مطالعه، حالت بهینه برای هدف تعریف شده مشخص می‌گردد. بهترین نتایج حاصل از اجرای هر کدام از روش‌های مختلف استفاده شده در جدول ۸- ارائه گردیده است. نتایج حاصل نشان می‌دهد که بهترین ترتیب الگوریتم‌های استفاده شده بر



۸ ضریب همبستگی تراوایی تخمینی با مقدار واقعی ۸۰ درصد آموزش و ۲۰ درصد آزمایش در روش SSVR

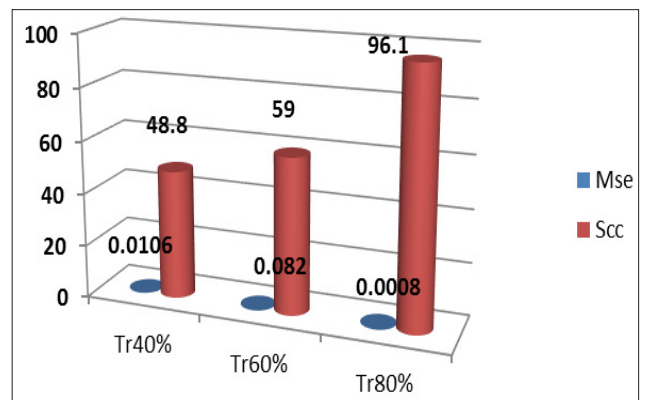
۲۰ درصدی داده‌هاست (مرحله‌ی ۸-) که مقادیر متغیرهای  $T$ ،  $d$  و  $C$  به ترتیب ۱ و ۵ و ۱۰ تنظیم شده‌اند. این حالت، بیشترین ضریب همبستگی ۸۱/۶۵ درصد و کمترین خطای میانگین مربعات ۰/۰۰۴ را دارد. بررسی جدول-۴ نشان می‌دهد که در این روش، اکثر حالات بهینه‌ی حاصل در مقدار متغیرهای  $T$  و  $d$  یکسان بوده و تقریباً بهترین مقدار  $C$  برای حالات بهینه ۱۰۰ خواهد بود. همچنین با افزایش تعداد نمونه‌های آموزشی نتایج بهتری به دست آمده است.

### ۳-۳- نتایج حاصل از اجرای رگرسیون بردار پشتیبان با استفاده از تابع هسته‌ی سیگموئیدی (SSVR)

نتایج حاصل از جدول ۵- نشان می‌دهد که بهینه‌ترین تخمین برای مجموعه‌ی داده‌ی آموزشی ۸۰ درصدی با مجموعه‌ی آموزشی ۲۰ درصدی است (مرحله‌ی ۷-) که در آن مقادیر متغیرهای  $C$ ،  $g$  و  $T$  به ترتیب ۱۸۲ و ۰/۰۱ و ۰/۰۱ تنظیم شده است. بهترین ضریب همبستگی برای این مرحله ۴۸/۱۷ درصد و کمترین خطای میانگین مربعات ۰/۰۱۱۷ می‌باشد. لازم به ذکر است متغیر  $g$ ، در محدوده‌ی [۰-۱]، متغیر  $C$  (هزینه)، در بازه‌ی [۲۰۰-۱۰۰] و متغیر  $T$  (ضریب صفر)، در محدوده‌ی [۰-۲] نتایج بهتری به دست می‌دهند. در مجموع با توجه به انتخاب مجموعه داده‌ی آموزشی مختلف، نوسان دقت تخمین با استفاده از روش SVR و تابع هسته‌ی سیگموئیدی زیاد است. خطای میانگین مربعات نیز از حدود ۱/۱ تا ۳/۱ درصد به دست آمد.

### ۳-۴- نتایج حاصل از اجرای رگرسیون بردار پشتیبان با استفاده از تابع هسته‌ی پایه‌ی شعاعی (SSVR)

با توجه به جدول ۶- مشاهده می‌شود که نتایج پس از اجرای روش SVR با تابع هسته‌ی پایه‌ی شعاعی (RBF) به دست آمده است. بر اساس نتایج حاصل بهینه‌ترین تخمین با ضریب همبستگی ۹۴/۶۴ درصد برای مجموعه‌ی داده‌ی آموزشی ۸۰ درصدی با مجموعه‌ی آزمایشی ۲۰ درصدی (مرحله‌ی ۱۲-) بوده که کمترین مقدار خطای میانگین مربعات این مرحله برابر ۰/۰۰۱



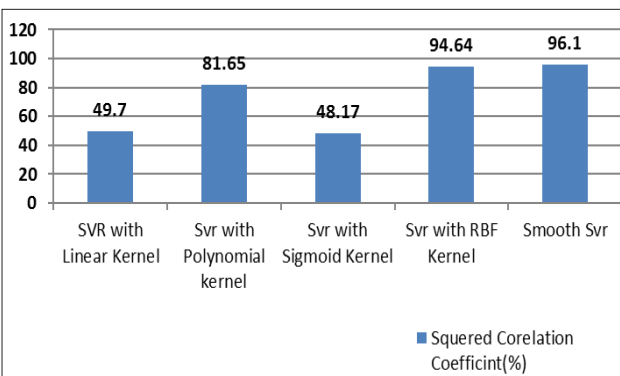
۷ مقایسه‌ی ضریب همبستگی و خطای میانگین مربعات با استفاده از مجموعه داده‌های مختلف پس از اجرای رگرسیون بردار پشتیبان یکنواخت با تابع هسته‌ی پایه‌ی شعاعی



SVRL، SVRP، و SVRS نیز به ترتیب با ضرایب همبستگی ۰/۴۷۹ و ۰/۸۱۶ و ۰/۴۸۲ قادر به پیش‌بینی تراوایی سازند برای این مجموعه از داده‌ها بودند.

۸ نتایج ضریب همبستگی و خطای میانگین مربعات با استفاده از روش‌های SVR و SSVR

SVM TYPE	TRAIN (%)SET	TEST (%)SET	MSE	(%)SCC
SVR with Linear Kernel	Tr80	Tr20	۰/۰۱۱۵	۴۹/۷
SVR with Polynomial Kernel	Tr80	Tr20	۰/۰۰۰۴	۸۱/۶۵
SVR with Sigmoid Kernel	Tr80	Tr20	۰/۰۱۱۷	۴۸/۱۷
SVR with RBF Kernel	Tr80	Tr20	۰/۰۰۰۱	۹۴/۶۴
Smooth SVR	Tr80	Tr20	۰/۰۰۰۰۸	۹۶/۱



۹ مقایسه‌ی ضریب همبستگی روش‌های رگرسیون بردار پشتیبان با توابع هسته (خطی، پایه‌ی شعاعی و چندجمله‌ای) و رگرسیون بردار پشتیبان یکنواخت

اساس مقدار ضریب همبستگی و بهینه‌ترین مقادیر تخمینی با کمترین خطای میانگین مربعات عبارتند از: SVRL، SVRP، SVRR، SSVR و SVRS. بر اساس داده‌های مطالعه، نتایج حاصل جهت تخمین تراوایی مخزن، روش SSVR را پیشنهاد می‌دهند. البته این موضوع دلیلی بر مطلق بودن و عمومیت این نتیجه (یعنی برتری روش SSVR بر سایر روش‌ها) نیست و بنابراین در مطالعه‌ی هر مخزن، قبل از هر چیز باید جهت انتخاب روش برتر، الگوریتم‌ها ارزیابی و بهینه‌سازی شوند.

### نتیجه‌گیری

تراوایی (نفوذپذیری) سازند، از مهم‌ترین خواص سنگ در مطالعه و ارزیابی مخازن زیرزمینی نفت و گاز است. در این مطالعه رویکرد جدیدی جهت تخمین تراوایی با استفاده از حالات مختلف الگوریتم ماشین بردار پشتیبان معرفی شده است. برای این منظور، الگوریتم‌های SVRL، SSVR، SVRP، SVRR، SVRS اجرا گردید.

بهترین جواب برای تخمین تراوایی، ضریب همبستگی ۹۴/۶۴ برای روش SVR با تابع هسته‌ی RBF و ضریب همبستگی ۹۶/۱ با روش SSVR به دست آمد. لازم به ذکر است که موفقیت هر مدل، کاملاً تابع کیفیت داده‌ها و شرایط مخزن مورد مطالعه است. بنابراین طبیعتاً در سایر موارد ممکن است نتایج متفاوتی مشاهده شود.

در تمامی روش‌ها از مجموعه داده‌ای یکسان شامل سری کامل داده‌های پتروفیزیکی (نمودارهای مقاومتی، صوت، هسته‌ای، تخلخل و اشباع) مخزن آماری در یکی از میادین جنوب غرب ایران استفاده شد. مقادیر واقعی تراوایی حاصل از اندازه‌گیری‌های آزمایشگاهی روی نمونه‌های مغزه‌ی سنگ هستند. بر اساس بهینه‌سازی انجام شده، روش‌های

### پانویس‌ها

<sup>1</sup>farsh.s@gmail.com

<sup>2</sup>mosayyeb.kamari@gmail.com

<sup>3</sup>Permeability

<sup>4</sup>Support vector regression(SVR)

<sup>5</sup>Smooth support vector regression (SSVR)

<sup>6</sup>Mapping

<sup>7</sup>Feature space

<sup>8</sup>Tube

<sup>9</sup>Matlab Software

<sup>10</sup>ToolBox

<sup>11</sup> Squared correlation coefficient (Scc)

<sup>12</sup>Mean Squared Error (Mse)

### منابع

- [1] Siamak Nazari, Heidi A. Kuzma, 2011, "Predicting Permeability From Well Log Data And Core Measurements Using Support Vector Machines", SEG San Antonio 2011 Annual Meeting, Pp 20042008-.
- [2] Shiloh Shah. Rohan, 2007, "Support Vector Machines For Classification And Regression ", A Thesis Submitted To McGill University Montreal, 1086-.
- [3] Cristianini Nello., Shawe Taylor John., 2003, "An Introduction To Support Vector Machines And Other Kernel-Based Learning Methods", Cambridge University Press, Chapter 6.
- [4] Te-Ming Hung, Vojislav Kecman, Ivica Kopriva, 2006, "Kernel Base Algorithms For Mining Huge Data Set", Springer, Chapter 2.
- [5] Ethem Alpaydn, 2010, "Introduction To Machine Learning Second Edition", The MIT Press Cambridge, Pp 21361-.
- [6] Bishop.M, Octobr 2007, "Pattern Recognition And Machine Learning", Springer, Vol 3 Pp 137154-, Vol 6 Pp 291311-, Vol 7 326356-.
- [7] Yuh-Jye Lee, Wen-Feng Hsieh, And Chien-Ming Huang, May 2005, "-SSVR: A Smooth Support Vector Machine For -Insensitive Regression ", Ieee Transactions On Knowledge And Data Engineering, Pp 678681-.
- [8] Vanderlooy.S, 12 November 2008, "Matlab Toolbox For Machine Learning", Maastricht University Of The Netherlands Technical Report 0803-, Vol 6.2.
- [9] Chih-Chung Chang, Chih-Jen Lin, March 4, 2013, "LIBSVM: A Library For Support Vector Machines", Department Of Computer Science National Taiwan University, Pp 135-.