Journal of Statistical Sciences, Autumn and Winter, 2023
Vol. 17, No. 2, pp 459-476
DOI: 10.61186/jss.17.2.4

A New Approach in Using Random Support Vector Machine Cluster in Analyzing Prostate Cancer Gene Expression Data

Moussavi, N., Golalizadeh, M.
Department of Statistics, Tarbiat Modares University, Tehran, Iran.

Corresponding author: N. Moussavi, niliamusavi2010@gmail.com

Received: 1/1/2023 **Revised:** 24/12/2023 **Accepted and Published Online:** 26/12/2023.

Introduction

Many statistical data analysis methods can help evaluate cancer progression among patients by creating a set of gene markers. However, one of the main problems in the statistical study of this type of data is the large number of genes versus the small number of samples. The situation is known as "big p and small n" among the scientific communities. Consequently, one should utilize some dimensionality reduction techniques for proper statistical analysis. One essential purpose of studying the gene data is to find the optimal number of genes to predict the desired classes accurately. Many machine learning tools were provided, so choosing an appropriate method is critical to providing an efficient statistical model. Support vector machine is a valuable technique to classify complex data such as gene expressions for prostate cancer. A new modified version of this tool called the random support vector machine cluster has been introduced in the machine learning communities. It is an ensemble learning approach and suitable to find the optimal feature set. The primary rationale of this technique is randomly projecting the original high-dimensional feature space onto multiple lower-dimensional feature subspaces and combining support vector machine classifiers. This paper will highlight the procedure for implementing this technique. It is shown that the main outcome of applying this tool to analyze the gene expression data for prostate cancer is twofold. It gives us not only the important genes but also a high level of classification precision.

Material and Methods

We use a Random Subsample Ensemble (RSE) to overcome the problem

caused by treating the high dimensional data. It is a variable selection based on the learning ensemble. Then, we utilize the Random Support Vector Machine Cluster (RSVMC) to classify the data and select the set of optimal variables. Note that we should repeat selecting the necessary variable procedure in invoking the SVM to allow the randomness of SVMC. To evaluate the model, we divide the data set into three common parts, i.e., training, validation, and testing samples. Moreover, we use the sigmoid kernel during the fitting step. We consider the accuracy, sensitivity, and specificity measures to showcase the model's superiority.

Results and Discussion
Our results, implemented on the prostate cancer data, show the RSVMC was able to identify thirteen patients with prostate cancer correctly. However, it made the mistake of recognizing two persons with having the disease while they did not have it. Regarding accuracy, sensitivity, and specificity measures, our method reached about ninety-three hundred and eighty-eight percent values, respectively. There is still room to implement our approach in the multi-class classification problem and compare it with other variable selection techniques, such as the regularization strategy.

Conclusion
The new idea presented in this paper, i.e., RSVMC, is a powerful tool to select an optimal subset of the optimal variable and then use it in a classification problem by invoking the support vector machine technique. Such a strategy will lead to the high efficiency of the model as well as provide a smooth and relevant interpretation of the essential genes. Moreover, it gains a high actual positive rate, leading to correctly identifying the patients who have prostate cancer.

Keywords: Ensemble learning, Dimensionality reduction, Classification, Random support vector machine cluster, Optimal feature set.

Mathematics Subject Classification (2010): 62H12, 62H30.

©The Author(s). The Publisher is Iranian Statistical Society.
This is an open access article distributed under the terms and conditions of (CC BY-NC 4.0)

مجله علوم آماری، پاییز و زمستان ۱۴۰۲
جلد ۱۷، شماره ۲، ص ۴۵۹ -- ۴۷۶
DOI: 10.61186/jss.17.2.4
مقاله پژوهشی

رویکردی نوین در بکارگیری روش دسته ماشین بردار پشتیبان تصادفی در تحلیل داده‌های بیان ژن سرطان پروستات

موسوی، ن. و گلعلی‌زاده، م.

نویسنده مسئول : نیلیا موسوی، niliamusavi2010@gmail.com

تاریخ دریافت: ۱۴۰۱/۱۰/۱۱ تاریخ بازنگری: ۱۴۰۲/۱۰/۳ تاریخ پذیرش و انتشار: ۱۴۰۲/۱۰/۵

گروه آمار، دانشگاه تربیت مدرس.

چکیده: پیشرفت سرطان در بین بیماران را می‌توان از طریق ایجاد مجموعه‌ای از نشانگرهای ژن با روش‌های تحلیل آماری داده‌ها بررسی کرد. اما یکی از مشکلات اساسی در مطالعه آماری این نوع داده‌ها وجود تعداد زیاد ژن‌ها در مقابل تعداد کم نمونه‌هاست. بنابراین، استفاده از روش‌های کاهش ابعاد برای حذف و یافتن تعداد بهینه‌ای از ژن‌ها برای پیش‌بینی صحیح رده‌های موردنظر، امری ضروری است. از طرفی، انتخاب یک روش کاهش ابعاد مناسب، می‌تواند به استخراج اطلاعات ارزشمند و افزایش کارایی یادگیری کمک کند. در این پژوهش از رویکرد یادگیری دسته‌ای به نام دسته ماشین بردار پشتیبان تصادفی برای یافتن مجموعه ویژگی بهینه، استفاده می‌شود. در تحلیل داده‌های واقعی مقاله حاضر، نشان داده می‌شود با تبدیل داده‌های بُعد بالا به زیرفضاهایی با بُعد پایین‌تر و ترکیب مدل‌های ماشین بردار پشتیبان، علاوه بر یافتن مجموعه‌ای از ژن‌های موثر در بروز سرطان پروستات، دقت رده‌بندی نیز افزایش می‌یابد.

واژه‌های کلیدی: یادگیری دسته‌ای، کاهش ابعاد، رده‌بندی، دسته ماشین بردار پشتیبان تصادفی، مجموعه ویژگی بهینه.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62H12، 62H30.

©نویسند(گان). ناشر انجمن آمار ایران است.
این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

۱ مقدمه

یادگیری ماشین شاخه‌ای از علوم کامپیوتر است و به عنوان زیرمجموعه‌ای از هوش مصنوعی شناخته می‌شود. در حالی که تمرکز هوش مصنوعی بر این است که ماشین را هوشمند و قادر به فکر کردن منطقی همانند انسان و حل مسائل کند، یادگیری ماشین در پی ایجاد سیستم‌ها و الگوریتم‌های رایانه‌ای است تا ماشین قادر به یادگیری تجربیات گذشته باشد. به عبارتی دیگر یادگیری ماشین از تجربیات قبلی استفاده می‌کند تا عملکرد سیستم را با استفاده از محاسبات بهبود بخشد. از آنجایی که هوش بدون توانایی یادگیری بدست نمی‌آید، یادگیری ماشین نقش مهمی در هوش مصنوعی بازی می‌کند.

در آمار، مسائل یادگیری به دو دسته یادگیری راهنماییده[1] و یادگیری ناراهنماییده[2] تقسیم می‌شوند. در میان روش‌های یادگیری راهنماییده، ماشین بردار پشتیبان[3] (SVM) کلاسی از ابزارهای قدرتمند است که به طور فزاینده‌ای در مسائل رده‌بندی، تشخیص الگو، هوش مصنوعی و بهینه‌سازی مورد استفاده قرار می‌گیرد. از نقطه‌نظر تاریخی، معرفی ساختار اولیه SVM براساس فعالیت‌های واپنیک و چروونکیس در دهه ۱۹۶۰ و حول نظریه یادگیری آماری شکل گرفت (هان و همکاران، ۲۰۱۲) و پس از به نظریه Vapnik–Chervonenkis یا به اختصار VC معروف شد. در واقع این الگوریتم یک تعمیم غیرخطی از الگوریتم تصویر (پرتره) است که در دهه شصت در روسیه توسط واپنیک و لرنر (۱۹۶۳) و واپنیک و چروونکیس (۱۹۶۴) توسعه یافت. لازم به اشاره است که نظریه VC در سه دهه اخیر توسط واپنیک و چروونکیس (۱۹۷۴)، واپنیک (۱۹۸۲) و واپنیک (۱۹۹۵) توسعه عظیمی یافته است. بنا به هان و همکاران (۲۰۱۲)، اولین نوشته علمی درباره SVM توسط بوسر و همکاران (۱۹۹۲) ارائه شد. سپس، واپنیک و همکاران (۱۹۹۷) نسخه‌ای از SVM که روشی از یادگیری راهنماییده برای مدل‌بندی رگرسیونی و توسیعی بر الگوریتم رده‌بندی ماشین بردار پشتیبان است، ارائه دادند. این ابزار ترکیبی به رگرسیون بردار پشتیبان[4] (SVR) معروف است. نکته قابل توجه این است که SVM برای پیدا کردن ابرصفحه بهینه از محاسبات ریاضی و مسائل بهینه‌سازی استفاده می‌کند. در این راستا فلتچر (۱۹۸۷) روش‌هایی برای حل مسائل بهینه‌سازی درجه دوم محدودکننده ارائه داد. همچنین پلات (۱۹۹۹) الگوریتم حداقل بهینه‌سازی متوالی را پیشنهاد داد که به طور خاص برای آموزش SVM طراحی شده است (ویتن و همکاران، ۲۰۱۷). اسمولا و شولکوپوف (۲۰۰۴) شرح مفصلی از کارکرد SVM در رگرسیون نوشته‌اند که مطالعه آن به خواننده علاقه‌مند توصیه می‌شود.

امروزه با گسترش بهره‌گیری از مجموعه داده‌های بُعد بالا و مواجهه با مشکلات مرتبط با ابعاد بالا، تحلیل داده‌ها و مطالعات نظری تغییر کرده و موجب توسعه تفکر آماری شده است. یکی از اساسی‌ترین چالش‌ها در مواجهه با ابعاد بالا، رویارویی با تعداد زیادی ویژگی در مقابل تعداد کم نمونه‌هاست. بنابراین، استفاده از رویکرد کاهش ابعاد در مسائل یادگیری ماشین حائز اهمیت است و معمولا پیش از مراحل رده‌بندی به کار گرفته می‌شود. این روش‌ها، در

[1] Supervised Learning
[2] Unsupervised Learning
[3] Support Vector Machine
[4] Support Vector Regression

حذف ویژگی‌های زائد، افزایش دقت یادگیری و تفسیرپذیری و فهم بهتر نتایج تاثیرگذارند (روحی و همکاران، ۱۴۰۲). در این پژوهش، برای غلبه بر مشکلات رده‌بندی داده‌های بُعد بالا، از رویکرد دسته زیرنمونه تصادفی[1] (RSE) به‌عنوان یک روش انتخاب ویژگی مبتنی بر یادگیری دسته‌ای، استفاده شده است. به‌علاوه، به‌منظور رده‌بندی داده‌های بُعد بالا و انتخاب مجموعه ویژگی بهینه، از مدل دسته ماشین بردار پشتیبان تصادفی[2] (RSVMC) بهره گرفته شد. این روش، مشابه رویکرد دسته زیرنمونه تصادفی، با نمونه‌گیری تصادفی همزمان از مشاهدات و فضای ویژگی بُعد بالا و با استفاده از مدل SVM به عنوان طبقه‌بند، موجب بهبود کارایی رده‌بندی می‌شود. همچنین، در پی استفاده از این رویکرد، با انتخاب مجموعه ویژگی بهینه، می‌توان به اطلاعات مفیدی از مسئله تحت مطالعه دست یافت. برای کسب اطلاعات بیشتر در خصوص دو ابزار RSE و RSVMC به بی و همکاران (۲۰۱۸) مراجعه شود.

ساختار مقاله حاضر به شرح زیر تدوین شده است. در بخش نخست مفاهیم اولیه مربوط به ماشین بردار پشتیبان ارائه شده است. بخش دوم دربرگیرنده جزئیاتی از رویکرد RSE و RSVMC است. در آخرین بخش نتایج تحلیل مثال واقعی برای ارزیابی عملکرد مسئله حاضر ارائه شده است. در انتها، علاوه بر بحث و نتیجه‌گیری کلی، پیشنهاداتی برای تحقیقات آتی مرتبط با موضوع مقاله حاضر ارائه خواهد شد.

۲ مروری بر ماشین بردار پشتیبان

رویکرد SVM روشی برای رده‌بندی خطی و غیرخطی داده‌ها است. الگوریتم SVM با استفاده از نگاشت غیرخطی، داده‌های آموزشی اصلی را به بُعد بالاتر انتقال می‌دهد طوری که در فضای جدید بتوان ابرصفحه جداکننده بهینه خطی و به عبارتی دیگر مرز تصمیم را بدست آورد. این الگوریتم با استفاده از بردارهای پشتیبان و حاشیه[3] که توسط بردارهای پشتیبان تعریف می‌شود، ابرصفحه جداکننده را می‌یابد. همانطور که در هان و همکاران (۲۰۱۲) اشاره شده است، می‌توان ماشین بردار پشتیبان را از دو منظر خطی و غیرخطی مطالعه کرد. برای فهم موضوع، در ادامه شرح مختصری از هر یک از آن‌ها ارائه می‌شود.

۲.۱ ماشین بردار پشتیبان خطی

برای شناسایی و تعیین مرز در روش SVM به یک مجموعه اصلی از نقاط نیاز است. این نقاط همان بردارهای پشتیبان هستند که مرز را پشتیبانی می‌کنند و معمولا دربرگیرنده یک سطر از داده‌ها با ویژگی‌های مختلف‌اند. این مرز، ابرصفحه نامیده می‌شود و به عنوان یک مثال ساده در دو بُعد، می‌تواند یک خط راست یا منحنی و در سه بُعد یک صفحه یا سطح پیچیده نامرتب باشد.

[1] Random Subset Ensemble
[2] Random Support Vector Machine Cluster
[3] Margin

فرض کنید در یک مسئله رده‌بندی دوتایی مجموعه یادگیری از داده‌ها به صورت نمادین

$$\mathcal{L} = \{(x_i, y_i) : i = ۱, \ldots, n\}; \quad x_i \in \mathcal{R}^p, y_i \in \{-۱, +۱\}$$

در اختیار است. در مسئله رده‌بندی دوتایی[1]، هدف اساسی استفاده از $\mathcal{L}$ برای ساختن تابع $f : \mathcal{R}^p \to \mathcal{R}$ است طوری که

$$C(x) = \text{sign}(f(x)), \quad x \in \mathcal{R}^p$$

یک طبقه‌بند باشد، که در آن تابع علامت (sign) به صورت

$$\text{sign}(\beta) = \begin{cases} ۱ & \beta \geq ۰, \\ -۱ & \beta < ۰ \end{cases}$$

تعریف می‌شود. تابع جداکننده[2] f هر نقطه جدید x در مجموعه آزمایشی را به یکی از دو گروه $\sqcap_+$ یا $\sqcap_-$ رده‌بندی می‌کند. به بیانی دیگر، هدف یافتن تابع f ای است که نقاط مثبت در مجموعه آزمایشی را به $\sqcap_+$ و نقاط منفی را به $\sqcap_-$ اختصاص دهد. در ساده‌ترین حالت فرض کنید که نقاط مثبت و منفی از مجموعه یادگیری $\mathcal{L}$ توسط یک ابرصفحه، جداپذیر باشند. معادله ابرصفحه جداکننده را به صورت $\{x : f(x) = \beta_۰ + x^T\beta = ۰\}$ در نظر بگیرید. اگر d_- کوتاه‌ترین فاصله از ابرصفحه جداکننده به نزدیک‌ترین نقاط منفی و d_+ کوتاه‌ترین فاصله از همان ابرصفحه به نزدیک‌ترین نقاط مثبت باشد، حاشیه ابرصفحه جداکننده به صورت $d = d_- + d_+$ تعریف می‌شود. در صورتی که فاصله ابرصفحه با نزدیک‌ترین مشاهدات ماکسیمم شود، ابرصفحه جداکننده بهینه یا طبقه‌بند حاشیه بیشین است (آیزنمن، ۲۰۰۸). اگر داده‌های یادگیری دو کلاس به طور خطی جداپذیر باشند، آنگاه β و $\beta_۰$ ای وجود دارند طوری که گزاره‌های شرطی

$$\beta_۰ + x_i^T\beta \geq +۱, \quad if \quad y_i = +۱, \tag{۱}$$

$$\beta_۰ + x_i^T\beta \leq -۱, \quad if \quad y_i = -۱. \tag{۲}$$

قابل نگارش هستند. بیان این نکته ضروری است که در حالت تساوی برای روابط (۱) و (۲) بردار داده‌ها روی ابرصفحه‌های $H_{+۱} : (\beta_۰ - ۱) + x_i^T\beta = ۰$ و $H_{-۱} : (\beta_۰ + ۱) + x_i^T\beta = ۰$ قرار می‌گیرند. نقاطی از مجموعه $\mathcal{L}$ که روی ابرصفحه‌های $H_{+۱}$ و $H_{-۱}$ قرار گرفته باشند، بردارهای پشتیبان نامیده می‌شوند و معمولا شامل درصد کمی از کل نقاط نمونه هستند. اگر $x_{+۱}$ روی ابرصفحه $H_{+۱}$ و $x_{-۱}$ روی $H_{-۱}$ قرار گرفته باشند آنگاه از

[1] Binary Classification Problem
[2] Separating function

جمع طرفین روابط $\beta_\circ + x_{+1}^T\beta = +1$ و $\beta_\circ + x_{-1}^T\beta = -1$ تساوی $\beta_\circ = -\frac{1}{2}\left\{x_{+1}^T\beta + x_{-1}^T\beta\right\}$ به دست می‌آید.

برای پیدا کردن فاصله عمودی ابرصفحه‌های H_{+1} و H_{-1}، نقاط x_{+1} و x_{-1} را به ترتیب روی ابرصفحه‌های H_{+1} و H_{-1} در نظر گرفته، خواهیم داشت:

$$\begin{cases} f(x_{+1}) := \beta_\circ + x_{+1}^T\beta = +1, \\ f(x_{-1}) := \beta_\circ + x_{-1}^T\beta = -1. \end{cases} \tag{۳}$$

اکنون با توجه به فرمول استاندارد برای فاصله دو خط موازی، فاصله دو ابرصفحه برابر

$$d = \frac{|f(x_{+1}) - f(x_{-1})|}{\|\beta\|} = \frac{2}{\|\beta\|},$$

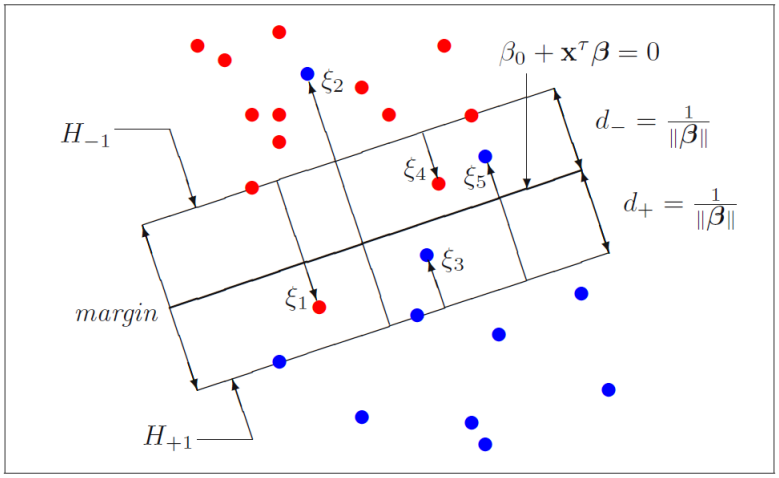
است. درنتیجه حاشیه ابرصفحه جداکننده از تساوی $d = d_{+1} + d_{-1} = \frac{2}{\|\beta\|}$ قابل محاسبه است. هدف از اجرای الگوریتم SVM یافتن ابرصفحه‌ای است که حاشیه $d = \frac{2}{\|\beta\|}$ را با توجه به شرط

$$y_i(\beta_\circ + x_i^T\beta) \geq 1, \quad i = 1, \ldots, n \tag{۴}$$

بیشینه سازد، که این امر هم‌ارز با پیداکردن $\beta_\circ$ و β ای است که $\|\beta\|$ مینیمم شود. چنین ابرصفحه‌ای، همان ابرصفحه بهینه است. انتظار می‌رود که اگر همه نقاط نمونه به درستی رده‌بندی شوند، آنگاه برای هر نمونه (x_i, y_i) به ازای $i = 1, ..., n$ نامساوی (۴) همواره برقرار باشد. دلیل گذاشتن این محدودیت آن است که نقاط در حاشیه ابرصفحه جداکننده قرار نگیرند و به درستی رده‌بندی شوند. در عمل بعید به نظر می‌رسد داده‌های دو گروه همواره جداپذیر باشند. به عبارتی دیگر، در برخی مواقع مشاهدات در جاهایی همپوشانی دارند که بسته به میزان همپوشانی، باعث بروز مشکلاتی در قانون رده‌بندی می‌شود. یک رهیافت برای داده‌هایی که دارای نرخی از همپوشانی هستند، ایجاد فرمول‌بندی قابل انعطاف مسئله است. برای رسیدن به این هدف، به ازای هر مشاهده (x_i, y_i) از مجموعه $\mathcal{L}$ متغیر کمکی[1] (ξ_i)، به عنوان مولفه‌ای از بردار $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T \geq 0$، طوری تعریف می‌شود که محدودیت (۴) به نامساوی $y_i(\beta_\circ + x_i^T\beta) + \xi_i \geq 1$ تبدیل شود. نمایشی از آنچه در چنین وضعیتی رخ می‌دهد در شکل ۱ نشان داده شده است. در این شکل، نقاط پشتیبان، آن نقاطی هستند که روی ابرصفحه‌های H_{+1} و H_{-1} قرار گرفته‌اند. متغیرهای کمکی ξ_1 و ξ_4 مربوط به نقاط قرمزند، این نقاط محدودیت ابرصفحه H_{-1} را نقض می‌کنند و ξ_2، ξ_3 و ξ_5 متغیرهای کمکی مربوط به نقاط آبی‌اند و این نقاط محدودیت ابرصفحه H_{+1} را نقض می‌کنند. اکنون باید ابرصفحه بهینه به گونه‌ای تعیین شود که هم حاشیه $\frac{2}{\|\beta\|}$ و هم تابع ساده محاسباتی متغیرهای کمکی را کنترل کند.

[1] Slack variable

این تابع به صورت $g_\sigma(\boldsymbol{\xi}) = \sum_{i=1}^{n} \xi_i^\sigma$, تعریف می‌شود. لازم به ذکر است که مقادیر معمول برای σ اعداد یک



شکل ۱. جداپذیری داده‌ها براساس الگوی غیرخطی.

یا دو است و رویکردهای متناظر به ترتیب به نرم L_1[1] و نرم L_2[2] معروفند (جیمز و همکاران، ۲۰۱۴). برای وضوح مطالب، در ادامه، حالت $\sigma = 1$ در نظر گرفته و نتایج حاصل از حل نامساوی مورد اشاره ارائه می‌شود. توجه شود که چون در این حالت تابع هدف ($\|\beta\|$) تابعی محدب نیست، مینیمم‌سازی آن از طریق مشتق‌گیری امکان‌پذیر نیست. لذا، بهینه‌سازی $\frac{1}{2}\|\beta\|^2$ که تابعی محدب با همان مقدار مطلوب است، مدنظر قرار گرفته و در نتیجه مسئله به کمک روش‌های برنامه‌ریزی درجه دوم حل خواهد شد. به طور دقیق‌تر، شخص با مسئله بهینه‌سازی حاشیه نرم L_1 روبرو می‌شود که در واقع معادل حل مسئله

$$\text{minimize} \quad \{\frac{1}{2}\|\beta\|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{n}} \xi_{\mathrm{i}}\}, \tag{۵}$$

به ازای β_0، β و ξ است به شرطی که

$$\xi_i \geq 0, \quad y_i(\beta_0 + x_i^T\beta) \geq 1 - \xi_i, \quad i = 1, \ldots, n.$$

کمیت $C \geq 0$ در رابطه (۵)، پارامتر تنظیم[3] است که اندازه متغیرهای کمکی را کنترل می‌کند و $\sum_{i=1}^{n} \xi_i$ مرز بالایی

[1] $L_1 - \text{norm}$
[2] $L_2 - \text{norm}$
[3] Regularization parameter

برای تعداد نمونه‌های آموزشی‌ای است که به اشتباه رده‌بندی شده‌اند (یانگ، ۲۰۱۹).

می‌توان مسئله بهینه‌سازی مقید (۵) را با استفاده از ضرایب لاگرانژ به صورت رابطه نامقید

$$F_p(\beta,\xi,\alpha,\eta)=\frac{1}{2}\|\beta\|^2+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left\{y_i(\beta_0+x_i^T\beta)-(1-\xi_i)\right\}$$
$$-\sum_{i=1}^{n}\eta_i\xi_i, \tag{۶}$$

با شرایط $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)^T\geq 0$ و $\boldsymbol{\eta}=(\eta_1,\ldots,\eta_n)^T\geq 0$ نوشت. توجه شود که $\boldsymbol{\alpha}$ و $\boldsymbol{\eta}$ ضرایب لاگرانژ یا متغیرهای دوگان نامیده می‌شوند (خرم، ۱۳۹۰). از آنجایی که هدف رسیدن به مقدار بهینه تابع $F_p=F_p(\beta,\xi,\alpha,\eta)$ برحسب پارامترهای رابطه است، با ثابت اختیار کردن η و α و مشتق‌گیری از F_p نسبت به β_0، β و $\boldsymbol{\xi}$ خواهیم داشت:

$$\frac{\partial F_p}{\partial\beta_0}=-\sum_{i=1}^{n}\alpha_i y_i, \tag{۷}$$

$$\frac{\partial F_p}{\partial\beta}=\beta-\sum_{i=1}^{n}\alpha_i y_i x_i, \tag{۸}$$

$$\frac{\partial F_p}{\partial\xi_i}=C-\alpha_i-\eta_i,\quad i=1,\ldots,n. \tag{۹}$$

با برابر صفر قرار دادن روابط (۷)، (۸) و (۹) داریم:

$$\sum_{i=1}^{n}\alpha_i y_i=0,\quad \beta=\sum_{i=1}^{n}\alpha_i y_i x_i,\quad \alpha_i=C-\eta_i,\quad i=1,\ldots,n. \tag{۱۰}$$

با جایگذاری (۱۰) در (۶) خواهیم داشت:

$$F_p=\frac{1}{2}\|\beta\|^2-\beta^T\beta+C\sum_{i=1}^{n}\xi_i+\sum_{i=1}^{n}\alpha_i\underbrace{-\sum_{i=1}^{n}\alpha_i\xi_i-\sum_{i=1}^{n}\eta_i\xi_i}_{-C\sum_i\xi_i}. \tag{۱۱}$$

سپس با ساده‌سازی روابط، مسئله بیشینه‌سازی دوگان به صورت

$$F_D(\boldsymbol{\alpha})=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i^T x_j), \tag{۱۲}$$

تبدیل می‌شود که در واقع تابعی از α_i خواهد بود و به همین دلیل مسئله دوگان نامیده می‌شود. در واقع ضرایب α_i متغیرهای دوگان هستند و هر کدام از نقاط آموزشی متغیر دوگان مربوط به خود را دارد. با توجه به شرط $C - \alpha_i - \eta_i = ۰$ و $\eta_i \geq ۰$ خواهیم داشت $۰ \leq \alpha_i \leq C$ که در واقع محدودیت مسئله دوگان است. ضروری است جواب‌های بهینه در شرایط KKT [۱] صدق کنند. لازم به اشاره است که برای رسیدن به جواب‌های بهینه در مسائل برنامه‌ریزی خطی این شروط شرایط لازم و کافی تلقی می‌شوند (خرم، ۱۳۹۰). شرایط مورد اشاره به همراه معادلات منتهی به جواب‌های بهینه به صورت

$$\begin{aligned}
&y_i(\beta_۰ + x_i^T\beta) - (۱ - \xi_i) \geq ۰,\\
&\xi_i \geq ۰, \quad \alpha_i \geq ۰, \quad \eta_i \geq ۰,\\
&\alpha_i \left\{y_i(\beta_۰ + x_i^T\beta) - (۱ - \xi_i)\right\} = ۰,\\
&\xi_i(\alpha_i - C) = ۰.
\end{aligned} \tag{۱۳}$$

هستند، که در (۱۳) متغیر کمکی ξ_i وقتی می‌تواند مخالف صفر باشد که $\alpha_i = C$. شایان ذکر است که در اینجا فقط ضرایب غیرصفر α_i در راه‌حل کلی نقش دارند. با توجه به شرط KKT رابطه (۱۳)، زمانی که $\alpha_i \neq ۰$ نامساوی $y_i(\beta_۰ + x_i^T\beta) \geq ۱ - \xi_i$ که همان محدودیت مسئله است و داده‌های دو طبقه را نشان می‌دهد، به تساوی تبدیل می‌شود. پس تنها داده‌های آموزشی با $\alpha_i > ۰$ می‌توانند در راه‌حل مسئله نقش داشته باشند که در اصل بردارهای پشتیبان را شکل می‌دهند. در نهایت، برآورد $\boldsymbol{\alpha}$ از رابطه (۱۲) به دست می‌آید و سپس با توجه به رابطه (۱۰) برآورد β قابل محاسبه خواهد بود.

۲.۲ ماشین بردار پشتیبان غیرخطی

در عمل موارد بیشماری پیش می‌آید که مسائل مورد مطالعه غیرخطی هستند که در آن صورت ماشین بردار پشتیبان خطی از کارایی لازم برخوردار نیست. ایده ماشین بردار پشتیبان غیرخطی در واقع یافتن ابرصفحه جداکننده بهینه در فضای ویژگی بُعد بالاست. انتظار می‌رود بُعد فضای ویژگی، مانع بزرگی برای ایجاد ابرصفحه جداکننده بهینه باشد. این مسئله نخستین بار توسط کرتس و واپنیک (۱۹۹۵) با استفاده از توابع هسته مورد مطالعه قرار گرفت. در واقع نقاط با استفاده از توابع هسته به فضای جدیدی با بُعد متفاوت نگاشت می‌شوند که این فضای جدید، فضای ویژگی نامیده می‌شود و نقاط نگاشت یافته در فضای ویژگی به صورت خطی جداپذیر خواهند بود (آیزنمن، ۲۰۰۸).

[1] The Karush-Kuhn-Tucker conditions

۳ کاهش ابعاد

امروزه با گسترش مجموعه داده‌های متعدد با ابعاد بالا رویکرد کاهش بُعد نقش مهمی را در مسائل یادگیری ماشین ایفا می‌کند. اگر تعداد ویژگی‌ها خیلی زیاد و هر ویژگی قابلیت رده‌بندی کمی داشته باشد، بهتر است، مجموعه ویژگی‌ها به طور خطی یا غیرخطی، به مجموعه‌ای کاهش‌یافته از ویژگی‌ها تبدیل شود. روش‌های زیادی برای از بین بردن ویژگی‌های زائد وجود دارد که هر یک معایب و مزایای خاص خود را دارند. اگر چه حذف ویژگی‌های زائد، یک نگاه تقریبا نسبی به مسئله انتخاب ویژگی است اما انجام آن منجر به بهبود چشمگیری در روش‌های محاسباتی مرتبط با SVM خواهد شد (جیمز و همکاران، ۲۰۱۴). کاهش ابعاد می‌تواند توسط دو رویکرد انتخاب و استخراج ویژگی صورت گیرد. در ادامه، از بین روش‌های کاهش بعد مرتبط با SVM به روشی نوین که کاربرد آن در مقاله حاضر تشریح شده است، پرداخته می‌شود. عملکرد این روش از نقطه‌نظر موفقیت در اجرا، ارائه نتایج صحیح و سرراست بودن شایسته توجه است.

۳.۱ انتخاب ویژگی مبتنی بر یادگیری دسته‌ای

در سال‌های اخیر، نوع جدیدی از روش‌های انتخاب ویژگی با نام انتخاب ویژگی مبتنی بر یادگیری دسته‌ای توسط سائی و همکاران (۲۰۰۸) پیشنهاد و مورد مطالعه قرار گرفته است. این رویکرد، انتخاب ویژگی و یادگیری دسته‌ای را ترکیب می‌کند و به دلیل عملکرد خوب آن، در یادگیری راهنماییده به کار گرفته می‌شود. در این روش، طبقه‌بندهای متعددی ایجاد و سپس نتایج رده‌بندی آن‌ها به صورت مناسبی ادغام می‌شود. وانگ و چیانگ (۲۰۱۰) نشان دادند که عملکرد این روش معمولا دقیق‌تر از نتایج طبقه‌بندهای جداگانه است و به طور موثر راه‌حلی باثبات را ارائه می‌دهد و موجب بهبود کارایی یادگیری می‌شود. ایده کلی این روش آن است که فرآیند انتخاب ویژگی چندین بار تکرار می‌شود تا دسته‌های متنوع ویژگی ایجاد شود تا از آن طریق بتوان خروجی‌ها را به طور مناسبی ترکیب کرد (گوآن و همکاران، ۲۰۱۴).

۳.۲ دسته زیرنمونه تصادفی

این ابزار یکی از انواع روش‌های زیرفضای تصادفی است و به عنوان رهیافتی از یادگیری دسته‌ای، توسط سرپن و پاتیکال (۲۰۰۹) برای مواجهه با مشکلاتی که در تحلیل داده‌های بعد بالا روی می‌دهد، معرفی شده است. در روش RSE یک مسئله پیچیده بُعد بالا به چندین زیرمسئله با بُعد پایین‌تر تبدیل و بدین ترتیب جنبه‌های پیچیدگی محاسباتی مسئله اصلی از بین می‌رود. نکته حائز اهمیت این روش آن است که فرآیند اجرا با استفاده از روش‌های نمونه‌گیری تصادفی انجام می‌شود. از منظر محاسباتی، هر زیرفضای ویژگی با بُعد پایین‌تر برای آموزش طبقه‌بند اصلی دسته، استفاده می‌شود و پیش‌بینی‌های طبقه‌بندهای اصلی با استفاده از یک روش ایستا، با هم ترکیب می‌شوند. آموزش طبقه‌بندهایی با زیرفضاهای ویژگی متفاوت، همبستگی بین طبقه‌بندهای اصلی را تا حدی کاهش می‌دهد. بنا به

پاتیکال و سرپن (۲۰۱۲)، RSE تبدیلات متفاوتی از فضای ویژگی را به هر طبقه‌بند اختصاص می‌دهد که منجر به تنوع طبقه‌بندها می‌شود.

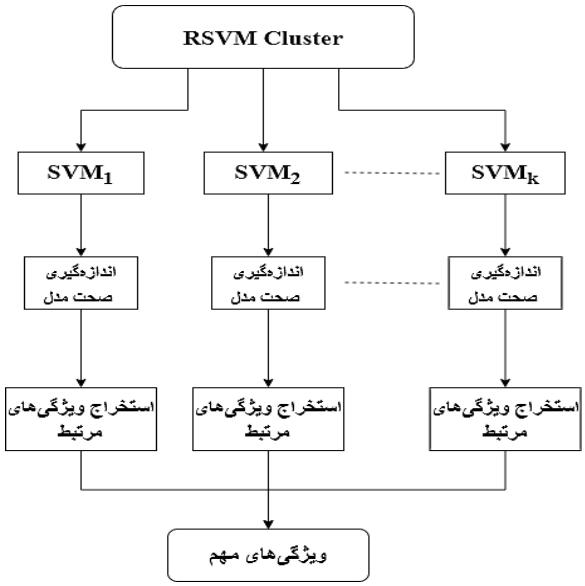
۳.۳ روش دسته ماشین بردار پشتیبان تصادفی

در روش RSVMC ویژگی‌ها و نمونه‌ها برای استفاده در SVM به طور تصادفی انتخاب و تعداد زیادی مدل SVM مدنظر قرار گرفته و در نهایت دقت تحلیل از طریق ترکیب این مدل‌ها بهبود می‌یابد. در حقیقت، در این روش از رهیافت یادگیری دسته‌ای با استفاده از طبقه‌بند SVM، استفاده می‌شود. با پیروی از بی و همکاران (۲۰۱۸) مراحل اجرای رویکرد RSVMC به شرح زیر است:

در مرحله اول مجموعه نمونه (N) به طور تصادفی به سه جزء مجموعه آموزشی (N_1)، اعتبارسنجی (N_2) و آزمایشی (N_3)، طوریکه $N = N_1 + N_2 + N_3$ تقسیم می‌شود. فرض کنید بُعد داده‌ها D باشد. به زبانی دیگر، فرض کنید ماتریس داده‌ها (X) از بُعد $N \times D$ باشد. آنگاه، نمونه‌ای تصادفی به حجم n از N_1 و d از ویژگی‌ها $(d < D)$ انتخاب و مدل SVM روی مجموعه حاصل اعمال می‌شود. این روند k بار تکرار و در نهایت k مدل SVM تشکیل و از نتیجه‌اش برای ساختن یک خوشه یا دسته استفاده می‌شود. پس از ساختن یک دسته، از مجموعه اعتبارسنجی N_2 برای بهینه‌سازی پارامترها بهره برده می‌شود. وقتی نمونه جدیدی ارائه شود، k ماشین بردار پشتیبان آن را طبقه‌بندی می‌کنند و رده‌بندی هر نمونه در مجموعه آزمایشی N_3 براساس دسته SVM تصادفی پیش‌بینی و مقدار پیش‌بینی شده با مقادیر واقعی مقایسه می‌شود. توجه شود که در این حالت تعداد موقعیت‌های سازگار با N_c نمایش داده و دقت دسته SVM تصادفی از طریق نسبت N_c/N_3 محاسبه شود. از آنجایی که هر مدل SVM با انتخاب بخشی از ویژگی‌ها به طور تصادفی تنظیم می‌شود، لذا در برخورد با داده‌های بُعد بالا بسیار کارآمد است. به علاوه، چون نمونه‌ها و ویژگی‌های هر SVM به طور تصادفی انتخاب می‌شوند، مدل‌های حاصل از رده‌بندی تا حد زیادی متفاوت از هم خواهند بود. به عبارتی دیگر، چون رویکرد RSVMC منجر به تنوع مدل‌ها می‌شود انتظار می‌رود دید بهتری از رده‌بندی ارائه کند. این موضوع باعث می‌شود که RSVMC از عملکرد بهتری در مقایسه با روش‌های معمولی SVM برخوردار باشد.

شکل ۲ مراحل اجرای روش RSVMC را نشان می‌دهد. برای انتخاب مجموعه ویژگی‌های مهم میزان دقت هر مدل محاسبه و از دقت هر مدل SVM به عنوان معیاری برای ارزیابی کیفیت ویژگی‌ها استفاده می‌شود. ویژگی‌هایی که سهم بسزایی در دقت SVM دارند، ویژگی‌های مهم نامیده می‌شوند. با این نگاه، در رویکرد RSVMC، ابتدا دقت هر مدل SVM محاسبه شده و سپس دقت مدل‌ها در دسته از بزرگ‌ترین تا کوچک‌ترین مرتب می‌شوند. بعد از آن ویژگی‌های مورد استفاده در هر مدل استخراج و یک ماتریس ویژگی ساخته می‌شود. در نهایت، تکرار هر ویژگی در ماتریس ویژگی شمارش و ویژگی‌ها با تکرار بیشتر به عنوان ویژگی‌های مهم شناخته می‌شوند.

برای یافتن مجموعه ویژگی بهینه از میان ویژگی‌های مهم دوباره عمل انتخاب ویژگی تکرار و با محاسبه دقت مدل RSVMC، مجموعه ویژگی با بیشترین کارایی به عنوان مجموعه ویژگی بهینه انتخاب می‌شود. با توجه به



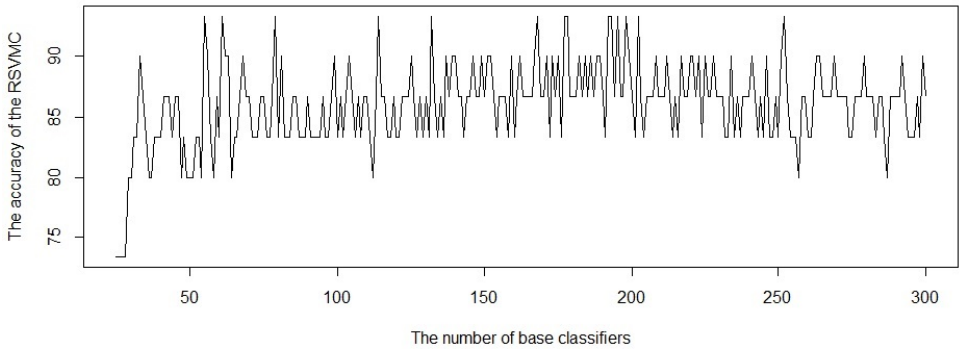
شکل ۲. انتخاب مجموعه ویژگی مهم در روش RSVMC.

انتخاب زیرمجموعه‌های تصادفی ویژگی و زیرنمونه‌های تصادفی، می‌توان گفت انتخاب ویژگی در رویکرد ماشین بردار پشتیبان تصادفی قالبی شبیه RSE دارد.

۴ کاربرد در داده‌های ژنتیک

در این بخش، با استفاده از یک مثال واقعی عملکرد مناسب رویکرد RSVMC نشان داده می‌شود. داده‌های موردنظر به نام مجموعه داده prostate، مربوط به بیان ژن تومور پروستات است، که جزئیاتی از آن در ادامه می‌آید. این مجموعه داده که متشکل از ۱۰۲ نمونه شامل ۵۲ بیمار مبتلا به سرطان پروستات و ۵۰ فرد سالم با ۶۰۳۳ ویژگی (ژن) است، در کتابخانه spls در نرم‌افزار R (تیم مرکزی آر، ۲۰۱۹) قابل دسترسی است. ابتدا برای ایجاد یک دسته RSVMC، ۱۰۲ نمونه، به ۶۲ نمونه آموزشی، ۳۰ مورد آزمایشی و ۱۰ نمونه اعتبارسنجی تقسیم شدند. با پیروی از بی و همکاران (۲۰۱۸)، تعداد ویژگی‌ها و نمونه‌های انتخابی را به ترتیب $\sqrt{۶۰۳۳} \approx ۷۸$ و ۶۱ در نظر گرفته و در برازش مدل موردنظر از تابع هسته سیگموید استفاده کردیم. در اینجا تعداد تکرار طبقه‌بندهای SVM از ۲۵ تا ۳۰۰ در نظر گرفته شده و برای انتخاب مجموعه‌ای از ویژگی‌ها، معیار انتخاب ویژگی‌ها را ۷۰ درصد قرار دادیم. باید اشاره شود که این معیار با توجه به مسئله موردنظر و به صورت تجربی تعیین می‌شود و قابل تغییر است. شکل ۳ صحت رده‌بندی مدل RSVMC را با توجه به تغییر تعداد طبقه‌بندها، نمایش می‌دهد. تعداد بهینه مدل‌های SVM در یک

دسته با توجه به تعداد ژن‌ها در مجموعه ویژگی بهینه، صحت و حساسیت مدل موردنظر، ۶۱ در نظر گرفته شده است.



شکل ۳. تعداد بهینه مدل‌های ماشین بردار پشتیبان در داده‌های بیان ژن تومور پروستات.

جدول ۱ میزان صحت پیشگویی‌ها را برای دو گروه مورد مطالعه نشان می‌دهد. مشاهده می‌شود که در میان ۳۰ نمونه آزمایشی، ۱۳ نفر، مبتلا به سرطان پروستات هستند که همگی به درستی رده‌بندی شده‌اند، همچنین مدل موردنظر، از میان ۱۷ فرد سالم ۱۵ نفر را درست پیش‌بینی کرده است.

جدول ۱. ماتریس درهم‌ریختگی حاصل از مدل RSVMC بر روی مجموعه داده بیان ژن تومور پروستات.

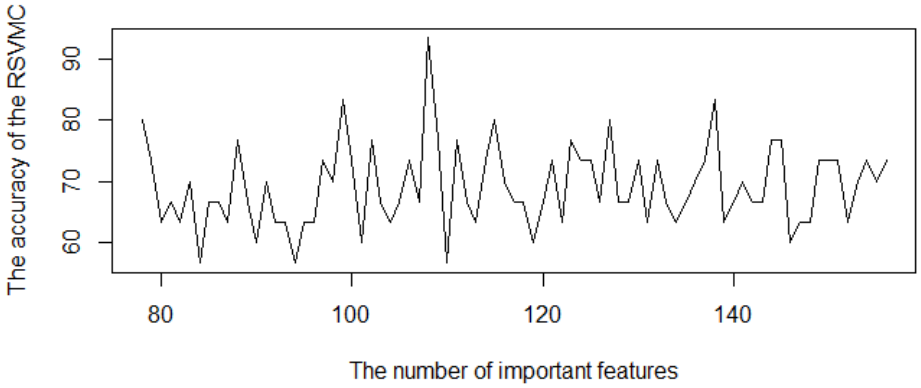
	وضعیت پیش‌بینی افراد	
وضعیت واقعی افراد	دارای سرطان	سالم
دارای سرطان	۱۳	۰
سالم	۲	۱۵

معیارهای ارزیابی (هان و همکاران، ۲۰۱۲) پس از انتخاب مجموعه ویژگی بهینه در مدل RSVMC، محاسبه و مقادیر

$$\text{Accuracy} = (\frac{۱۵+۱۳}{۳۰}) \times ۱۰۰ = ۹۳/۳۳$$
$$\text{Sensitivity} = (\frac{۱۳}{۰+۱۳}) \times ۱۰۰ = ۱۰۰$$
$$\text{Specificity} = (\frac{۱۵}{۲+۱۵}) \times ۱۰۰ = ۸۸/۲۳$$

حاصل شده‌اند. همانطور که ملاحظه می‌شود نسبت نمونه‌هایی که به درستی رده‌بندی شده‌اند تقریبا ۹۳ درصد است. شکل ۴ مجموعه بهینه‌ای از ژن‌ها که در بروز سرطان پروستات حائز اهمیت هستند را نشان می‌دهد. لازم به اشاره است زمانی که مجموعه‌ای ۱۱۷ تایی از ژن‌ها را از میان ویژگی‌های مهم انتخاب می‌کنیم میزان صحت و حساسیت

مدل RSVMC، به بالاترین مقدار خود می‌رسد.



شکل ۴. انتخاب مجموعه ویژگی بهینه از میان ویژگی‌های مهم.

بحث و نتیجه‌گیری

روش SVM از محبوبیت بالایی در حوزه یادگیری ماشین برخوردار است و تلاش‌های بسیاری برای بهبود عملکرد آن بویژه در تحلیل داده‌های بعد بالا صورت گرفته است. در سال‌های اخیر، مدل‌های بسیاری برای رده‌بندی داده‌های بیان ژن سرطان پروستات توسط محققین ارائه شده است. به عنوان مثال، گوناواتی و پرمالاتا (۲۰۱۴) از این مجموعه داده و الگوریتم‌های KNN، SVM و روش اعتبارسنجی متقابل ۵ بخشی استفاده کرده‌اند و به دقت ۸۵/۷۱ درصد رسیده‌اند. همچنین، دشتبان و بالافار (۲۰۱۷) از تقسیم داده‌ها به نمونه‌های آموزشی و آزمایشی و به کارگیری روش SVM کارایی مدل خود را ۹۱/۲ درصد گزارش کردند. در این مقاله، رویکرد جدیدی با عنوان RSVMC مطرح شد. نشان داده شد که، در روش RSVMC مشابه رویکرد RSE با انتخاب زیرمجموعه‌های ویژگی و نمونه‌های آموزشی به طور تصادفی، و با استفاده از طبقه‌بند SVM، علاوه بر افزایش کارایی کلی مدل رده‌بندی، با یافتن مجموعه ویژگی بهینه و تفسیر ژن‌های موثر در بروز سرطان پروستات، می‌توان به اطلاعات دقیق‌تری از داده‌ها رسید. همچنین در این مدل نرخ مثبت صحیح به بالاترین مقدار خود رسیده که حاکی از عملکرد خوب آن در تشخیص بیماران و یافتن مجموعه ژن‌های موثر در بروز سرطان پروستات است. تحقیقات آتی که بر پایه این موضوع می‌تواند انجام شود، عبارتند از: استفاده از رویکرد RSVMC برای رده‌بندی داده‌های چند کلاسه، به کارگیری دیگر الگوریتم‌های داده‌کاوی به عنوان طبقه‌بند در روش RSE و مقایسه آن‌ها با رویکرد SVM، استفاده از روش‌های انتخاب ویژگی مبتنی بر ماشین بردار پشتیبان و مقایسه آن با روش RSVMC. واضح است که بررسی کارایی هر کدام از رویکردهای پیشنهادی از هر دو منظر کاربردی و نظری نیز موضوع جالب توجهی برای تحقیق خواهد بود.

تقدیر و تشکر

نویسندگان مقاله کمال قدردانی و تشکر را از پیشنهادات ارزنده داوران، سردبیر و ویراستار محترم مجله که باعث ارائه بهتر و افزایش سطح کیفی مقاله شده است، دارند.

مراجع

خرم، ا. (۱۳۹۰)، برنامه‌ریزی خطی و جریان‌های شبکه‌ای، نشر کتاب دانشگاهی، تهران.

روحی، ا.، جهادی، ف.، روزبه، م. و زال زاده، س. (۱۴۰۲)، تحلیل داده‌های با بعد بالا با استفاده از رگرسیون بردار پشتیبان تعمیم یافته، رگرسیون تابعی، رگرسیون ستیغی و لاسو، مجله علوم آماری، **۱۷**، ۱۰۲-۸۱.

Bi, X. A., Shu, Q., Sun, Q. and Xu, Q. (2018), Random Support Vector Machine Cluster Analysis of Resting-State fMRI in Alzheimer's Disease, *PLOS ONE*, **13**, 1-17.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992), A Training Algorithm for Optimal Margin Classifiers. In: Haussler, D. (Ed.), *Proceedings of the Annual Conference on Computational Learning Theory*, ACM, Pittsburgh, 144–152.

Cortes, C. and Vapnik, V. (1995), Support Vector Networks, *Machine Learning*, **20**, 273-297.

Dashtban, M. and Balafar, M. (2017), Gene Selection for Microarray Cancer Classification Using a New Evolutionary Method Employing Artificial Intelligence Concepts, *Genomics*, **109**, 91–107.

Fletcher, R. (1987), *Practical Methods of Optimization*, John Wiley and Sons, New York.

Guan, D., Yuan, W., Lee, Y. K., Najeebullah, K. and Rasel, M. K. (2014), A Review of Ensemble Learning Based Feature Selection, *IETE Technical Review*, **27**, 190–198.

Gunavathi, C. and Premalatha, K. (2014), Performance Analysis of Genetic Algorithm with KNN and SVM for Feature Selection in Tumor Classification. *Int J Comput Electr Autom Control Inform Eng*, **8**, 1490–1497.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002), Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, **46**, 389–422.

Han, J., Kamber, M., and Pei, J. (2012), *Classification: Advanced Methods, Data Mining Concepts and Techniques*, Morgan Kaufmann, Waltham.

Izenman, A. J. (2008), *Modern Multivariate Statistical Techniques, Multidimensional Scaling and Distance Geometry*, Springer, New York.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2014), *The Elements of Statistical Learning: with Applications in R*, Springer, New York.

Pathical, S. and Serpen, G. (2012), Hybrid Random Subsample Classifier Ensemble for High Dimensional Data Sets, *International Journal of Hybrid Intelligent Systems*, **9**, 91–103.

Platt, J. (1999), Fast Training of Support Vector Machines Using Sequential Minimal Optimization, In: Sch¨olkopf, B., Burges, C. J. C. and Smola, A.J. (Eds.) *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, 185-208.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Saeys, Y., Abeel, T. and Peer, Y. (2008), Robust Feature Selection Using Ensemble Feature Selection Techniques, In *Proceedings of the ECML PKDD*, **5212**, 31325.

Serpen, G. and Pathical, S. (2009), Classification in High-Dimensional Feature Spaces: Random Subsample Ensemble, In *Proceedings of International Conference on Machine Learning and Applications*, PP. 740-745.

Smola, A. J. and Schölkopf, B. (2004), A Tutorial on Support Vector Regression, *Statistics and Computing*, **14**, 199–222.

Vapnik, V. and Lerner, A. (1963), Pattern Recognition Using Generalized Portrait Method. *Automation and Remote Control*, **24**, 774–780.

Vapnik, V. and Chervonenkis, A. (1964), A Note on One Class of Perceptrons, *Automation and Remote Control*, **25**, 103-109.

Vapnik, V. and Chervonenkis, A. (1974), *Theory of Pattern Recognition*, Nauka, Moscow.

Vapnik, V. N. (1982), *Estimation of Dependences Based on Empirical Data*, Springer, Berlin.

Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, Springer, New York.

Vapnik, V., Golowich, S. and Smola, A. (1997), Support Vector Method for Function Approximation, Regression Estimation and Signal Processing, In: Mozer, M. C., Jordan, M. I. and Petsche, T. (Eds), *Advances in Neural Information Processing Systems*, 281-287.

Wang, B. and Chiang, H. D. (2010), ELITE: Ensemble of Optimal Input-pruned Neural Networks Using TRUST-TECH, *IEEE: Transaction on Neural Networks*, **22**, 96-109.

Witten, I. H., Frank, E., Hall, M. A. and Christopher, J. P. (2017), *Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Cambridge.

Yang, X.S. (2019), *Introduction to Algorithms for Data Mining and Machine Learning*, Elsevier, London.