Journal of Statistical Sciences, Spring and Summer, 2025
Vol. 19, No. 1, pp 205-226
DOI: 10.61186/jss.19.1.1

Multi-class Support Vector Machine with Uncertain Inputs by Probabilistic Constraints

Mohammadi, T.[1], Jabbari, H.[1], Effati, S.[2]
[1]Department of Statistics, Ferdowsi University of Mashhad, Mashhad , Iran.
[2]Department of Applied Mathematics, Ferdowsi University of Mashhad, Mashhad, Iran.

Corresponding author: Hadi Jabbari, jabbarinh@um.ac.ir

Received: 12/11/2024 **Revised:** 24/1/2025 **Accepted and Published Online:** 27/11/2025.

Introduction

Support Vector Machine (SVM) is a robust supervised learning algorithm primarily used for classification and regression. It is initially invented for binary classification. However, many real-world problems involve multi-class classification, where the goal is to distinguish between more than two classes. Various strategies have been proposed to address this challenge, including decomposition methods that reduce the multi-class problem into multiple binary subproblems, as well as direct approaches that optimize for all classes simultaneously. Two of the methods that distinct all classes simultaneously are Cramer-Singer and simplified multi-class SVM (SimMSVM). The challenge in multi-class SVM lies in accurate inputs while the data are uncertain in the real world due to measurement errors. Some approaches include Crammer and Singer's method or SimMSVM. Despite having nice strengths and performance, they consider inputs in an accurate state. This work has explored the theoretical foundations and practical implementations of SimMSVM in which the inputs are random variables and the constraints are probabilistic. The constraints have been released from the probabilistic state by leveraging statistical theories, and the expectation value has been entered into the model. Hence, the model has been converted into ordinary quadratic programming. This paper will show that the model with uncertain inputs has a better procedure than the traditional one.

Material and Methods
This paper applies the expected value to release the constraints from the probabilistic state. Then, the sample mean is used to estimate the expected value in the unknown population state. Also, the Monte Carlo method is implemented to simulate the data, and then the bootstrap resampling method is used to generate random samples to estimate the expected value. Moreover, the confusion matrix statistical indices and accuracy criterion are used to evaluate the proposed and competing models.

Results and Discussion
Using three kernel functions, linear, polynomial, and Gaussian, we investigate the accuracy values of the three models. In the simulation subsection, we examine the performance of the three proposed models, basic and Cramer-Singer, using the statistical indices of mean and standard deviation, and we compare the three models using the Friedman test statistic. Also, in the real example subsection, we evaluate these models in different parameter ranges for three real data sets of seeds, ecoli, and leukemia.

Conclusion
Given that most of the data we encounter in the real world is not precise and certain and has inherent flaws due to measurement errors, this paper proposes a model that addresses this issue. The results show that the model based on random inputs is far better than the models based on definite and precise inputs.

Keywords: Multi–class SVM, Stochastic input, Probabilistic constraint, Expectation.

Mathematics Subject Classification (2010): 62H30, 62G35.

مجله علوم آماری، بهار و تابستان ۱۴۰۴
جلد ۱۹، شماره ۱، ص ۲۰۵ -- ۲۲۶
DOI: 10.61186/jss.19.1.1
مقاله پژوهشی

ماشین بردار پشتیبان چند–کلاسه با ورودی‌های تصادفی با محدودیت‌های احتمالی

تارا محمدی[۱]، هادی جباری نوقابی[۱] و سهراب عفتی[۲]

[۱]گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد
[۲]گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

نویسنده مسئول : هادی جباری نوقابی، jabbarinh@um.ac.ir

تاریخ دریافت: ۱۴۰۳/۸/۲۲ تاریخ بازنگری: ۱۴۰۳/۱۱/۵ تاریخ پذیرش و انتشار: ۱۴۰۳/۱۱/۸

چکیده: ماشین بردار پشتیبان به عنوان یک الگوریتم با نظارت در ابتدا برای حالت دودویی ابداع شد، سپس به علت کاربردهای آن، الگوریتم های چند–کلاسه نیز طراحی شدند و همچنان به عنوان یک پژوهش در حال بررسی است. اخیرا مدل‌هایی برای بهبود روش‌های چند–کلاسه ارائه گردیده است. اغلب آن‌ها حالتی را بررسی می‌کنند که در آن ورودی‌ها غیرتصادفی هستند در حالی که در دنیای واقعی با داده‌های غیرقطعی و نادقیق مواجه هستیم. لذا در این مقاله به بررسی مدلی پرداخته شده است که در آن ورودی‌ها تصادفی و محدودیت‌های مسئله نیز احتمالی هستند. با استفاده از قضایای آماری و با استفاده از امیدریاضی، محدودیت‌های مسئله از حالت احتمالی خارج شده است. سپس از روش برآورد گشتاوری، برای برآورد امیدریاضی استفاده شده است. با استفاده از شبیه‌سازی مونت کارلو به تولید داده‌های مصنوعی پرداخته و از روش بازنمونه‌گیری بوت استرپ نمونه‌ها را به عنوان ورودی به مدل داده و دقت مدل مورد بررسی قرار گرفته است. در نهایت مدل پیشنهادی را با داده‌های واقعی آموزش داده و دقت آن با شاخص‌های آماری ارزیابی شده است. نتایج حاصل از شبیه‌سازی و مثال واقعی برتری کارایی مدل پیشنهادی را بر مدل‌های مبتنی بر ورودی‌های قطعی نشان می‌دهد.

واژه‌های کلیدی: ماشین بردار پشتیبان چند–کلاسه، ورودی تصادفی، محدودیت احتمالی، امیدریاضی.
کد موضوع‌بندی ریاضی (۲۰۱۰): 62H30، 62G35.

۱ مقدمه

ماشین بردار پشتیبان[1] یک الگوریتم با نظارت و از شاخه‌های یادگیری ماشین به شمار می‌رود. این الگوریتم در ابتدا توسط وپنیک (۱۹۹۵) برای کلاس‌های دودویی طراحی شد. او به همراه کورتس (۱۹۹۵) مدل خطی را به حالت غیرخطی ارتقا دادند. مدل ارائه شده توسط وپنیک فقط به کلاس‌های دودویی محدود می‌شد، در حالی که اغلب مواقع با داده‌های بیش از دو کلاس سروکار داریم. در میان مدل‌های چند–کلاسه، برخی کلاس‌ها را به صورت دودویی و برخی به صورت همزمان تفکیک می‌کنند. مدل‌های ارائه شده توسط شیلکوپ و همکاران (۱۹۹۵)، وپنیک (۱۹۹۹)، نِر و همکاران (۱۹۹۰)، پلات و همکاران (۱۹۹۹) و هستی و تیبشیرانی (۱۹۹۷) از جمله مدل‌های چند–کلاسه‌ای هستند که کلاس‌ها را به صورت دودویی تفکیک می‌کنند. همچنین مدل‌های ارائه شده توسط وپنیک (۱۹۹۹) و وستون و واتکینز (۱۹۹۹) از جمله مدل‌هایی هستند که کلاس‌ها را به صورت همزمان در نظر می‌گیرند. گونن و همکاران (۲۰۰۸) مدلی به نام ماشین بردار پشتیبان احتمال پسین چند–کلاسه ارائه کرد. آن‌ها مقادیر احتمال پسینی را برای هر کلاس در ساختار محدودیت مدل وستون و واتکینز (۱۹۹۹) در نظر گرفتند. در این مدل، اثر نقاط پرت کاهش می‌یابد. همچنین در روش آن‌ها، با کاهش تعداد ماشین‌های بردار پشتیبان، پیچیدگی مکانی و زمانی کاهش می یابد، زیرا محاسبات کمتری مورد نیاز است.

مدلی که توسط هیی و همکاران (۲۰۱۲) منتشر شده است یک ماشین بردار پشتیبان چند–کلاسه ساده شده را ارائه می‌کند که مدل کرامر و سینگر (۲۰۰۱) را با معرفی یک کران خطای جدید ارتقا می‌دهد و سبب افزایش دقت و عملکرد بهتر می‌گردد. همچنین مدل دیگری به نام ماشین بردار پشتیبان با محدودیت‌های احتمالی برای ورودی‌های نامعین توسط یزدی و همکاران (۲۰۰۷) پیشنهاد گردیده است که بر اساس توابع چگالی احتمال عمل می کند. در مدل آن‌ها ابتدا با لحاظ کردن یک خطای تصادفی در محدودیت مسئله اولیه، از حالت قطعی به حالت غیرقطعی تبدیل شد سپس با استفاده از روابط آماری و استفاده از تعریف تابع توزیع محدودیت را از حالت احتمالی خارج کرد. در نهایت نتایج نشان داد کارایی مدل آن‌ها نسبت به مدل استاندارد با ورودی‌های قطعی بالاتر است. علاوه‌براین مدل‌ها، یک روش چند–کلاسه با محدودیت‌های احتمالی و داده‌های غیرقطعی توسط بوش و همکاران (۲۰۱۳) ارائه شد که در آن از طبقه‌بندی کننده‌های تصادفی دودویی و تصادفی چند–کلاسه یک در برابر همه استفاده شد. کارایی مدل پیشنهادی با عملکرد بهتری نسبت به مدل قطعی به دست آمد. همچنین یک مدل احتمالی توسط لی و همکاران (۲۰۱۳) برای داده‌های متاثر از اغتشاش منتشر شد که در آن مسئله اولیه بر اساس تابع چگالی احتمال وزن‌ها است. آن‌ها نشان دادند که در حضور اغتشاش می تواند کارایی بیشتری نسبت به حالت معمولی داشته باشد. مدل چند–کلاسه جدید دیگری توسط چِن و همکاران (۲۰۱۶) پیشنهاد شد. در این مدل با به کارگیری یک الگوریتم متوالی برای کران خطا، سبب افزایش سرعت روند آموزش گردید. عباس زاده و عفتی (۲۰۱۸) یک رگرسیون بردار پشتیبان با ورودی و خروجی تصادفی پیشنهاد کردند که محدودیت‌های آن به صورت احتمالی است. محدودیت‌های احتمالی سبب استحکام مدل و کارایی بهتری نسبت به مدل استاندارد گردید. مدل ارائه شده توسط

[1] Support Vector Machine

عباس زاده و عفتی (۲۰۱۹) یک مدل دودویی بر اساس محدودیت‌های احتمالی است. در آن نمونه‌های آموزشی به عنوان متغیرهای تصادفی و محدودیت مسئله نیز به صورت احتمالی در نظر گرفته شد. در مدل آن‌ها با استفاده از روش برآورد گشتاوری محدودیت مسئله از حالت احتمالی خارج گردید. نتایج نشان داد که کارایی آن به مراتب بهتر از مدل استاندارد است.

مدل دیگری توسط دُن و یاکوب (۲۰۲۰) پیشنهاد شده است که به دلیل پراکندگی داده‌ها در فضای با ابعاد بالا انجام شد. این کار با تقسیم کل مجموعه داده آموزشی به زیرمجموعه‌های مجزا بر اساس تقسیمات دودویی انجام شده است. در فرایند مدل‌سازی با حذف بعضی تقسیمات و تکرار این فرایند تنها دو تصمیم دودویی وجود خواهد داشت. آزمایشات نشان داد مدل آن‌ها عملکرد بهتری نسبت به بعضی از مدل‌های فعلی دارد. همچنین یک مدل چند–کلاسه با نام حداقل مربعات رگرسیون بردار پشتیبان با k کلاس توسط موسایی و هِلادیک (۲۰۲۲) پیشنهاد شده است. برای به حداقل رساندن متغیرهای کمکی در مسئله اولیه، محدودیت‌های نامساوی را با محدودیت‌های مساوی جایگزین کرده و ℓ_2 را به جای ℓ_1 اعمال کرده است. این تغییر باعث شد الگوریتم سرعت بیشتر و عملکرد بهتری داشته باشد.

المعصومه و همکاران (۲۰۲۲) داده‌های هیستوگرامی را با ماشین بردار پشتیبان احتمالی مبتنی بر امیدریاضی و مبتنی بر کمیت و با محدودیت‌های احتمالی ارائه کردند. روش آن‌ها برای هر دو حالت دودویی و چند–کلاسه انجام شد که عملکرد بهتری نسبت به الگوی معمولی داشت. یک مدل چند–کلاسه توسط صالح و همکاران (۲۰۲۳) با استفاده از طرح یک در برابر همه پیشنهاد شده است. این مدل، مسئله را به زیرمسئله های دودویی تقسیم می‌کند که در آن دو ابر صفحه غیرموازی را پیدا می‌کند. مدل آن‌ها به خوبی توانست در تقسیم‌بندی کلاس‌ها کارایی بیشتری نسبت به بعضی مدل‌های پیش از خود به دست آورد. در میان مدل‌های ارائه شده اخیر، بنیتز و همکاران (۲۰۲۴) خروجی‌های احتمالی را از طریق برآورد نقطه‌ای تولید کردند. مدل آن‌ها به گونه‌ای طراحی شده است که نسبت به هزینه حساس است. همچنین مقادیر احتمالات از طریق برآوردهای بوت استرپ انجام می‌شود. برای بهبود حساسیت، آن‌ها از دو رویکرد مختلف استفاده کردند، روش اول مبتنی بر بوت استرپ که اندازه‌گیری حساسیت توسط احتمالات کلاس پسین کنترل می‌شود. روش دوم مشابه اول انجام می شود، اما طبقه‌بندی متفاوتی را در نظر می‌گیرد. اخیرا مدلی توسط موسوی و گلعلی‌زاده (۱۴۰۲) ارائه شد که در آن به دلیل چالش ابعاد روش‌های تحلیل داده‌های ژنی برای ارزیابی سرطان از تکنیک‌های کاهش ابعاد و ماشین بردار پشتیبان برای حل این مشکل استفاده شد. در مدل آن‌ها به نام ماشین بردار پشتیبان تصادفی خوشه‌ای با نمایش تصادفی فضای ویژگی و ترکیب طبقه‌بندی‌کننده‌ها، دقت بالایی در شناسایی ژن‌های مهم و پیش‌بینی سرطان پروستات ارائه شد. بیشتر اوقات در زندگی واقعی با داده‌های معلوم و دقیقی مواجه نیستیم و در عمل به دلایلی مانند خطای اندازه‌گیری، خطای نمونه‌گیری با داده‌هایی که دقیق و قطعی نیستند، سروکار داریم. از سوی دیگر، در موارد مکرر، داده‌ها بیش از دو کلاس هستند، بنابراین تصمیم گرفتیم هر دو مشکل را به طور همزمان برطرف کنیم. انگیزه اصلی این مقاله این است که برای حل این مشکلات، مدلی را برای ورودی های تصادفی و تقسیم‌بندی به سه کلاس و بیشتر ارائه کند.

در این مقاله یک مدل چند–کلاسه با ورودی‌های تصادفی و محدودیت‌های احتمالی در نظر گرفتیم. در بخش

۲ ساختار مدل را تعریف و با استفاده از قضایای آماری، آن را از حالت احتمالی خارج می‌کنیم. سپس با روش‌های آماری به برآورد پارامترهای مدل می‌پردازیم. در بخش ۳ به شبیه‌سازی و در بخش ۴ به ارائه مثال واقعی جهت بررسی عملکرد مدل با نمونه‌هایی از مجموعه داده‌های واقعی از پایگاه داده یادگیری ماشین دووآ و گرَف (۲۰۱۹) می‌پردازیم. بطور کلی نتایج شبیه‌سازی و مثال واقعی نشان می‌دهد که کارایی مدل پیشنهادی به مراتب بهتر از مدل‌های با ورودی‌های قطعی است.

۲ ساختار مدل بر اساس محدودیت‌های احتمالی

این مقاله نسخه بهبودیافته مدل ارائه شده توسط هیی و همکاران (۲۰۱۲) را ارائه می‌کند که حالت بهبودیافته مدل کرامر و سینگر (۲۰۰۱) است، با این تفاوت که مدل کرامر و سینگر (۲۰۰۱) مجموعه فشرده‌ای از محدودیت‌ها است که تعداد متغیرها در مسئله دوگان $\ell \times k$ است، و این مقدار حتی برای مجموعه داده‌های کوچک پیچیدگی محاسباتی بالایی دارد. الگوی ارائه شده توسط هیی و همکاران (۲۰۱۲) روشی ساده برای کاهش محدودیت پیشنهاد می‌کند. در الگوی هیی و همکاران (۲۰۱۲) یک مسئله برنامه‌ریزی درجه دوم با ℓ متغیر خواهیم داشت. اکنون مسئله اولیه کرامر و سینگر (۲۰۰۱) را بیان می‌کنیم، مجموعه داده‌های آموزشی $S = \{(x_۱, y_۱), \ldots, (x_\ell, y_\ell)\}$ را در نظر بگیرید، که در آن $x_i \in \mathrm{R}^n$ نمونه iام با n ویژگی از کلاس $y_i = ۱, \ldots, k$ است. بنابراین مسئله اولیه به صورت

$$\begin{aligned} \min \quad & \frac{۱}{۲} \sum_{m=۱}^{k} \mathbf{w}_m^T \mathbf{w}_m + C \sum_{i=۱}^{\ell} \xi_i \\ \text{s.t.} \quad & \mathbf{w}_{y_i}^T \phi(x_i) - \mathbf{w}_t^T \phi(x_i) \geq ۱ - \delta_{y_i,t} - \xi_i, \\ & \xi_i \geq ۰, i = ۱, \ldots, \ell, t = ۱, \ldots, k \neq y_i. \end{aligned}$$

بیان می‌شود، که در آن $C > ۰$ پارامتر جریمه و $\xi_i \geq ۰$ متغیرهای کمکی هستند و

$$\delta_{ij} = \begin{cases} ۱ & i = j, \\ ۰ & i \neq j \end{cases}$$

دلتای کرونکر و $\mathbf{w}^T = [w_۱, \cdots, w_n]^T$ بردار وزن است. همچنین $\phi(\cdot) : R^n \to R^{n'}$ یک تبدیل غیرخطی است که از طریق ضرب نقطه‌ای، فضای ورودی را از n به فضای ویژگی با ابعاد بالاتر n' تبدیل می‌کند. تابع تصمیم[1]

[1] Decision function

بهینه در رابطه

$$\arg\max_{m} f_m(\boldsymbol{x}) = \mathbf{w}_m^T \phi(\boldsymbol{x}). \tag{۱}$$

تعریف می‌شود. اگرچه مدل چند–کلاسه کرامر و سینگر (۲۰۰۱) مجموعه فشرده‌ای از محدودیت‌ها را ارائه می‌دهد و تعداد متغیرها در مسئله دوگان آن‌ها $(\ell \times k)$ است که منجر به پیچیدگی محاسباتی زیادی می‌شود بنابراین مدل ارائه شده توسط هیی و همکاران (۲۰۱۲) شیوه کرامر و سینگر (۲۰۰۱) را دنبال می‌کند و یک روش ساده‌سازی شده برای کاهش محدودیت‌های کرامر و سینگر (۲۰۰۱) پیشنهاد می‌کند. با انجام این کار، به یک مسئله برنامه‌ریزی درجه دوم با ℓ متغیر تبدیل می‌شود. فرض کنید برای نمونه آموزشی x_i عبارت

$$\zeta_{i,m} = ۱ - f_{y_i}(x_i) - f_m(x_i), \quad m \in \{۱, \ldots, k\} \setminus y_i.$$

نشان دهنده حاشیه زوجی نادرست بین کلاس واقعی y_i و هر کلاس دیگر m است (وستون و واتکینز، ۱۹۹۹) که در آن به جای "۱" از "۲" استفاده شده است. در تابع زیان وستون و واتکینز (۱۹۹۹) تمامی حاشیه‌های زوجی نادرست $(\zeta_{i,m} > ۰)$ به شکل $\xi_i^{(۱)} = \sum_{m \neq y_i} [\zeta_{i,m}]_+$ جمع می‌شود، که در آن $[\cdot]_+ \equiv max(\cdot, ۰)$. اما در مدل کرامر و سینگر (۲۰۰۱) زیان نمونه تنها حداکثر حاشیه زوجی نادرست به حالت $\xi_i^{(۲)} = [\max_{m \neq y_i} \zeta_{i,m}]_+$ بیان می‌شود. برای کاهش اندازه مسئله، تعداد محدودیت‌ها در هیی و همکاران (۲۰۱۲) به جای $(\ell \times k)$ باید متناسب با ℓ باشد. لذا آن‌ها کران جدید $\xi_i^{(۳)} = [\frac{۱}{k-۱} \sum_{m=۱, m \neq y_i}^{k} \zeta_{i,m}]_+$ را معرفی کردند. بنابراین رابطه بین این زیان‌ها $\xi_i^{(۱)} \geq \xi_i^{(۲)} \geq \xi_i^{(۳)}$ است. در نهایت مدل ارائه شده توسط هیی و همکاران (۲۰۱۲) در رابطه

$$\begin{aligned} \min \quad & \frac{۱}{۲} \sum_{m=۱}^{k} \mathbf{w}_m^T \mathbf{w}_m + C \sum_{i=۱}^{\ell} \xi_i \\ \text{s.t.} \quad & \mathbf{w}_{y_i}^T \phi(x_i) - \frac{۱}{k-۱} \sum_{m \neq y_i} \mathbf{w}_m^T \phi(x_i) \geq ۱ - \xi_i, \\ & \xi_i \geq ۰, i = ۱, \ldots, \ell. \end{aligned} \tag{۲}$$

بیان می‌شود. برای به دست آوردن مقادیر بهینه متغیرها، تابع دوگان لاگرانژ را به صورت

$$\begin{aligned} L(\mathbf{w}, \xi, \alpha, \lambda) = & \frac{۱}{۲} \sum_{m=۱}^{k} \mathbf{w}_m^T \mathbf{w}_m + C \sum_{i=۱}^{\ell} \xi_i - \sum_{i=۱}^{\ell} \alpha_i \xi_i - \sum_{i=۱}^{\ell} \lambda_i \xi_i \\ & - \sum_{i:m=y_i} \alpha_i \mathbf{w}_{y_i}^T \phi(\mathbf{x}_i) + \frac{۱}{k-۱} \sum_{i:m \neq y_i} \alpha_i \mathbf{w}_m^T \phi(\mathbf{x}_i) + \sum_{i=۱}^{\ell} \alpha_i \end{aligned}$$

تعریف می‌کنیم، که در آن $\alpha_i \geq ۰$ و $\lambda_i \geq ۰$ ضرایب لاگرانژ هستند. مقادیر بهینه متغیرهای مسئله اولیه به ترتیب

$$\mathbf{w}_m = \sum_{i:y_i=m} \alpha_i \mathbf{x}_i - \frac{۱}{k-۱} \sum_{i:y_i \neq m} \alpha_i \mathbf{x}_i, \tag{۳}$$

$$C\mathbf{e} = \boldsymbol{\alpha} + \boldsymbol{\lambda}, \tag{۴}$$

هستند، که در آن $\mathbf{e}$ بردار یکه است. با جایگذاری روابط (۳) و (۴) در (۱) تابع تصمیم

$$f_m(\mathbf{x}) = \sum_{i:y_i=m} \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) - \frac{۱}{k-۱} \sum_{i:y_i \neq m} \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) \tag{۵}$$

حاصل خواهد شد، که در آن $K(\mathbf{x}_i, \mathbf{x}_j) = \phi^T(\mathbf{x}_i)\phi(\mathbf{x}_j)$ تابع هسته است. با ساده‌کردن رابطه (۵) تابع تصمیم بهینه به صورت

$$\arg\max_m f_m(\mathbf{x}) = \arg\max_m \sum_{i:y_i=m} \alpha_i K(\mathbf{x}_i, \mathbf{x}_j), \tag{۶}$$

تبدیل می‌شود. اکنون بر اساس رابطه (۲) در حالتی که ورودی‌ها تصادفی هستند، مدل پیشنهادی ارائه می‌شود. پس مجموعه داده‌های آموزشی به شکل $\tilde{S} = \{(\mathbf{X}_۱, y_۱), \ldots, (\mathbf{X}_\ell, y_\ell)\}$ تبدیل می‌شود، که در آن $\mathbf{X}_i = [X_۱, \ldots, X_n]^T$ بردار متغیر تصادفی ورودی است. همچنین در مسئله برنامه‌ریزی غیرقطعی یک حالت در نظر گرفتن محدودیت‌های مسئله به صورت احتمالی است. برای ضمانت تعمیم مدل (۲)، با اضافه کردن $b_m^۲$ به عبارت تنظیم کننده $\mathbf{w}_m^T\mathbf{w}$، تابع هدف مسئله اولیه در رابطه

$$\begin{aligned}
\min \quad & \frac{۱}{۲} \sum_{m=۱}^{k} \left(\mathbf{w}_m^T \mathbf{w}_m + b_m^۲\right) + C \sum_{i=۱}^{\ell} \xi_i \\
\text{s.t.} \quad & Pr\left((\mathbf{w}_{y_i}^T \phi(\mathbf{X}_i) + b_{y_i}) - \frac{۱}{k-۱} \sum_{m \neq y_i} (\mathbf{w}_m^T \phi(\mathbf{X}_i) + b_m) \geq ۱ - \xi_i\right) \geq \gamma, \\
& \xi_i \geq ۰, i = ۱, \ldots, \ell,
\end{aligned}$$

بیان می‌شود، که در آن γ احتمال به ازای اثر نمونه iام است.

قضیه ۱. (بارنت و همکاران، ۲۰۰۲) اگر U متغیری تصادفی در بازه $[c,a]$ با تابع توزیع تجمعی $F(u) = P(U \leq u)$ باشد، نامساوی $|P(U \leq u) - \frac{a-E(U)}{a-c}| \leq \frac{۱}{۲} + \frac{|u-\frac{c+a}{۲}|}{a-c}$ یا

$$\left| P(U \geq u) - \frac{E(U)-c}{a-c} \right| \leq \frac{۱}{۲} + \frac{\left|u-\frac{c+a}{۲}\right|}{a-c}. \tag{۷}$$

برقرار است.

قضیه ۲. فرض کنید $\mathbf{X}_i = [X_{i۱}, \ldots, X_{in}]^T$ برداری تصادفی با امیدریاضی معلوم $E(\mathbf{X}_i) = [E(X_{i۱}), \ldots, E(X_{in})]^T$ باشد. شرط کافی برای برقراری محدودیت

$$Pr\big((\mathbf{w}_{y_i}^T \phi(\mathbf{X}_i) + b_{y_i}) - \frac{۱}{k-۱} \sum_{m \neq y_i} (\mathbf{w}_m^T \phi(\mathbf{X}_i) + b_m) \geq ۱ - \xi_i\big) \geq \gamma,$$

عبارت است از

$$(\mathbf{w}_{y_i}^T E(\phi(\mathbf{X}_i)) + b_{y_i}) - \frac{۱}{k-۱} \sum_{m \neq y_i} (\mathbf{w}_m^T E(\phi(\mathbf{X}_i)) + b_m) \geq ۱ - \xi_i + ۲a\gamma.$$

برهان: فرض کنید $U_i = (\mathbf{w}_{y_i}^T \phi(\mathbf{X}_i) + b_{y_i}) - \frac{۱}{k-۱} \sum_{m \neq y_i} (\mathbf{w}_m^T \phi(\mathbf{X}_i) + b_m)$، برای یک مقدار معین u بطوریکه $-u \leq U_i - \xi_i \leq U_i + \xi_i \leq u$ باشد و با استفاده از رابطه (۷) و $c = -a$، نامساوی

$$-\frac{۱}{۲} - \frac{|u|}{۲a} + \frac{E(U_i+\xi_i)+a}{۲a} \leq Pr(U_i + \xi_i \geq u) \leq \frac{۱}{۲} + \frac{|u|}{۲a} + \frac{E(U_i+\xi_i)+a}{۲a}.$$

به دست می‌آید. با در نظر گرفتن $u = ۱$، داریم:

$$-\frac{۱}{۲a} + \frac{E(U_i)+\xi_i}{۲a} \leq Pr(U_i + \xi_i \geq ۱) \leq ۱ + \frac{۱}{۲a} + \frac{E(U_i)+\xi_i}{۲a},$$

و در نهایت چون $P(U_i + \xi_i \geq ۱) \geq \gamma$ داریم $-\frac{۱}{۲a} + \frac{E(U_i)+\xi_i}{۲a} \geq \gamma$ یا $E(U_i) \geq ۱ - \xi_i + ۲a\gamma$. بنابراین محدودیت مسئله اولیه از حالت احتمالی خارج و به

$$(\mathbf{w}_{y_i}^T E(\phi(\mathbf{X}_i)) + b_{y_i}) - \frac{۱}{k-۱} \sum_{m \neq y_i} (\mathbf{w}_m^T E(\phi(\mathbf{X}_i)) + b_m) \geq ۱ - \xi_i + ۲a\gamma.$$

تبدیل می‌شود.

اکنون برای به دست آوردن تابع دوگان با استفاده از تابع لاگرانژ[1]، مسئله اولیه نسبت به متغیرهای $\mathbf{w}$، b و ξ_i و ضرایب لاگرانژ مینیمم می‌شود. برای به دست آوردن مقادیر بهینه متغیرها، تابع

$$
\begin{aligned}
L(\mathbf{w}, b, \xi, \alpha, \lambda) = \frac{۱}{۲}\sum_{m=۱}^{k}\left(\mathbf{w}_m^T\mathbf{w}_m + b_m^۲\right) + C\sum_{i=۱}^{\ell}\xi_i - \sum_{i=۱}^{\ell}\alpha_i\xi_i - \sum_{i=۱}^{\ell}\lambda_i\xi_i \\
- \sum_{i:m=y_i}\alpha_i\left(\mathbf{w}_{y_i}^T E(\phi(\mathbf{X}_i)) + b_{y_i}\right) + \frac{۱}{k-۱}\sum_{i:m\neq y_i}\alpha_i\left(\mathbf{w}_m^T E(\phi(\mathbf{X}_i)) + b_m\right) \\
+ \sum_{i=۱}^{\ell}\alpha_i + ۲a\sum_{i=۱}^{\ell}\alpha_i\gamma.
\end{aligned}
$$

تعریف می‌شود. از مشتق L نسبت به $\mathbf{w}_m$، b_m و ξ شرایط لازم بهینگی به صورت

$$
\begin{aligned}
&\mathbf{w}_m = \sum_{i:y_i=m}\alpha_i E(\phi(\mathbf{X}_i)) - \frac{۱}{k-۱}\sum_{i:y_i\neq m}\alpha_i E(\phi(\mathbf{X}_i)), \\
&b_m = \sum_{i:y_i=m}\alpha_i - \frac{۱}{k-۱}\sum_{i:y_i\neq m}\alpha_i = ۰, \qquad (۸) \\
&C\mathbf{e} = \boldsymbol{\alpha} + \boldsymbol{\lambda} = ۰.
\end{aligned}
$$

به دست می‌آید. با جایگذاری روابط (۸) در تابع لاگرانژ برای $i \neq j = ۱, \ldots \ell$ و بسط تابع لاگرانژ داریم

$$
\begin{aligned}
L(\alpha) = \frac{۱}{۲}\sum_{i:y_i=m}\alpha_i^۲ E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) \\
- \frac{۱}{k-۱}\sum_{i:y_i=m}\alpha_i\sum_{i:y_i\neq m}\alpha_i E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) \\
+ \frac{۱}{۲(k-۱)^۲}\sum_{i:y_i\neq m}\alpha_i^۲ E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) \\
+ \frac{۱}{۲}\sum_{i:y_i=m}\alpha_i^۲ - \frac{۱}{k-۱}\sum_{i:y_i=m}\alpha_i\sum_{i:y_i\neq m}\alpha_i + \frac{۱}{۲(k-۱)^۲}\sum_{i:y_i\neq m}\alpha_i^۲,
\end{aligned}
$$

[1] Lagrangian function

$$
\begin{aligned}
. \quad & - \sum_{i:y_i=m} \alpha_i^2 E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) + \frac{1}{k-1} \sum_{i:y_i\neq m} \alpha_i^2 E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) \\
& - \sum_{i:y_i=m} \alpha_i^2 + \frac{1}{k-1} \sum_{i:y_i\neq m} \alpha_i^2 + \frac{1}{k-1} \sum_{i:y_i=m} \alpha_i^2 E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) \\
& - \frac{1}{(k-1)^2} \sum_{i:y_i\neq m} \alpha_i^2 E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) + \frac{1}{k-1} \sum_{i:y_i=m} \alpha_i^2 \\
& - \frac{1}{(k-1)^2} \sum_{i:y_i\neq m} \alpha_i^2 + \sum_{i=1}^{\ell} \alpha_i + 2a \sum_{i=1}^{\ell} \alpha_i \gamma. \qquad (۹)
\end{aligned}
$$

در نهایت با خلاصه کردن رابطه (۹)، مسئله دوگان به به صورت

$$
\begin{aligned}
\max \; & -\frac{1}{2}\frac{k}{k-1} \sum_{i:y_i=m} \alpha_i^2 \left(E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) + \mathbf{e}\right) \\
& - \frac{1}{2}\frac{-k}{(k-1)^2} \sum_{i:y_i\neq m} \alpha_i^2 \left(E^T(\phi(\mathbf{X}_i))E(\phi(\mathbf{X}_j)) + \mathbf{e}\right) + \sum_{i=1}^{l} \alpha_i (1 + 2a\gamma) \\
\text{s.t.} \quad & \mathbf{0} \leq \boldsymbol{\alpha} \leq C\boldsymbol{e}. \qquad (۱۰)
\end{aligned}
$$

تبدیل می‌شود، که در آن $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_n]^T$ بردار ضرایب لاگرانژ است. همچنین برای (۶) داریم:

$$
\arg\max_m f_m(\mathbf{x}) = \arg\max_m \sum_{i:y_i=m} \boldsymbol{\alpha} K(E(\mathbf{X}_i), E(\mathbf{X}_j)). \qquad (۱۱)
$$

که در آن

$$
K(E(\mathbf{X}_i), E(\mathbf{X}_j)) = \phi^T(E(\mathbf{X}_i))\phi(E(\mathbf{X}_j)). \qquad (۱۲)
$$

تابع هسته است. با تعریف ماتریس هسین [1] به صورت

$$
\mathbf{G} = \begin{cases} \frac{k}{k-1} K(E(\mathbf{X}_i), E(\mathbf{X}_j)); & y_i = m \\ \frac{-k}{(k-1)^2} K(E(\mathbf{X}_i), E(\mathbf{X}_j)); & y_i \neq m \end{cases}
$$

[1] Hessian matrix

مسئله دوگان (۱۰)، به صورت

$$\begin{aligned} \max \quad & -\frac{۱}{۲}\boldsymbol{\alpha}^T\mathbf{G}\boldsymbol{\alpha} + \mathbf{e}(۱+۲a\gamma)\boldsymbol{\alpha} \\ \text{s.t.} \quad & \mathbf{0} \le \boldsymbol{\alpha} \le C\mathbf{e}. \end{aligned} \tag{۱۳}$$

تبدیل می‌شود. برای تابع تصمیم (۱۱) داریم

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}_m^T E(\phi(\mathbf{X}_i)) + b_m \\ &= \Big(\sum_{i:y_i=m} \alpha_i E(\phi(\mathbf{X}_i)) - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i E(\phi(\mathbf{X}_i)) \Big) E(\phi(\mathbf{X}_j)) \\ &+ \sum_{i:y_i=m} \alpha_i - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i, \end{aligned}$$

که در نهایت:

$$\begin{aligned} f(\mathbf{x}) = & \sum_{i:y_i=m} \alpha_i K(E(\mathbf{X}_i), E(\mathbf{X}_j)) - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i K((\mathbf{X}_i), E(\mathbf{X}_j)) \\ & + \sum_{i:y_i=m} \alpha_i - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i. \end{aligned} \tag{۱۴}$$

چون میانگین جامعه نامعلوم است، باروش گشتاوری برآورد می‌شود. اگر نمونه‌های

$$\mathbf{x}_۱ = [x_{۱۱}, \ldots, x_{۱n}]^T, \mathbf{x}_۲ = [x_{۲۱}, \ldots, x_{۲n}]^T, \ldots, \mathbf{x}_t = [x_{t۱}, \ldots, x_{tn}]^T.$$

وجود داشته باشد، بردار برآورد امیدریاضی $\bar{\mathbf{x}}_j = \left[\frac{۱}{N_j}\sum_{t=۱}^{N_j} x_{t۱}, \ldots, \frac{۱}{N_j}\sum_{t=۱}^{N_j} x_{tn}\right]$ است، که در آن N_jها حجم نمونه‌های ممکن است. بنابراین تابع هسته (۱۲)، به $K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) = \phi^T(\bar{\mathbf{x}}_i)\phi(\bar{\mathbf{x}}_j)$ و بردار برآورد وزن‌ها در رابطه (۸) به $\widehat{\mathbf{w}}_m = \sum_{i:y_i=m} \alpha_i\phi(\bar{\mathbf{x}}_i) - \frac{۱}{k-۱}\sum_{i:y_i\neq m} \alpha_i\phi(\bar{\mathbf{x}}_i)$ تبدیل می‌شود.

$$\begin{aligned} \max & -\frac{۱}{۲}\frac{k}{k-۱} \sum_{\{i:y_i=m\}} \alpha_i^۲ \left(K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) + \boldsymbol{e}\right) \\ & -\frac{۱}{۲}\frac{-k}{(k-۱)^۲} \sum_{\{i:y_i\neq m\}} \alpha_i^۲ (K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) + \mathbf{e}) + \sum_{i=۱}^{\ell} \alpha_i(۱+۲a_i) \\ \text{s.t.} \quad & \mathbf{0} \le \boldsymbol{\alpha} \le C\boldsymbol{e}. \end{aligned}$$

بیان می‌شود و در نهایت برای برآورد مدل (۱۳) و همچنین برآورد تابع تصمیم بهینه (۱۴) به ترتیب داریم

$$\begin{aligned} \max \quad & -\frac{۱}{۲}\boldsymbol{\alpha}^T\widehat{\boldsymbol{G}}\boldsymbol{\alpha} + \mathbf{e}(۱ + ۲a\gamma)\boldsymbol{\alpha} \\ \text{s.t.} \quad & \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{e}. \end{aligned}$$

و

$$\begin{aligned} f(\mathbf{x}) = & \sum_{i:y_i=m} \alpha_i K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) \\ & + \sum_{i:y_i=m} \alpha_i - \frac{۱}{k-۱} \sum_{i:y_i\neq m} \alpha_i. \end{aligned}$$

۳ مطالعه شبیه‌سازی

فرآیند شبیه‌سازی با روش مونت کارلو[1] در (موونی، ۱۹۹۷) انجام شده است. روش بازنمونه‌گیری بوت استرپ به کار برده شد و از هر کلاس نمونه‌برداری شد، سپس میانگین داده‌ها محاسبه شده و دقت مدل به دست آمد. همچنین از سه تابع هسته خطی، چندجمله‌ای و گاوسی

$$K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) = \bar{\mathbf{x}}_i^T\bar{\mathbf{x}}_j, K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) = (\bar{\mathbf{x}}_i^T\bar{\mathbf{x}}_j + ۱)^p, K(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) = \exp(-\mu\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|^۲)$$

استفاده شد. فرایند شبیه‌سازی مطابق با گام‌های زیر انجام می‌شود.

گام ۱: با استفاده از روش بازنمونه‌گیری بوت استرپ از هر کلاس نمونه‌برداری می‌شود.

گام ۲: میانگین نمونه‌ها محاسبه و به عنوان ورودی در مدل پیشنهادی در نظر گرفته می‌شود.

گام ۳: محاسبه معیار دقت [2] به عنوان یکی از معیارهای ماتریس آشفتگی [3] با استفاده از رابطه

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{۱۵}$$

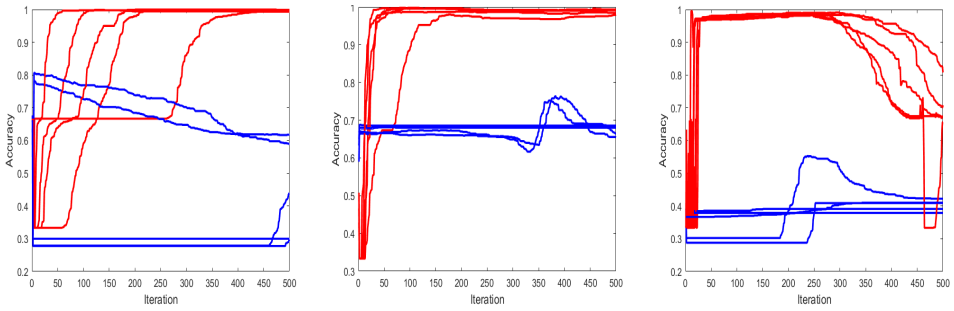
انجام می‌شود، که در آن TP ، TN ، FP و FN به ترتیب مثبت درست، منفی درست، مثبت کاذب و منفی کاذب هستند. بر اساس رابطه (۱۵) دقت مدل‌ها تا چهار رقم اعشار به دست آمد.

گام ۴: مقادیر پارامترها با $C = ۱, \mu = ۰/۵, p = ۲, \gamma = ۰/۷$ تنظیم شده است. دقت به دست آمده از هر

[1] Monte Carlo
[2] Accuracy
[3] Confusion matrix

دو مدل در ۵۰۰ تکرار و برای ۵ اجرا، در شکل ۱ برای سه تابع هسته خطی، چندجمله‌ای و گاوسی نشان داده شده است. رنگ قرمز، دقت مدل پیشنهادی و رنگ آبی، دقت مدل پایه را نشان می‌دهد. همان طور که دیده می‌شود تقریبا در تمام اجراها دقت مدل پیشنهادی برای هر سه تابع هسته در سطح بالاتری از مدل پایه قرار گرفته است.

گام ۵: برای ارزیابی سه مدل کرامر–سینگر، پایه و پیشنهادی، به ازای هر سه تابع هسته از آزمون فریدمن استفاده می‌شود. با توجه به نتایج این آزمون مشخص می‌شود که آیا تفاوتی بین مدل‌ها وجود دارد؟



(ج) (ب) (الف)

شکل ۱. دقت مدل‌های پیشنهادی و پایه در داده‌های شبیه‌سازی شده برای تابع هسته الف– خطی، ب– چندجمله‌ای، ج– گاوسی.

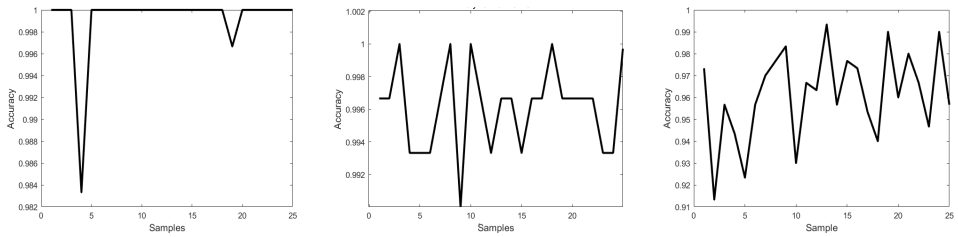
در جدول ۱ مقادیر مختلف γ در بازه $[۰/۱, ۰/۹]$ در نظر شده است و برای هر مقدار، ده تکرار برای هر دو مدل اجرا شده است و میانگین آن‌ها به دست آمده است. مقادیر سایر پارامترها با $C = ۱$، $\mu = ۰/۵$، $p = ۲$ تنظیم شده است. مشاهده می‌شود به ازای مقادیر کوچک γ، مدل پیشنهادی دقت کم تری دارد و پیش‌بینی برچسب صحیح کلاس دارای خطای بیش تری است. بنابراین مدل ارائه شده به ازای مقادیر $\gamma \leq ۰/۵$ برتری قابل توجهی ندارد، اما به ازای مقادیر $\gamma > ۰/۵$، بهتر از مدل پایه عمل کرده است. مقدار آماره آزمون فریدمن برای مقایسه میانگین دقت‌های به دست آمده به ازای سه تابع هسته در جدول ۱ برابر ۴۰/۹۶۳ و p–مقدار کمتر از ۰/۰۱ بیانگر تفاوت بین دقت سه مدل است. در شکل ۲ نمودار دقت برای هر سه تابع هسته در مدل پیشنهادی و ۲۵ نمونه و به ازای مقادیر پارامترهای $\gamma = [۰/۲, ۰/۳, ۰/۴, ۰/۵, ۰/۶, ۰/۷, ۰/۸, ۰/۹]$، $C = [۱, ۵, ۱۵, ۲۵, ۵۰]$ و $\mu = [۰/۰۵, ۰/۰۹, ۰/۱۵, ۰/۵, ۰/۷, ۰/۸]$ ارائه شده است. بطور کلی با توجه به نمودارها در تابع هسته خطی (آ) کوچکترین مقدار دقت تقریبا ۰/۹۱ و بزرگ ترین آن بیش از ۰/۹۹ است. در تابع هسته چندجمله‌ای (ب) کوچک ترین مقدار دقت برابر ۰/۹۹ و بزرگ ترین آن ۱ است. همچنین در تابع هسته گاوسی (ج) کم ترین مقدار دقت تقریبا ۰/۹۸ و بیشترین آن ۱ است که نشان می‌دهد مدل با هسته گاوسی بهترین عملکرد را دارد.

جدول ۱. جدول مقادیر معیارهای آماری دقت داده‌های شبیه‌سازی در هر دو مدل به ازای سه تابع هسته.

پارامتر	مدل	تابع خطی		تابع چندجمله‌ای		تابع گاوسی	
γ		میانگین	انحراف‌معیار	میانگین	انحراف‌معیار	میانگین	انحراف‌معیار
۰٫۱	کرامر–سینگر	۰٫۷۸۲۴	۰٫۰۱۳۲	۰٫۹۱۱۸	۰٫۰۲۳۱	۰٫۹۰۱۱	۰٫۰۲۱۶
۰٫۱	پایه	۰٫۷۹۲۱	۰٫۰۱۲۴	۰٫۸۹۸۷	۰٫۰۲۱۲	۰٫۹۱۲۱	۰٫۰۱۲۷
۰٫۱	پیشنهادی	۰٫۸۱۳۲	۰٫۰۲۴۱	۰٫۹۵۱۳	۰٫۰۳۰۱	۰٫۹۳۲۸	۰٫۰۳۱۸
۰٫۲	کرامر–سینگر	۰٫۸۲۲۱	۰٫۰۲۲۹	۰٫۹۰۱۳	۰٫۰۳۱۷	۰٫۹۰۵۴	۰٫۰۳۰۹
۰٫۲	پایه	۰٫۸۱۴۵	۰٫۰۲۲۱	۰٫۹۱۱۱	۰٫۰۲۰۱	۰٫۹۳۳۱	۰٫۰۲۰۱
۰٫۲	پیشنهادی	۰٫۸۳۰۴	۰٫۰۲۳۸	۰٫۹۴۵۴	۰٫۰۲۶۲	۰٫۹۱۳۴	۰٫۰۳۱۵
۰٫۳	کرامر–سینگر	۰٫۸۱۱۹	۰٫۰۲۶۴	۰٫۹۰۲۹	۰٫۰۲۸۱	۰٫۹۳۴۱	۰٫۰۲۷۵
۰٫۳	پایه	۰٫۸۰۲۲	۰٫۰۱۸۳	۰٫۹۲۲۵	۰٫۰۲۱۱	۰٫۹۵۷۱	۰٫۰۱۹۷
۰٫۳	پیشنهادی	۰٫۸۴۲۵	۰٫۰۲۳۷	۰٫۹۵۰۱	۰٫۰۲۲۵	۰٫۹۴۱۴	۰٫۰۲۱۴
۰٫۴	کرامر–سینگر	۰٫۷۹۲۵	۰٫۰۱۹۸	۰٫۸۶۴۳	۰٫۰۲۲۵	۰٫۹۲۱۱	۰٫۰۱۵۱
۰٫۴	پایه	۰٫۸۱۳۳	۰٫۰۲۶۸	۰٫۸۸۸۶	۰٫۰۲۳۱	۰٫۹۷۶۸	۰٫۰۳۱۷
۰٫۴	پیشنهادی	۰٫۸۵۱۱	۰٫۰۳۲۱	۰٫۹۶۳۶	۰٫۰۲۰۵	۰٫۹۵۷۴	۰٫۰۲۱۱
۰٫۵	کرامر–سینگر	۰٫۸۵۱۱	۰٫۰۲۸۶	۰٫۹۴۱۶	۰٫۰۱۹۷	۰٫۹۲۲۹	۰٫۰۳۰۲
۰٫۵	پایه	۰٫۸۳۲۵	۰٫۰۲۴۱	۰٫۹۳۲۱	۰٫۰۳۱۹	۰٫۹۴۲۴	۰٫۰۲۵۷
۰٫۵	پیشنهادی	۰٫۸۹۳۱	۰٫۰۳۹۷	۰٫۹۸۲۵	۰٫۰۲۱۲	۰٫۹۶۲۲	۰٫۰۱۲۹
۰٫۶	کرامر–سینگر	۰٫۸۲۹۰	۰٫۰۲۶۹	۰٫۹۳۲۱	۰٫۰۳۱۱	۰٫۹۰۱۷	۰٫۰۱۴۸
۰٫۶	پایه	۰٫۷۹۶۱	۰٫۰۲۵۴	۰٫۹۰۹۱	۰٫۰۳۷۱	۰٫۹۳۲۶	۰٫۰۲۳۵
۰٫۶	پیشنهادی	۰٫۹۱۴۸	۰٫۰۳۸۵	۰٫۹۹۱۶	۰٫۰۱۹۸	۰٫۹۷۹۸	۰٫۰۱۱۹
۰٫۷	کرامر–سینگر	۰٫۷۷۳۶	۰٫۰۲۲۱	۰٫۹۳۶۷	۰٫۰۲۸۹	۰٫۹۶۵۳	۰٫۰۲۷۱
۰٫۷	پایه	۰٫۷۹۴۴	۰٫۰۱۳۴	۰٫۹۱۱۱	۰٫۰۲۰۸	۰٫۹۷۱۴	۰٫۰۳۲۹
۰٫۷	پیشنهادی	۰٫۹۲۳۳	۰٫۰۱۸۷	۰٫۹۹۷۶	۰٫۰۱۰۳	۰٫۹۹۶۷	۰٫۰۱۵۳
۰٫۸	کرامر–سینگر	۰٫۸۲۱۹	۰٫۰۱۹۲	۰٫۹۱۲۰	۰٫۰۲۴۶	۰٫۹۷۲۵	۰٫۰۱۲۵
۰٫۸	پایه	۰٫۸۴۲۱	۰٫۰۲۲۷	۰٫۹۱۷۵	۰٫۰۲۴۵	۰٫۹۸۹۱	۰٫۰۲۷۸
۰٫۸	پیشنهادی	۰٫۹۲۷۴	۰٫۰۱۹۶	۰٫۹۹۸۹	۰٫۰۱۰۸	۰٫۹۹۹۷	۰٫۰۱۱۱
۰٫۹	کرامر–سینگر	۰٫۸۰۳۸	۰٫۰۱۲۴	۰٫۹۱۶۶	۰٫۰۲۷۸	۰٫۹۴۸۳	۰٫۰۱۲۶
۰٫۹	پایه	۰٫۸۲۹۷	۰٫۰۲۱۳	۰٫۹۲۲۵	۰٫۰۲۲۹	۰٫۹۶۴۱	۰٫۰۲۶۹
۰٫۹	پیشنهادی	۰٫۹۱۲۶	۰٫۰۱۷۴	۰٫۹۹۹۲	۰٫۰۱۱۷	۰٫۹۹۶۹	۰٫۰۱۰۹

۴ مثال کاربردی

در این بخش با استفاده از مجموعه داده های دانه‌ها و اکولی در دووآ و گرَف (۲۰۱۹) و همچنین مجموعه داده سرطان خون در http://csse.szu.edu.cn/staff/zhuzx/Datasets.html، دقت هر سه مدل مقایسه



(ج) (ب) (الف)

شکل ۲. نمودار دقت به ازای ۲۵ نمونه برای سه تابع هسته الف- خطی، ب- چندجمله‌ای، ج- گاوسی

شده است.

آ) جدول ۲ نتایج دقت برای مجموعه داده دانه‌ها به ازای طیف مختلف پارامترهای C، μ، γ و $p = ۲$ است. در توابع هسته خطی و چندجمله‌ای مدل پایه به علت مستقل بودن از پارامتر γ تغییر زیادی دیده نمی‌شود، اما این تغییرات بر روی مدل پیشنهادی به علت وابستگی به γ مشهود است. به ازای برخی مقادیر کوچک γ، دقت مدل پیشنهادی کم تر از خود در مقایسه با مقادیر بزرگ γ است که می‌تواند به علت فرایند نمونه‌گیری باشد. همان طور که مشاهده می‌شود در تابع هسته گاوسی به ازای مقادیر مختلف پارامترهای μ و γ، دقت مدل پیشنهادی بهتر از مدل پایه است. برای مقادیر بزرگ γ، مدل ارائه شده دقت بهتری نسبت به مدل پایه دارد. پس به طور کلی الگوی پیشنهادی، عملکرد بهتری نسبت به الگوی پایه و نیز نسبت به خودش در مقایسه با مقادیر کوچک γ به ازای هر سه تابع هسته دارد.

ب) جدول ۳ مقادیر دقت را برای مجموعه داده اکولی نشان می‌دهد که بر مبنای پارامترهای C، μ، γ و $p = ۲$ است. در تابع هسته خطی، مدل پیشنهادی نسبت به مدل پایه تقریبا عملکرد بهتری دارد. به ازای تابع هسته چندجمله‌ای، مدل پیشنهادی به طور کلی دقت بیش تری نسبت به مدل پایه دارد. برای تابع هسته گاوسی، همان طور که از جدول مشهود است، مقادیر دقت مدل پیشنهادی هم از مدل پایه و هم از دو تابع هسته خطی و چندجمله‌ای بهتر است. بنابراین می توان گفت تابع هسته گاوسی برای مدل پیشنهادی بهتر از مدل پایه است.

ج) جدول ۴ مقادیر دقت را برای مجموعه داده سرطان خون به ازای پارامترهای C، μ، γ و $p = ۲$ نشان می‌دهد. به ازای مقادیر کوچک γ و تابع هسته خطی، مدل پیشنهادی عملکرد ضعیف‌تری نسبت به مدل پایه دارد، اما برای مقادیر بزرگ تر پارامتر γ، مدل پیشنهادی دقت بهتری دارد. برای تابع هسته چندجمله‌ای، مدل پیشنهادی بطور کلی از مدل پایه عملکرد بهتری دارد. به ازای تابع هسته گاوسی نیز، مدل پیشنهادی عملکرد بسیار بهتری نسبت به مدل پایه دارد.

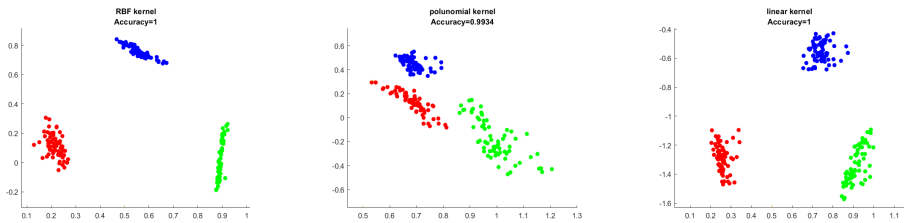
ح) شکل ۳ برای نحوه طبقه‌بندی مجموعه داده دانه‌ها به ترتیب برای توابع هسته خطی، چندجمله‌ای و گاوسی با پارامترهای $C = ۱$، $\mu = ۰/۱$، $\gamma = ۰/۷$ و $p = ۲$ رسم شده است. برای توابع هسته خطی و گاوسی دقت کامل به دست آمده است و به ازای هسته چندجمله‌ای، دقت برابر ۰/۹۹۳۴ است.

د) شکل ۴ متعلق به دسته‌بندی مجموعه داده اکولی است. مقادیر پارامترها $C = ۱$، $\mu = ۰/۵$، $\gamma = ۰/۵$ و $p = ۳$ هستند. برای تابع هسته خطی، دقت مدل برابر ۰/۹۹۸۳ به دست آمده است. مقادیر دقت برای هسته‌های چندجمله‌ای و گاوسی به ترتیب ۰/۹۸۴۶ و ۰/۹۷۷۵ به دست آمده است.

ر) شکل ۵ مربوط به دسته بندی مجموعه داده سرطان خون برای توابع هسته خطی، چندجمله‌ای و گاوسی با پارامترهای $\mu = ۰/۷$، $\gamma = ۰/۶۵$ و $p = ۳$ است. مقادیر دقت مدل برای توابع هسته خطی، چندجمله‌ای و گاوسی به ترتیب برابر ۰/۹۹۷۲، ۰/۹۹۸۹ و ۱ به دست آمده است.

جدول ۲. جدول مقادیر دقت مجموعه داده دانه‌ها برای هر دو مدل به ازای سه تابع هسته.

پارامترها			تابع خطی		تابع چندجمله‌ای		تابع گاوسی	
C	μ	γ	پایه	پیشنهادی	پایه	پیشنهادی	پایه	پیشنهادی
۵	۰/۰۵	۰/۳	۰/۹۵۴۲	۰/۹۷۱۶	۰/۹۸۷۳	۰/۹۷۵۳	۰/۹۷۱۰	۰/۹۶۴۴
۱۰	۰/۰۹	۰/۵	۰/۹۵۹۱	۰/۹۷۷۸	۰/۹۸۵۲	۰/۹۷۳۶	۰/۹۸۰۸	۰/۹۹۲۵
۲۵	۰/۱	۰/۶	۰/۹۶۱۷	۰/۹۸۶۸	۰/۹۷۸۳	۰/۹۹۱۹	۰/۹۸۶۷	۰/۹۹۴۷
۱	۰/۱۵	۰/۷	۰/۹۵۲۲	۰/۹۷۴۳	۰/۹۸۰۹	۰/۹۸۳۸	۰/۹۸۰۱	۰/۹۹۵۸
۱	۰/۲	۰/۸	۰/۹۷۲۹	۰/۹۸۹۱	۰/۹۸۹۴	۱	۰/۹۵۳۶	۰/۹۷۴۸
۱۵	۰/۷	۰/۹	۰/۹۸۱۴	۱	۰/۹۸۷۸	۰/۹۹۷۳	۰/۹۷۱۵	۰/۹۸۹۱
۵	۰/۰۱	۰/۸	۱	۰/۹۹۳۸	۰/۹۷۶۳	۰/۹۹۲۵	۰/۹۹۰۱	۱
۵۰	۰/۹	۰/۷	۰/۹۸۸۱	۰/۹۹۵۴	۰/۹۶۱۲	۱	۰/۹۸۲۳	۱
۳۰	۱/۷	۰/۴	۰/۹۹۵۱	۰/۹۸۶۱	۰/۹۷۰۴	۰/۹۸۰۴	۰/۹۶۵۷	۰/۹۷۱۳
۱۰	۲/۵	۰/۹	۰/۹۶۴۸	۰/۹۷۹۱	۰/۹۸۵۶	۱	۰/۹۳۷۱	۰/۹۵۱۵



(ج) (ب) (الف)

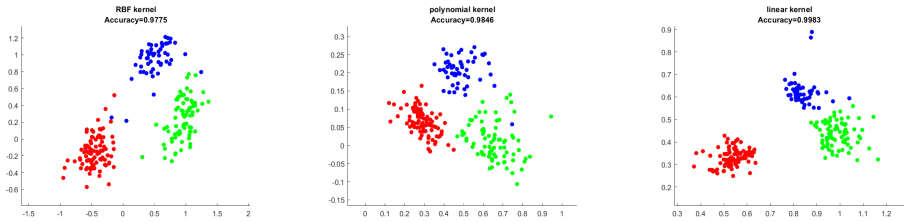
شکل ۳. دسته‌بندی کلاس‌ها برای مجموعه داده دانه‌ها. الف- خطی، ب- چندجمله‌ای، ج- گاوسی

جدول ۳. جدول مقادیر دقت مجموعه داده اکولی برای هر دو مدل به ازای سه تابع هسته.

پارامترها			تابع خطی		تابع چندجمله‌ای		تابع گاوسی	
C	μ	γ	پایه	پیشنهادی	پایه	پیشنهادی	پایه	پیشنهادی
۵	۰٫۰۵	۰٫۳	۰٫۹۲۱۱	۰٫۹۳۶۴	۰٫۹۷۷۵	۰٫۹۶۳۲	۰٫۹۸۲۷	۰٫۹۷۳۳
۱۰	۰٫۰۹	۰٫۵	۰٫۹۴۵۱	۰٫۹۵۳۶	۰٫۹۸۲۵	۰٫۹۷۱۲	۰٫۹۸۱۹	۰٫۹۷۱۰
۲۵	۰٫۱	۰٫۶	۰٫۹۲۴۱	۰٫۹۶۱۷	۰٫۹۶۱۸	۰٫۹۸۳۱	۰٫۹۶۳۷	۰٫۹۸۲۱
۱	۰٫۱۵	۰٫۷	۰٫۹۴۳۶	۰٫۹۸۱۹	۰٫۹۷۳۹	۰٫۹۹۵۸	۰٫۹۸۱۳	۱
۱	۰٫۲	۰٫۸	۰٫۹۸۳۳	۰٫۹۹۴۵	۰٫۹۷۳۴	۱	۰٫۹۶۴۹	۰٫۹۹۷۱
۱۵	۰٫۷	۰٫۹	۰٫۹۹۳۱	۱	۰٫۹۸۲۴	۰٫۹۸۷۸	۰٫۹۹۱۸	۱
۵	۰٫۰۱	۰٫۸	۱	۰٫۹۹۱۲	۰٫۹۸۲۶	۱	۰٫۹۹۵۴	۱
۵۰	۰٫۹	۰٫۷	۰٫۹۹۰۱	۰٫۹۹۵۶	۱	۰٫۹۸۹۹	۰٫۹۸۹۵	۰٫۹۹۶۸
۳۰	۱٫۷	۰٫۴	۰٫۹۸۱۱	۰٫۹۹۲۳	۰٫۹۸۷۷	۰٫۹۹۶۴	۰٫۹۸۲۱	۰٫۹۸۸۵
۱۰	۲/۵	۰٫۹	۰٫۹۸۱۹	۰٫۹۹۵۷	۰٫۹۸۱۶	۰٫۹۹۶۵	۰٫۹۴۳۳	۰٫۹۶۵۲

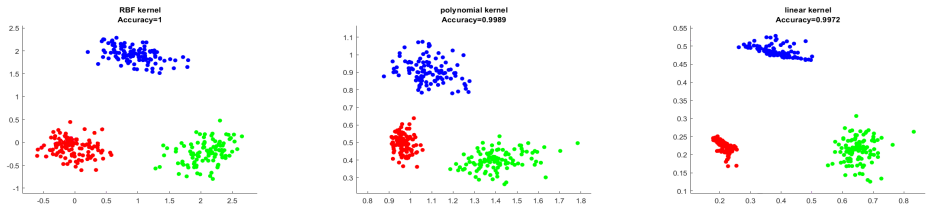
جدول ۴. جدول مقادیر دقت مجموعه داده سرطان خون برای هر دو مدل به ازای سه تابع هسته.

پارامترها			تابع خطی		تابع چندجمله‌ای		تابع گاوسی	
C	μ	γ	پایه	پیشنهادی	پایه	پیشنهادی	پایه	پیشنهادی
۵	۰٫۰۵	۰٫۳	۰٫۹۷۱۶	۰٫۹۵۹۹	۰٫۹۸۱۲	۰٫۹۷۲۳	۰٫۹۶۳۳	۰٫۹۶۹۵
۱۰	۰٫۰۹	۰٫۵	۰٫۹۷۴۲	۰٫۹۵۲۰	۰٫۹۷۱۷	۰٫۹۶۱۹	۰٫۹۸۷۸	۰٫۹۸۱۱
۲۵	۰٫۱	۰٫۶	۰٫۹۹۰۱	۰٫۹۷۹۸	۰٫۹۸۱۳	۰٫۹۸۱۳	۰٫۹۸۸۹	۰٫۹۹۱۸
۱	۰٫۱۵	۰٫۷	۰٫۹۶۱۸	۱	۰٫۹۸۵۹	۰٫۹۹۸۱	۰٫۹۸۱۷	۰٫۹۹۰۱
۱	۰٫۲	۰٫۸	۰٫۹۷۷۶	۰٫۹۸۳۲	۰٫۹۷۷۹	۱	۰٫۹۸۸۹	۰٫۹۹۱۳
۱۵	۰٫۷	۰٫۹	۰٫۹۹۶۶	۰٫۹۹۲۵	۰٫۹۸۹۲	۱	۰٫۹۹۴۱	۱
۵	۰٫۰۱	۰٫۸	۰٫۹۹۰۳	۰٫۹۹۵۱	۰٫۹۸۹۱	۰٫۹۹۹۰	۰٫۹۸۹۸	۰٫۹۹۲۸
۵۰	۰٫۹	۰٫۷	۰٫۹۷۶۳	۱	۰٫۹۸۹۹	۰٫۹۹۹۳	۰٫۹۷۲۶	۱
۳۰	۱٫۷	۰٫۴	۰٫۹۹۴۶	۰٫۹۸۱۱	۰٫۹۸۰۱	۰٫۹۹۳۵	۰٫۹۹۱۳	۰٫۹۹۶۱
۱۰	۲٫۵	۰٫۹	۰٫۹۸۸۸	۰٫۹۹۸۶	۰٫۹۹۱۴	۰٫۹۹۸۹	۰٫۹۹۷۲	۰٫۹۸۱۵



(ج) (ب) (الف)

شکل ۴. دسته‌بندی کلاس‌ها برای مجموعه داده اکولی. الف- خطی، ب- چندجمله‌ای، ج- گاوسی



(ج) (ب) (الف)

شکل ۵. دسته‌بندی کلاس‌ها برای مجموعه داده سرطان‌خون. الف- خطی، ب- چندجمله‌ای، ج- گاوسی

۵ بحث و نتیجه‌گیری

در دنیای واقعی اغلب با داده‌هایی سروکار داریم که به علت خطای ناشی از اندازه‌گیری یا نمونه‌گیری یا مدل‌سازی دقیق نیستند. در نتیجه نیاز به ارائه مدل‌هایی بر پایه الگوی غیرقطعی است. در این مقاله مدلی برای ورودی‌های غیرقطعی پیشنهاد شده است. در الگوی پیشنهادی با اضافه کردن مقداری اریبی در عبارت تنظیم کننده، تابع هدف اولیه به یک مسئله برنامه‌ریزی درجه دوم تبدیل گردید و اضافه کردن مقدار اریبی در محدودیت آن، سبب راحتی فرایند دسته‌بندی کلاس‌ها شده است. در این رویکرد از میانگین نمونه‌ای به عنوان برآوردی مناسب برای امیدریاضی استفاده شد. با توجه به نتایج بخش‌های ۳ و ۴، کارایی بهتر مدل با ورودی‌های غیرقطعی نسبت به مدل با ورودی‌های دقیق نشان داده شد. ضعف مدل پیشنهاد شده، سرعت آن در هنگام پردازش داده‌های با حجم بالا است. همچنین تاثیرپذیری میانگین از داده‌های پرت از دیگر ضعف‌های مدل پیشنهادی است که می‌تواند در مطالعات آینده مدنظر قرار گیرد.

۶ تقدیر و تشکر

نویسندگان این مقاله صمیمانه از داوران، سردبیر و ویراستار محترم به خاطر بازخورد ارزشمند و نظرات سازنده تشکر می کنند. پیشنهادات آن‌ها در ارتقای کیفیت نسخه نهایی این مقاله بسیار مؤثر بوده است. تحقیق نویسنده مسئول

از حمایت مالی دانشگاه فردوسی مشهد (۳/۶۰۷۵۴) برخوردار بوده است.

مراجع

موسوی، ن. و گلعلی‌زاده م. (۱۴۰۲)، رویکردی نوین در بکارگیری روش دسته ماشین بردار پشتیبان تصادفی در تحلیل داده‌های بیان ژن سرطان پروستات، مجله علوم آماری، ۱۷(۲)، ۴۵۹-۴۷۶.

Abaszade, M., and Effati, S. (2018), Stochastic Support Vector Regression with Probabilistic Constraints, *Applied Intelligence*, **48**, 243-256.

Abaszade, M., and Effati, S. (2019), A New Method for Classifying Random Variables Based on Support Vector Machine, *Journal of Classification*, **36**, 152-174.

Al-Ma'shumah, F., Razmkhah, M., and Effati, S. (2022), Expectation-Based and Quantile-Based Probabilistic Support Vector Machine Classification for Histogram-Valued Data, *International Journal on Electrical Engineering and Informatics*, **14**(1).

Barnett, N. S., Dragomir, S. S., and Agarwal, R. P. (2002), Some Inequalities for Probability, Expectation, and Variance of Random Variables Defined over a Finite Interval, *Computers & Mathematics with Applications*, **43**(10-11), 1319-1357. Pergamon.

Benítez-Peña, S., Blanquero, R., Carrizosa, E., and Ramírez-Cobo, P. (2024), Cost-sensitive Probabilistic Predictions for Support Vector Machines, *European Journal of Operational Research*, **314**(1), 268-279.

Bosch, P., López, J., Ramírez, H., and Robotham, H. (2013), Support Vector Machine under Uncertainty: An Application for Hydroacoustic Classification of Fish-schools in Chile, *Expert Systems with Applications*, **40**(10), 4029-4034.

Chen, W. J., Shao, Y. H., Li, C. N., and Deng, N. Y. (2016), MLTSVM: A Novel Twin Support Vector Machine to Multi-label Learning, *Pattern Recognition*, **52**, 61-74.

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

Crammer, K., and Singer, Y. (2001), On the Algorithmic Implementation of Multi-class Kernel-based Vector Machines, *Journal of Machine Learning Research*, **2**, 265-292.

Don, D. R., and Iacob, I. E. (2020), DCSVM: Fast Multi-class Classification using Support Vector Machines, *International Journal of Machine Learning and Cybernetics*, **11**(2), 433-447.

Dua, D., and Graff, C. (2019), UCI Machine Learning Repository. *University of California, Irvine, School of Information and Computer Sciences*.

Gonen, M., Tanugur, A. G., and Alpaydin, E. (2008), Multi-class Posterior Probability Support Vector Machines, *IEEE Transactions on Neural Networks*, **19**(1), 130-139.

Hastie, T., and Tibshirani, R. (1997), Classification by Pairwise Coupling, *The Annals of Statistics*, **26**(2), 451-471.

He, X., Wang, Z., Jin, C., Zheng, Y., and Xue, X. (2012), A Simplified Multi-class Support Vector Machine with Reduced Dual Optimization, *Pattern Recognition Letters*, **33**(1), 71-82.

Knerr, S., Personnaz, L., and Dreyfus, G. (1990), Single-layer Learning Revisited: A Stepwise Procedure for Building and Training a Neural Network,*Neurocomputing: Algorithms, Architectures and Applications*, 41-50.

Li, H. X., Yang, J. L., Zhang, G., and Fan, B. (2013), Probabilistic Support Vector Machines for Classification of Noise Affected Data, *Information Sciences*, **221**, 60-71.

Madjarov, G., Gjorgjevikj, D., and Chorbev, I. (2009), A Multi-class SVM Classifier Utilizing Binary Decision Tree, *Informatica*, **33**, 225-233.

Mooney, C. Z. (1997), Monte Carlo Simulation (No. 116), *Sage*.

Moosaei, H., and Hladík, M. (2022), Least Squares Approach to K-SVCR Multi-class Classification with Its Applications, *Annals of Mathematics and Artificial Intelligence*, **90**(7), 873-892.

Platt, J., Cristianini, N., and Shawe-Taylor, J. (1999), Large Margin DAGs for Multi-class Classification, *NIPS'99: Proceedings of the 13th International Conference on Neural Information Processing Systems*, **12**, 547-553..

Sahleh, A., Salahi, M., and Eskandari, S. (2023), Multi-class Nonparallel Support Vector Machine, *Progress in Artificial Intelligence*, **12**(4), 349-361.

Schiilkop, P. B., Burgest, C., and Vapnik, V. (1995), Extracting Support Data for a Given Task, *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*. AAAI Press, Menlo Park, CA, 252-257.

Vapnik, V. (1995), The Nature of Statistical Learning Theory, *NY: Springer-Verlag*.

Vapnik, V. (1999), An Overview of Statistical Learning Theory, *IEEE Transactions on Neural Networks*, (**10**)5, 988-999.

Weston, J., and Watkins, C. (1999), Support Vector Machines for Multi-class Pattern Recognition. *In Esann*, **99**, 219-224.

Yazdi, H. S., Effati, S., and Saberi, Z. (2007), The Probabilistic C,onstraints in the Support Vector Machine *Applied Mathematics and Computation*, **194**(2), 467-479.