



An Improvement on Twin Parametric-Margin Support Vector Machine

Ali Sohle¹, Maziar Salahi¹, Sadegh Eskandari²

¹ Department of Applied Mathematics, Faculty of Mathematical Sciences, Gilan University, Rasht, Iran

² Department of Computer Science, Faculty of Mathematical Sciences, Gilan University, Rasht, Iran

Abstract

Twin Parametric-Margin Support Vector Machine (TPMSVM) is one of the efficient tools in machine learning for binary classification especially for data with heteroscedastic noise. In this model, the goal is to find a pair of hyperplanes indirectly through a pair of nonparallel parametric-margin hyperplanes by solving two smaller sized quadratic programming problems for data classification such that each hyperplane is as far as possible from the samples of the other class in negative half-space made by a hyperplane. However, this does not guarantee that all of the samples will be far and be on the other side of the hyperplane. In this paper, we propose an enhancement of TPMSVM which keeps farther the samples of one class from the parametric margin hyperplane of other class by replacing a variable in the objective function. Numerical experiments on UCI datasets show that the enhanced model performs better compared to two similar models for both linear and nonlinear cases.

Keywords: Twin SVM, Heteroscedastic noise model, Classification, Parametric-margin model.

Paper Type:

Receive: 0

Revise:

Accept:

* Corresponding Author

Email Address:

DOI: <http://dx.doi.org/10.22105/dmor.2021.281369.1372>



بهبودی در ماشین بردار پشتیبان دوگانه حاشیه-پارامتری

علی سهله^۱، مازیار صلاحی^۱، صادق اسکندری^۲

۱ گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه گیلان، رشت، ایران

۲ گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه گیلان، رشت، ایران

ماشین بردار پشتیبان دوگانه حاشیه-پارامتری یکی از مدل‌های موثر در یادگیری ماشین برای جداسازی داده‌های به خصوص داده‌های با اختلال ناهمگن است. در این مدل، هدف پیدا کردن دو ابرصفحه به طور غیرمستقیم از طریق دو ابرصفحه حاشیه‌ای غیر موازی با حل دو مسئله بهینه سازی درجه دوم مقید با ابعاد کمتر برای جداسازی داده‌ها است، به طوری که ابرصفحه هر کلاس از داده‌های کلاس دیگر در طرف نیم فضای منفی حاصل از ابرصفحه تا جای ممکن دور باشد. اما این عبارت تضمین کننده دور بودن تمام نقاط از ابرصفحه و قرار داشتن آنها در نیم فضای منفی نیست. برای رفع این مشکل، در این مقاله یک نسخه بهبودیافته از آن ارائه می‌گردد که نقاط یک کلاس را از ابرصفحه حاشیه-پارامتری کلاس دیگر تا جای ممکن با جایگزین کردن متغیر جدید در تابع هدف دور می‌کند. آزمایش‌های عددی بر روی داده‌های کتابخانه‌ای UCI نشان دهنده عملکرد بهتر مدل پیشنهادی در مقایسه با دو مدل مشابه در حالت‌های خطی و غیرخطی از نظر دقت است.

واژه‌های کلیدی: ماشین بردار پشتیبان دوگانه، مدل واریانس ناهمگن، کلاس‌بندی، مدل حاشیه-پارامتری.

نوع مقاله:

پذیرش:

اصلاح: ---/---/---

دریافت: ---/---/---

۱- مقدمه

یادگیری ماشین^۱ یکی از زمینه‌های علمی است که باعث می‌شود رایانه‌ها بدون نیاز به یک برنامه یک موضوع خاص را یاد بگیرند و در مواجهه با رفتار و نمونه جدید تصمیم بگیرند. به عنوان زیر مجموعه‌ای از هوش مصنوعی، الگوریتم‌های یادگیری ماشین یک مدل ریاضی بر اساس داده‌های آموزش ایجاد می‌کنند تا پیش‌بینی یا تصمیم‌گیری بدون برنامه‌ریزی صریح، انجام شود. یادگیری ماشین در بسیاری زمینه‌ها از جمله مهندسی، کسب و کار (پرواست^۲ و همکاران ۲۰۱۳)، زبان‌شناسی و پزشکی (پیریانسی^۳ و همکاران ۲۰۲۰) کاربرد دارد. یکی از مدل‌های کارا در جداسازی داده‌های دودویی در زمینه یادگیری ماشین، ماشین بردار پشتیبان^۴ (SVM) (کرز^۵ ۱۹۹۵، وپینک^۶ ۱۹۹۶، وپینک ۱۹۹۸) است. با وجود عملکرد مناسب SVM، هزینه محاسباتی آن برای حل مسئله‌های برنامه‌ریزی درجه

³ Machine learning

³ Piryonesi

³ Provos

⁴ Support Vector Machine

⁵ Cortez

³ Vapink

دوم مقیاس- بزرگ بالا است. لذا به منظور بهبود این مشکل، جایادوا^۱ (۲۰۰۷) مدل SVM دوگانه (TWSVM) را ارائه داد که دو ابرصفحه جداکننده را از طریق حل دو مسئله درجه دوم با ابعاد کمتر تعیین می‌کند و این خود منجر به کاهش زمان یادگیری می‌شود. ابرصفحه‌های جداکننده به گونه‌ای تعیین می‌شوند که هر ابرصفحه به داده‌های یک کلاس نزدیک بوده و از کلاس دیگر یک واحد دور باشد. گسترش‌های زیادی از TWSVM ارائه شده‌اند، به عنوان مثال نصیری^۲ و همکاران (۲۰۲۰) مدل بهبود یافته نزدیکترین K همسایگی^۳ (KNN) منظم شده TWSVM را ارائه دادند که هزینه محاسباتی پیدا کردن KNN برای هر داده و اثر نقاط پرت در خروجی مدل را کاهش می‌دهد. جیمenez کاستانو^۴ و همکاران (۲۰۲۰) مدل بهبود یافته خودکار TWSVM را ارائه کردند که به منظور مقابله با داده‌های نامتعادل، مدل TWSVM را با نمایش هسته^۵ ادغام می‌کند.

یکی دیگر از مدل‌های کارا در جداسازی داده‌ها به خصوص داده‌هایی که دارای اختلال ناهمگن هستند توسعه ای از SVM با نام ماشین بردار پشتیبان حاشیه-پارامتری^۶ (par v-SVM) است که توسط هاو^۷ براساس مدل v-SVM (اسکلپف^۸ و همکاران ۲۰۰۰) ارائه شد. یکی از اشکالات این مدل همانند مدل کلاسیک SVM، زمان بالای حل مسئله‌های برنامه ریزی درجه دوم است. به منظور رفع این مشکل، پنگ^۹ (۲۰۱۰) بر پایه par v-SVM و مدل TWSVM، مدل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری^{۱۰} (TPMSVM) را معرفی کرد که به طور غیر مستقیم دو ابرصفحه جداکننده را از طریق حل دو مسئله درجه دوم با ابعاد کمتر تعیین می‌کند و مزیت par v-SVM را در رابطه با داده‌های با اختلال ناهمگن داراست. گوپتا^{۱۱} و همکاران ۲۰۱۷ مدل لاگرانژ فازی TPMSVM را ارائه دادند که اثر نقاط پرت را با تخصیص دادن وزن به هر داده کاهش می‌دهد. راستوگی^{۱۲} و همکاران ۲۰۱۸ مدل استوار TPMSVM را معرفی کردند که به طور خودکار حاشیه-پارامتری غیر حساس را تنظیم می‌کند. وانگ^{۱۳} و همکاران ۲۰۱۸، بهبودی از TPMSVM هموار ارائه دادند که با تبدیل کردن دو مسئله برنامه ریزی درجه دوم به دو مسئله مینیمم سازی نامقید، کارایی مدل را افزایش می‌دهد. راستوگی و همکاران ۲۰۱۸ نیز مدل TPMSVM مبتنی بر زاویه را ارائه دادند که دو ابرصفحه حاشیه-پارامتری را با ماکسیمم کردن زاویه بین دو ابرصفحه بدست می‌آورد.

یکی از عوامل اثر گذار در دقت کلاس بندی ابرصفحه‌های بدست آمده در مدل TPMSVM، تعداد داده‌هایی است که در طرف نیم فضای منفی حاصل از ابرصفحه کلاس مخالف قرار می‌گیرد. این عمل از طریق مینیمم کردن مجموع تصاویر نقاط کلاس مخالف بر ابرصفحه انجام می‌شود. اما این عبارت تضمین کننده دور بودن بیشتر نقاط از ابرصفحه و قرار داشتن آنها در نیم فضای منفی نیست. در این مقاله با تغییری در تابع هدف مدل TPMSVM، مدلی ارائه می‌شود که در مقایسه با مدل اصلی تعداد نقاط بیشتری از یک کلاس در طرف نیم فضای منفی حاصل از ابرصفحه کلاس دیگر قرار داشته باشند. ادامه مقاله به صورت زیر سازماندهی شده است. در بخش اول به بررسی مدل کلاسیک SVM، TWSVM، par v-SVM و Twin par v-SVM پرداخته می‌شود. در بخش دوم مدل بهبودیافته ارائه داده می‌شود و در بخش سوم نیز نتایج عددی و تحلیل آن براساس روش‌های آماری مرسوم در یادگیری ماشین ارائه می‌شود.

¹ Jayadeva

² Nasiri

³ K-nearest-neighbor

⁴ Jimenez-Castano

⁵ Kernel

⁶ Parametric-margin v-support vector machine

⁷ Hao

⁸ Schölkopf

⁹ Peng

¹⁰ Twin parametric margin Support Vector Machine

¹¹ Gupta

¹² Rastogi

¹³ Wang

فرض کنیم نمونه‌های داده شده به صورت بردارهای سطری $x_i \in R^n$ ($i = 1, \dots, d$) هستند که هر درایه آن نشان دهنده ویژگی نمونه است و هر داده دارای برچسب کلاس $y_i \in \{-1, +1\}$ است. مجموعه اندیس‌های کلاس داده‌های مثبت و منفی را به ترتیب به صورت I^+ و I^- نشان می‌دهیم. در ادامه مدل‌های TWSVM، par ν -SVM و Twin par ν -SVM به اختصار معرفی می‌شوند.

TWSVM-1-1

جایادوا ۲۰۰۷ مدلی ارائه کرد که دو مسئله درجه دوم با ابعاد کوچکتر را به جای یک مسئله درجه دوم در مقایسه با مدل کلاسیک SVM حل می‌کند. در این مدل دو ابرصفحه غیر موازی از طریق حل دو مسئله تعیین می‌شود. فرض کنید $A \in R^{m_1 \times n}$ ماتریس نقاط کلاس مثبت و $B \in R^{m_2 \times n}$ ماتریس نقاط کلاس منفی باشند و m_1 ، m_2 و نیز n به ترتیب تعداد نمونه‌های کلاس مثبت، منفی و تعداد ویژگی‌های هر نمونه هستند. دو ابرصفحه از طریق حل دو مسئله زیر بدست می‌آیند:

$$\begin{aligned} \min_{\omega_1, b_1, \xi_2} \quad & \frac{1}{2} \|A\omega_1 + e_1 b_1\|^2 + c_1 e_2^T \xi_2 \\ \text{s. t.} \quad & -(B\omega_1 + e_2 b_1) + \xi_1 \geq e_2, \\ & \xi_2 \geq 0, \end{aligned} \quad (1)$$

و

$$\begin{aligned} \min_{\omega_2, b_2, \xi_1} \quad & \frac{1}{2} \|B\omega_2 + e_2 b_2\|^2 + c_2 e_1^T \xi_1 \\ \text{s. t.} \quad & A\omega_1 + e_2 b_1 + \xi_1 \geq e_2, \\ & \xi_1 \geq 0, \end{aligned} \quad (2)$$

که در آن c_1 و c_2 ضرایب هزینه برای متغیرهای خطای ξ_1 و ξ_2 هستند و e_1 ، e_2 نیز بردارهای تمام یک هستند. در مدل (۱) ((۲)) عبارت اول در تابع هدف، ابرصفحه کلاس مثبت (منفی) را تا جای ممکن به نقاط کلاس مثبت (منفی) نزدیک می‌کند و قید اول، نقاط کلاس منفی (مثبت) را به فاصله یک واحد از ابرصفحه دور نگه می‌دارد. با توجه به اینکه حل مسائل دوگان از نظر محاسباتی در مقایسه با (۱) و (۲) سریع‌تر است، برای یافتن ابرصفحه‌ها از حل مسائل دوگان استفاده می‌شود. دوگان مسئله‌های (۱) و (۲) به صورت زیر هستند (جایادوا ۲۰۰۷):

$$\begin{aligned} \max_{\alpha} \quad & e_2^T \alpha - \frac{1}{2} \alpha^T G^T (H^T H)^{-1} G \alpha \\ \text{s. t.} \quad & 0 \leq \alpha \leq c_1 e_2, \end{aligned} \quad (3)$$

$$\begin{aligned} \max_{\beta} \quad & e_1^T \beta - \frac{1}{2} \beta^T P^T (Q^T Q)^{-1} P \beta \\ \text{s. t.} \quad & 0 \leq \beta \leq c_2 e_1, \end{aligned} \quad (4)$$





که در آن $\beta \in R^{m_2}$ و $\alpha \in R^{m_1}$ ضرایب لاگرانژ هستند. همچنین $G = [B \ e_2]$ ، $H = [A \ e_1]$ ، $P = [A \ e_1]$ و $Q = [B \ e_2]$ هستند.

۱-۲- par v-SVM

دو مدل SVM کلاسیک و SVM v بر پایه این فرض هستند که اختلال در داده‌ها به صورت یکنواخت هستند، اما در واقعیت این فرض همیشه برقرار نیست (اسکلیف، ۲۰۰۰). برای حل این مشکل، هاو ۲۰۱۰ مدلی حاصل از تغییر SVM v را ارائه کرد. در این مدل، حاشیه-پارامتری $g(x) = xz + d$ به جای حاشیه تابعی در مدل SVM v در نظر گرفته می‌شود. ابرصفحه جداکننده $f(x) = x\omega + b$ در مدل par v-SVM جداسازی داده‌ها را انجام می‌دهد اگر و تنها اگر

$$\text{الف) } x_i\omega + b \geq x_iz + d \quad \text{برای } i \in I^+$$

$$\text{ب) } x_i\omega + b \geq -x_iz - d \quad \text{برای } i \in I^- \text{ برای پیدا کردن } f(x) \text{ و } g(x) \text{ مسئله زیر حل می‌شود:}$$

$$\begin{aligned} \min_{\omega, b, z, d, \xi} & \frac{1}{2} \|\omega\|^2 + c \left(v \left(\frac{1}{2} \|z\|^2 + d \right) + \frac{1}{m_1} e^T \xi \right) \\ \text{s. t. } & y_i(x_i\omega + b) \geq x_iz + d - \xi_i \end{aligned} \quad (5)$$

که در آن ξ متغیر کمبود و $\mathbf{v} \geq \mathbf{0}$ ، \mathbf{c} ضرایب جریمه هستند. دوگان مسئله فوق به صورت زیر است:

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i x_i x_j^T y_j \alpha_j + \frac{1}{2cv} \sum_{i=1}^l \sum_{j=1}^l \alpha_i x_i x_j^T \alpha_j \\ \text{s. t. } & \sum_{i=1}^l y_i \alpha_i = 0, \\ & \sum_{i=1}^l \alpha_i \geq cv, \\ & 0 \leq \alpha_i \leq \frac{c}{m}, \quad i = 1, \dots, l. \end{aligned} \quad (6)$$

فرض کنیم $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ جواب بهینه مسئله دوگان فوق باشد. جواب بهینه مسئله اولیه فوق به صورت زیر بدست می‌آید:

$$\omega^* = \sum_{i=1}^l y_i \alpha_i^* x_i, \quad z^* = \frac{1}{cv} \sum_{i=1}^l x_i \alpha_i^*. \quad (7)$$

جواب‌های b و d نیز با استفاده از شرایط کروش کان تاکر^۱ (KKT) به صورت زیر بدست می‌آیند:

^۱ Karush–Kuhn–Tucker (KKT)



که در آن نقاط x_i و x_j طوری
انتخاب می‌شوند که $\alpha_i^*, \alpha_j^* \in$
 $(0, \frac{c_1}{l})$.

$$\begin{aligned} b &= -\frac{1}{2}(x_i \omega^* + x_j \omega^* - x_i z^* + x_j z^*), \\ d &= \frac{1}{2}(x_i \omega^* - x_j \omega^* - x_i z^* - x_j z^*), \end{aligned} \quad (8)$$

TPMSVM-۱-۳

مدل par ν -SVM مانند مدل کلاسیک SVM دارای سرعت یادگیری پایینی است. برای حل این مشکل، پنگ ۲۰۱۱ براساس par ν -SVM و ایده TWSVM، مدلی با نام Twin par ν -SVM ارائه کرد که دو ابرصفحه $f_1(x) = x\omega_1 + b_1 = 0$ و $f_2(x) = x\omega_2 + b_2 = 0$ را به طور غیرمستقیم تعیین می‌کند. هر کدام تعیین کننده یک ابرصفحه حاشیه-پارامتری به ترتیب برای کلاس مثبت و منفی است. جداسازی داده‌ها توسط مدل Twin par ν -SVM صورت می‌گیرد اگر و تنها اگر

$$i \in I^+ \text{ برای } x_i \omega_1 + b_1 \geq 0 \text{ (الف)}$$

$$i \in I^- \text{ برای } x_i \omega_2 + b_2 \leq 0 \text{ (ب)}$$

در واقع ابرصفحه‌های $f_1(x)$ و $f_2(x)$ معادل $f(x) \pm g(x)$ در par ν -SVM هستند. برای بدست آوردن ابرصفحه‌های حاشیه‌ای کلاس مثبت و منفی دو مسئله زیر در نظر گرفته می‌شود:

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1} & \frac{1}{2} \|\omega_1\|^2 + \frac{v_1}{m_2} e_2^T (B\omega_1 + e_2 b_1) + \frac{c_1}{m_1} e_1^T \xi_1 \\ \text{s. t.} & A\omega_1 + e_1 b_1 \geq -\xi_1, \\ & \xi_1 \geq 0, \end{aligned} \quad (9)$$

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2, r} & \frac{1}{2} \|\omega_2\|^2 - \frac{v_2}{m_1} e_1^T (A\omega_2 + e_1 b_2) + \frac{c_2}{m_2} e_2^T \xi_2 \\ \text{s. t.} & B\omega_2 + e_2 b_2 \leq \xi_2, \\ & \xi_2 \geq 0, \end{aligned} \quad (10)$$

بهبودی در ماشین بردار پشتیبان دوگانه حاشیه-پارامتری

که در آن $v_1, v_2 \geq 0$ و $c_1, c_2 \geq 0$ ضرایب پارامترهای منظم سازی و ξ_1, ξ_2 متغیرهای کمبود هستند. در مسئله (۹) ((۱۰)) عبارت اول در تابع هدف پیچیدگی ابرصفحه حاشیه‌ای را کنترل می‌کند. عبارت دوم در تابع هدف، مجموع مقدار تصویر نقاط کلاس منفی (مثبت) بر $f_1(x)$ ($f_2(x)$) را مینیمم می‌کند. نتیجه این رویکرد این است که نقاط کلاس منفی (مثبت) تا جای ممکن از ابرصفحه حاشیه‌ای کلاس مثبت (منفی) دور می‌شود. قید مسئله نیز شرط قرار داشتن نقاط کلاس مثبت (منفی) در طرف مثبت (منفی) نیم فضای حاصل از ابرصفحه حاشیه کلاس مثبت (منفی) را تضمین می‌کند. به طریق مشابه می‌توان نقش پارامترها و عبارات مدل (۱۰) را بیان کرد.

دوگان مسئله‌های (۹) و (۱۰) نیز به صورت زیر است:

$$\begin{aligned} \min_{\alpha, \beta} & -\frac{1}{2} \alpha^T A A^T \alpha + \frac{v_1}{m_2} e_2^T B A^T \alpha \\ \text{s. t.} & e_1^T \alpha = v_1, \\ & 0 \leq \alpha \leq \frac{c_1}{m_1} e_1, \end{aligned} \quad (11)$$

$$\min_{\alpha, \beta} -\frac{1}{2} \beta^T B B^T \beta^T + \frac{v_2}{m_1} e_1^T A B^T \beta$$

$$\text{s. t. } e_2^T \beta = v_2,$$

$$0 \leq \beta \leq \frac{c_2}{m_2} e_2.$$
(۱۲)

با فرض اینکه $\alpha^* = (\alpha_1^*, \dots, \alpha_{m_1}^*)^T$ و $\beta^* = (\beta_1^*, \dots, \beta_{m_2}^*)^T$ جواب‌های مسئله‌های (۱۱) و (۱۲) باشند، جواب‌های مسئله اولیه (۹) و (۱۰) به صورت زیر بدست می‌آید:

$$\omega_1^* = A^T \alpha^* - \frac{v_1}{m_2} B^T e_2, \quad b_1^* = -\frac{1}{|N_+|} \sum_{i \in N_+} x_i \omega_1^*, \quad N_+ = \{i | 0 < \alpha_i \leq \frac{c_1}{m_1} e_1\},$$
(۱۳)

$$\omega_2^* = B^T \beta^* - \frac{v_2}{m_1} A^T e_1, \quad b_2^* = -\frac{1}{|N_-|} \sum_{i \in N_-} x_i \omega_2^*, \quad N_- = \{i | 0 < \beta_i \leq \frac{c_2}{m_2} e_2\}.$$
(۱۴)

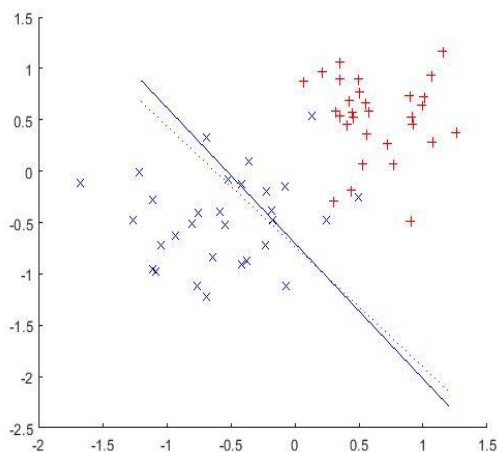
با استفاده از جواب‌های فوق، تابع تصمیم نیز به صورت زیر بدست می‌آید:

$$f(x) = \text{sign} \left((\hat{\omega}_1^* + \hat{\omega}_2^*)^T x + (\hat{b}_1^* + \hat{b}_2^*) \right),$$
(۱۶)

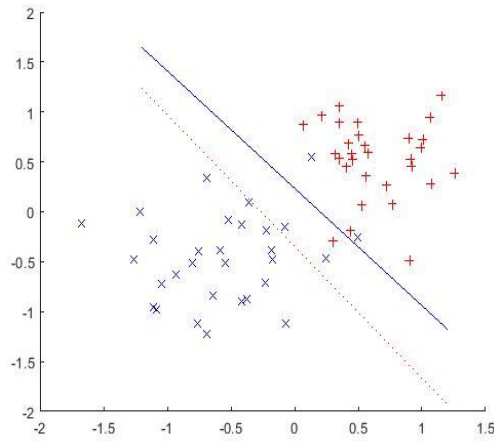
که در آن $\hat{\omega}_i^* = \omega_i^* / \|\omega_i^*\|$ و $\hat{b}_i^* = b_i^* / \|b_i^*\|$ (i=1,2) است.

۲- مدل بهبود یافته

در این بخش، مدل بهبود یافته TPMSVM (ITPMSVM) به منظور افزایش دقت کلاس بندی ارائه می‌شود. در مدل TPMSVM ابرصفحه یک کلاس از داده‌های کلاس دیگر تا جای ممکن از نقاط کلاس دیگر دور می‌شود که این کار از طریق مینیمم کردن مجموع تصویرهای نقاط کلاس دیگر بر ابرصفحه انجام می‌شود. به عنوان مثال، در شکل ۱ و ۲ که داده‌های کلاس I^+ و I^- با توزیع نرمال به ترتیب با میانگین ۰.۵- و ۰.۵ و واریانس ۰.۴ تولید شده‌اند، ملاحظه می‌شود در شکل ۱ تمام نقاط در طرف نیم‌فضای منفی حاصل از ابرصفحه قرار نگرفته‌اند. در مدل ITPMSVM، ابرصفحه‌ها همانند روش TPMSVM تعیین می‌شوند با این تفاوت که در این روش به جای افزایش حاشیه بین نقاط یک کلاس از ابرصفحه کلاس مخالف در قسمت منفی نیم‌فضای به وجود آمده حاصل از ابرصفحه، این هدف از طریق بهینه کردن متغیر به طور مستقیم در تابع هدف انجام می‌شود. بدین صورت با قرار گرفتن تعداد نقاط بیشتری در طرف نیم‌فضای منفی حاصل از ابرصفحه کلاس دیگر و تا جای ممکن دورتر از آن، کلاسبندی مناسب تری تعیین می‌شود (شکل ۲). مدل بهبود یافته در دو حالت خطی و غیرخطی در ادامه بحث می‌شود.



شکل ۱: نحوه کلاس بندی مدل TPMSVM



شکل ۲: نحوه کلاس بندی مدل ITPMSVM

۲-۱- حالت خطی ITPMSVM

مدل‌های بهبود یافته خطی روش پیشنهادی به صورت زیر هستند:

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1, r_1} & \frac{1}{2} \|\omega_1\|^2 - v_1 r_1 + \frac{c_1}{m_1} e_1^T \xi_1 \\ \text{s. t.} & A\omega_1 + e_1 b_1 \geq -\xi_1, \\ & B\omega_1 + e_2 b_1 \leq -r_1 e_2, \\ & \xi_1 \geq 0, \end{aligned} \quad (17)$$

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2, r_2} & \frac{1}{2} \|\omega_2\|^2 - v_2 r_2 + \frac{c_2}{m_2} e_2^T \xi_2 \\ \text{s. t.} & B\omega_2 + e_2 b_2 \leq \xi_2, \\ & A\omega_2 + e_1 b_2 \geq r_2 e_1, \\ & \xi_2 \geq 0, \end{aligned} \quad (18)$$

که در این دو مدل $v_1, v_2 \geq 0$ و $c_1, c_2 \geq 0$ ضرایب پارامترهای منظم سازی و ξ_1, ξ_2 متغیرهای کمبود هستند و متغیر r_1 (r_2) در تابع هدف نیز با بهینه شدن، نقاط کلاس منفی (مثبت) را تا جای ممکن از ابرصفحه حاشیه‌ای کلاس مثبت دور می‌کند. با توجه به سرعت حل بالای مدل‌های دوگان، همان بخش‌های قبل در ادامه مدل‌های دوگان مسائل فوق را ارائه می‌دهیم.

قضیه ۲-۱. دوگان مسئله (۱۷) به صورت زیر است

$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T A A^T \alpha^T - 2\alpha^T A B^T \beta^T + \beta^T B B^T \beta \\ \text{s. t.} & e_1^T \alpha = e_2^T \beta, \\ & e_2^T \beta = v_1 \beta, \\ & 0 \leq \alpha \leq \frac{c_1}{m_1} e_1, \\ & \beta \geq 0, \end{aligned} \quad (19)$$

اثبات. تابع لاگرانژ مسئله (۱۷) به صورت زیر است:

$$L(\omega_1, b_1, \xi_1, r, \alpha, \beta, \gamma) = \frac{1}{2} \|\omega_1\|^2 - v_1 r + \frac{c_1}{m_1} e_1^T \xi_1 - \alpha^T (A\omega_1 + e_1 b_1 + \xi_1) + \beta^T (B\omega_1 + e_2 b_1 + r e_2) - \gamma \xi_1, \quad (20)$$

که در آن $\alpha^T = (\alpha_1, \dots, \alpha_{m_1})^T$ و $\beta^T = (\beta_1, \dots, \beta_{m_2})^T$ بردار ضرایب لاگرانژ هستند. لذا دوگان مسئله (۱۷) به صورت زیر بدست می‌آید:

$$\max_{\alpha, \beta, \gamma} \inf_{\omega_1, b_1, \xi_1, r} L(\omega_1, b_1, \xi_1, r, \alpha, \beta, \gamma). \quad (21)$$

با استفاده از شرایط لازم و کافی بهینگی KKT برای مسئله فوق نتیجه می‌شود:

$$\frac{\partial L}{\partial \omega_1} = \omega_1 - A^T \alpha + B^T \beta = 0, \quad (22)$$

$$= 0, \quad \frac{\partial L}{\partial b_1} = -e_1^T \alpha + e_2^T \beta \quad (23)$$

$$= 0, \quad \frac{\partial L}{\partial \xi_1} = \frac{c_1}{m_1} e_1 - \alpha - \gamma \quad (24)$$

$$= 0, \quad \frac{\partial L}{\partial r} = -v_1 + e_2^T \beta \quad (24)$$

$$A\omega_1 + e_1 b_1 \geq -\xi_1, \quad (24)$$

$$B\omega_1 + e_2 b_1 \leq -re_2, \quad (25)$$

$$\xi_2 \geq 0, \quad (26)$$

$$\alpha \geq 0, \quad (27)$$

$$\beta \geq 0, \quad (28)$$

$$\gamma \geq 0. \quad (29)$$

از رابطه (24) و (29) نیز نتیجه می‌شود:

$$0 \leq \alpha \leq \frac{c_1}{m_1} e_1 \quad (30)$$

حال با جایگذاری روابط (22) تا (25) و رابطه (30) در (21)، (19) بدست می‌آید. □

قضیه ۲-۲. دوگان مسئله (۱۸) به صورت زیر است:

$$\begin{aligned} \max_{\alpha, \beta} \quad & \alpha^T B B^T \alpha^T - 2\alpha^T B A^T \beta^T + \beta^T A A^T \beta \\ \text{s.t.} \quad & e_1^T \alpha = e_2^T \beta, \\ & e_2^T \beta = v_1, \\ & 0 \leq \alpha \leq \frac{c_2}{m_2} e_2, \\ & \beta \geq 0, \end{aligned} \quad (31)$$

اثبات. مشابه اثبات قضیه ۱-۲ است، لذا از بیان جزییات صرف نظر می‌شود. □

۲-۲- حالت غیرخطی ITPMSVM



در دنیای واقعی معمولاً کلاسیک‌ترین نمونه‌های داده با استفاده از جداسازی خطی امکان پذیر نیست. برای حل این مشکل، یک نگاهت غیر خطی معرفی می‌شود که نقطه x را به فضایی با بعد بالاتر نگاهت می‌کند. تحت شرایط قضیه مرسر^۱ (مرسر ۱۹۹۰، وپینک ۱۹۹۶ و ۱۹۹۸) می‌توانیم از تابع هسته $K(x, x') = \varphi(x)\varphi(x')^T$ استفاده کنیم. یکی از توابع هسته متداول، هسته گوسی $K(x, x') = e^{-\|x-x'\|^2/\sigma^2}$ است. برای بسط روش پیشنهادی به حالت غیر خطی، به جای دو ابرصفحه خطی، دو ابرصفحه زیر در فضای ویژگی در نظر گرفته می‌شود:



۱۰۷

$$f_1(x) = \varphi(x)\omega_1 + b_1 = 0, \quad f_2(x) = \varphi(x)\omega_2 + b_2 = 0, \quad (32)$$

مشابه حالت خطی، دو ابرصفحه فوق با حل دو مسئله برنامه ریزی درجه دوم زیر بدست می‌آیند:

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1, r} & \frac{1}{2} \|\omega_1\|^2 - v_1 r + \frac{c_1}{m_1} e_1^T \xi_1 \\ \text{s. t.} & \quad \tilde{A}\omega_1 + e_1 b_1 \geq -\xi_1, \\ & \quad \tilde{B}\omega_1 + e_2 b_1 \leq -r e_2, \\ & \quad \xi_1 \geq 0, \end{aligned} \quad (33)$$

و

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2, r} & \frac{1}{2} \|\omega_2\|^2 - v_2 r + \frac{c_2}{m_2} e_2^T \xi_2 \\ \text{s. t.} & \quad \tilde{B}\omega_2 + e_2 b_2 \leq \xi_2, \\ & \quad \tilde{A}\omega_2 + e_1 b_2 \geq r e_1, \\ & \quad \xi_2 \geq 0, \end{aligned} \quad (34)$$

بهبودی در ماشین بردار پشتیبان دوگانه حاشیه-پارامتری

که در آن \tilde{A} و \tilde{B} ماتریس‌های حاصل از تحت نگاهت φ هستند. به منظور بهره‌گیری از سرعت بالای حل مدل‌های دوگان این مدل‌ها در قضیه بعدی ارائه می‌شوند.

$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T K(A, A)\alpha^T - 2\alpha^T K(A, B)\beta^T + \beta^T K(B, B)\beta \\ \text{s. t.} & \quad e_1^T \alpha = e_2^T \beta, \\ & \quad 0 \leq \alpha \leq \frac{c_1}{m_1} e_1, \\ & \quad \beta \geq 0, \end{aligned} \quad (35)$$

^۱Mercer

$$\begin{aligned} \max_{\alpha, \beta} \quad & \alpha^T K(B, B) \alpha^T - 2\alpha^T K(B, A) \beta^T + \beta^T K(A, A) \beta \\ \text{s. t.} \quad & e_2^T \alpha = e_1^T \beta, \\ & = v_1 \quad e_2^T \beta \\ & 0 \leq \alpha \leq \frac{c_1}{m_2} e_2, \\ & \beta \geq 0. \end{aligned} \quad (36)$$



قضیه ۲-۳. دوگان مسائل (۳۳) و (۳۴) به صورت زیر هستند:

اثبات. مشابه اثبات قضیه ۱-۲ است. □

۳- نتایج محاسباتی

در این بخش دقت کلاس بندی مدل پیشنهادی برای داده های مختلف که شامل داده های کتابخانه ای UCI برای جداسازی خطی و غیرخطی در مقایسه با مدل های TWSVM، TPMSVM تحلیل می شود. مشخصات داده های UCI استفاده شده برای مقایسه، در جدول ۱ ارائه و نتایج با استفاده از تست اصلاح شده فریدمن تحلیل شده است. پیاده سازی مدل ها در محیط برنامه نویسی متلب ۲۰۱۸ تحت سیستم عامل ویندوز ۷ و با پردازنده intel core i5 چهار گیگابایت حافظه انجام شده است.

جدول ۱: مشخصات داده های کتابخانه ای UCI

Data	Classes	Samples	Positive	Negative	Features
Sonar	2	208	97	111	60
Iris	3	435	267	168	16
Votes	2	270	120	150	13
Cancer	2	32	9	23	56
Housevotes	2	435	168	267	16
Shape	2	258	129	129	2
Liver disorder	2	345	145	200	6
Austrilian	2	690	307	383	14
Lung Cancer	2	32	9	23	56
Wdbc	2	569	212	357	30
Corner cancer	2	303	216	87	303
Ion sphere	2	351	225	126	34

۱-۳- تنظیم پارامترها

برای حالت غیرخطی از تابع هسته گوسی $K(x, x') = e^{-\|x-x'\|^2/\sigma^2}$ استفاده شده که پارامترهای C_1 و C_2 و نیز پارامتر σ برای مدل‌های TWSVM، TPMSVM و ITPMSVM از مجموعه $\{2^i \mid i = -10, \dots, +10\}$ انتخاب شده است. پارامترهای v_1 و v_2 برای TPMSVM و ITPMSVM نیز از بازه (0,1) انتخاب شده‌اند. به منظور کنترل برازش داده، آزمایش بر روی اعتبار سنجی متقابل k -دسته‌ای انجام شده است. دقت کلاس‌بندی به صورت زیر تعریف می‌شود:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (37)$$

در آن TP ، TN ، FP و FN به ترتیب تعداد نقاط مثبت درست، منفی درست، مثبت غلط، منفی غلط است.

۲-۳- مقایسه دقت جداسازی مدل‌ها

نتایج حل مدل‌ها در جدول ۲ و ۳ برای حالت خطی و غیرخطی برای داده‌های جدول ۱ ارائه شده است. در هر سطر که یک مدل عملکرد بهتری نسبت به بقیه مدل‌ها از لحاظ دقت داشته عدد پررنگ شده است. از جدول ۲ مشاهده می‌کنیم که در چهار مورد مدل پیشنهادی دقت بهتری دارد و تنها مدلی که از لحاظ عملکرد به مدل ارائه شده نزدیک است مدل TWSVM می‌باشد. مدل TPMSVM تنها در یک داده از مدل دیگر عملکرد بهتری دارد. اما در جدول ۳ عملکرد مدل پیشنهادی برای حالت غیرخطی تقریباً در همه موارد بهتر از سایر مدل‌ها است. در بخش بعد، به منظور اینکه نشان داده شود که مدل‌های مقایسه شده در حالت کلی از لحاظ عملکرد و دقت تمایز دارند از آزمون آماری رایج در مقایسه مدل‌های ماشین بردار پشتیبان استفاده می‌شود.

جدول ۲ دقت حاصل از اجرای مدل‌های TWSVM، TPMSVM و ITPMSVM بر روی داده‌های UCI در حالت خطی

Data	ITPMSVM	TPMSVM	TWSVM
sonar	78.357	77.779	74.710
iris	100	100	100
votes	96.064	94.931	96.564
cancer	95.851	96.165	95.557
housevotes	95.169	94.946	95.645
Shape	99.231	98.046	99.615
Liver Disorder	68.975	53.933	66.672
Austerilia	85.58	62.732	85.758
Lung cancer	78.333	76.667	27.500
wdbc	94.869	88.992	84.217



جدول ۳ دقت حاصل از اجرای مدل‌های TPMSVM، TWSVM و ITPMSVM بر روی داده‌های UCI در حالت غیرخطی

Data	ITPMSVM	TPMSVM	TWSVM
sonar	76.509	46.617	76.176
iris	100	100	100
votes	96.105	61.379	47.541
cancer	96.704	65.522	45.833
housevotes	95.782	38.319	42.278
Shape	97.662	50.000	96.908
Liver Disorder	68.975	53.933	66.672
Austerilia	66.348	51.979	44.492
Lung cancer	69.167	62.500	45
wdbc	90.498	37.000	37.258



۳-۳- آزمایش‌های آماری

برای تحلیل نتایج عددی، از تست فریدمن^۱ استفاده می‌شود که این تست چندین چندین الگوریتم را بدون اینکه فرضی بر توزیع داده‌ها باشد مقایسه می‌کند. در این آزمون، فرض اولیه مبتنی بر یکسان بودن عملکرد الگوریتم‌هاست. رد شدن فرض اولیه به این معنی است که در بین الگوریتم‌ها حداقل دو الگوریتم با هم اختلاف معنا داری دارند. این آزمون هر الگوریتم را برای هر داده رتبه بندی می‌کند. الگوریتمی که بهترین عملکرد را داشته باشد رتبه یک و الگوریتم بعدی از لحاظ عملکرد رتبه دو و برای بقیه بر همین اساس رتبه تخصیص می‌دهد. فرض کنیم r_{ij} رتبه الگوریتم i ام برای داده j ام باشد. این تست میانگین رتبه برای هر کلاس بندی، یعنی $\bar{r}_j = \sum_{i=1}^k r_{ij}$ را با هم مقایسه می‌کند که N تعداد داده‌ها و فرمول آن به صورت زیر است:

$$X_F^2 = \frac{12N}{k(k+1)} \sum_{j=1}^k \left(\bar{r}_j - \frac{k+1}{2} \right)^2, \quad (38)$$

که در آن k تعداد الگوریتم است. ایمان^۲ و داوینپرت^۳ ۱۹۹۸ نشان دادند که آماره X_F^2 محافظه کارانه است و آماره بهتری که از رابطه (۳۸) بهره می‌گیرد ارائه کردند که به صورت زیر تعریف می‌شود:

¹ Friedman
² Iman
³ Davenport

$$F_f = \frac{X_F^2(N-1)}{N(k-1) - X_F^2}, \quad (39)$$

که در آن F_f دارای توزیع F با درجه آزادی $(k-1), (k-1)(N-1)$ است. جدول ۴ و ۵ رتبه الگوریتم‌ها برای داده‌های UCI را نشان می‌دهد. با توجه به جدول ۴، $X_F^2 = 17.084$ و $F_f = 27.172$ و با توجه به جدول ۵، $X_F^2 = 3.1174$ و $F_f = 1.6421$ است که F_f دارای توزیع F با درجه آزادی $(2,22)$ است. مقدار بحرانی F برای $\alpha = 0.01$ و $\alpha = 0.05$ به ترتیب $۳/۴۴۳$ و $۵/۷۱۹$ است. چون مقدار F_f بیشتر از مقدار بحرانی برای حالت غیرخطی است پس عملکرد سه الگوریتم مقاسیه شده در حالت غیر خطی متفاوت است و از آنجایی که رتبه الگوریتم پیشنهادی کمتر از بقیه است پس عملکرد بهتری دارد.

جدول ۴: رتبه دقت داده‌های UCI برای مدل‌های TWSVM، TPMSVM و ITPMSVM برای حالت خطی

Data	ITPMSVM	TPMSVM	TWSVM
sonar	1	3	2
iris	2	2	2
votes	1	2	3
cancer	1	2	3
housevotes	1	2	3
Shape	1	3	2
Liver Disorder	1	3	2
Austerilia	1	2	3
Lung cancer	1	2	3
wdbc	1	3	2

جدول ۵: رتبه دقت داده‌های UCI برای مدل‌های TWSVM، TPMSVM و ITPMSVM برای حالت غیرخطی

Data	ITPMSVM	TPMSVM	TWSVM
sonar	1	2	3
iris	2	2	2
votes	2	3	1
cancer	2	1	3
housevotes	2	3	1
Shape	2	3	1



Liver Disorder	1	3	2
Austerilia	2	3	1
Lung cancer	1	2	3
wdbc	1	2	3

۴- نتیجه‌گیری

با توجه به این که از عوامل تعیین کننده میزان دقت کلاس بندی هر ابر صفحه در مدل TPMSVM تعداد داده‌های قرار گرفته یک کلاس در طرف نیم فضای منفی حاصل از ابر صفحه کلاس دیگر است و اینکه این عمل از طریق مینیمم کردن عبارتی انجام می‌شود که مجموع تصاویر نقاط کلاس مخالف بر ابر صفحه کلاس دیگر است، در این مقاله با جایگزینی این عبارت توسط متغیر جدید بهبودی از این مدل ارائه گردید. مزیت مدل بهبود یافته نسبت به مدل اولیه در این است که ابر صفحه بدست آمده به گونه ای تعیین می‌شود که تعداد نقاط بیشتری از یک کلاس در طرف منفی نیم فضای حاصل از ابر صفحه کلاس دیگر قرار بگیرند و ابر صفحه‌های مناسب‌تری برای جداسازی داده‌ها بدست آید. آزمایش‌هایی انجام شده بر روی داده‌های UCI و مقایسه مدل ارائه شده با TPMSVM و TWSVM، نشان دهنده این است که در حالت خطی تنها مدلی که از لحاظ عملکرد به مدل ارائه شده نزدیک است مدل TWSVM می‌باشد. اما عملکرد مدل پیشنهادی برای حالت غیر خطی تقریباً در همه موارد بهتر از سایر مدل‌ها است. با استفاده از اصلاح شده تست فریدمن مشاهده شد مدل بهبود یافته از لحاظ دقت در حالت غیر خطی در حالت کلی عملکرد بهتری نسبت به TPMSVM و همچنین TWSVM دارد و رتبه مدل پیشنهادی نسبت به مدل‌های مقایسه شده در دو حالت خطی و غیر خطی کمتر است. در تحقیقات آتی می‌توان مدل پیشنهادی را در حضور قیدهای احتمالی نیز مطالعه کرد.

منابع

1. S. M. Pirayonesi, T. E. El-Diraby, Data analytics in asset management: Cost-effective prediction of the pavement condition index, *Journal of Infrastructure Systems* 26 (1) (2020) 04019036.
2. F. Provost, T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, Inc (2013)
3. V.N. Vapnik, B. Boser, L. Guyon, A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ACM Press, Pittsburgh, 1992 144-152.
4. C. Cortes, V.N. Vapnik, Support vector networks, *Machine Learning* 20 (3) (1995) 273-297.
5. V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1996.
6. V.N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.
7. B. Schölkopf, I. Guyon, J. Weston, *Statistical learning and kernel methods in Bioinformatics*, Technical Report, 2000.
8. P. Hao, New support vector algorithms with parametric insensitive margin model, *Neural Networks* 23 (1) (2010) 60-73.
9. Jayadeva, R. Khemchandani, S. Chandra, Twin support vector machines for pattern classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (5) (2007) 905-910.
10. X. Peng, TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition, *Pattern Recognition* 44 (10-11) (2011) 2678-2692.
11. D. Gupta, P. Borah, M. Prasad, A fuzzy based Lagrangian twin parametric-margin support vector machine (FLTPMSVM), In *2017 IEEE symposium series on computational intelligence (SSCI)* (2017) 1-7.
12. Z. Wang, Y.H. Shao, L. Bai, L.C. Li, L.M. Liu, N.Y. Deng, Insensitive stochastic gradient twin support vector machines for large scale problems, *Information Sciences* 462 (1) (2018) 114-31.
13. R. Rastogi, P. Saigal, S. Chandra, Angle-based twin parametric-margin support vector machine for pattern classification, *Knowledge-Based Systems* 139 (2018) 64-77.



14. R. Rastogi, S. Sharma, S. Chandra, Robust parametric twin support vector machine for vector machine pattern classification, *Neural Processing Letters* 47 (1) (2018) 293-323.
15. J.A. Nasiri, A.M. Mir, An enhanced KNN-based twin support vector machine with stable learning rules, *Neural Computing and Applications* 16 (2020) 12949-69.
16. C. Jimenez-Castano, A. Alvarez-Meza, A. Orozco-Gutierrez, Enhanced automatic twin support vector machine for imbalanced data classification, *Pattern Recognition* 107 (2020) 107442.
17. B. Schölkopf, A. J. Smola, R Williamson & P. L. Bartlett (2000). New support vector algorithms. *Neural Computation*, 12(5), 1207-1245.
18. O.L. Mangasarian, E.W. Wild, Multisurface proximal support vector classification via generalized eigenvalues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (1) (2005) 69-74.
19. J. Demsar, Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research* 7 (2006) 1-30.
20. R. Rastogi, S. Sharma, S. Chandra. Robust parametric twin support vector machine for pattern classification. *Neural Processing Letters*. 47(1)(2018) 293-323.
21. M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, *The Annals of Mathematical Statistics* 11(1) (1940) 86-92
22. R.L. Iman, J.M. Davenport, Approximations of the critical region of the fbietkan statistic, *Communications in Statistics-Theory and Methods* 9 (6) (1980) 571-595.
23. J. Mercer, Functions of positive and negative type and the connection with the theory of integral equations, *Philosophical Transactions of the Royal Society of London, Series A, containing papers of a mathematical or physical character* 209 (441-458) (1990) 415-446.
24. N. Deng, Y. Tian, C. Zhang, *Support Vector Machines Optimization Based Theory, Algorithms, and Extensions*, Champman & Hall 2012.