


Paper Type: Original Article



An Improvement on Twin Parametric-Margin Support Vector Machine

Ali Sahleh¹, Maziar Salahi^{1*} , Sadegh Eskandari²

¹ Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran; alisahleh@gmail.com; salahim@guilan.ac.ir.

² Department of Computer Science, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran; eskandari@guilan.ac.ir.

Citation:



Sahleh, A., Salahi, M., & Eskandari, S. (2022). An improvement on twin parametric-margin support vector machine. *Journal of decisions and operations research*, 7(4), 503-514.

Received: 15/04/2021

Reviewed: 28/05/2021

Revised: 30/09/2021

Accepted: 03/10/2021

Abstract

Purpose: The aim of this paper is to present an enhanced variant of Twin Parametric-Margin Support Vector Machine (TPMSVM) that improves classification performance.

Methodology: By replacing a variable in the objective function, we keep the samples of one class farther from the parametric margin hyperplane of the other class.

Findings: The enhanced model is convex for both linear and nonlinear cases. Also, numerical experiments on UCI datasets show that the enhanced model performs better compared to two similar models for both linear and nonlinear cases.

Originality/Value: The previous studies of TPMSVM that increased the accuracy through approaches such as assigning weights to data sample, converting it into an unconstrained model and adding a new term in the objective function, did not guarantee that all samples will be far and on the negative side of the margin hyperplane. However, this study provides an approach to overcome this disadvantage of TPMSVM.

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

Corresponding Author: salahim@guilan.ac.ir

 <http://dorl.net/dor/20.1001.1.25385097.1401.7.4.1.2>



Licensee. **Journal of Decisions and Operations Research**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).



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.

چکیده

هدف: هدف در این مقاله ارائه یک نوع بهبودیافته از مدل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری به منظور بهبود عملکرد کلاس بندی است.

روش‌شناسی پژوهش: با جایگزین کردن متغیر جدید در تابع هدف، نقاط یک کلاس از ابرصفحه حاشیه-پارامتری کلاس دیگر تا جای ممکن دور می‌شود.

یافته‌ها: مدل بهبودیافته در هر دو حالت خطی و غیرخطی محدب است. همچنین آزمایش‌های عددی بر روی داده‌های کتابخانه‌ای UCI نشان دهنده عملکرد بهتر مدل پیشنهادی در مقایسه با دو مدل مشابه در مقایسه با دو مدل مشابه در حالت‌های خطی و غیرخطی از نظر دقت است.

اصالت/ارزش افزوده علمی: مطالعات قبلی انجام شده در مدل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری که از طریق روش‌های مانند وزن‌دار کردن داده‌ها، تبدیل کردن آن به مدل نامقید و یا اضافه کردن عبارت جدید در تابع هدف دقت مسئله را افزایش می‌دهند، تضمین‌کننده دور بودن تمام نقاط از ابرصفحه و قرار داشتن آن‌ها در نیم فضای منفی نیستند؛ اما این بررسی یک رویکرد جدید برای رفع این مشکل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری ارائه می‌دهد.

کلیدواژه‌ها: ماشین بردار پشتیبان دوگانه، مدل واریانس ناهمگن، کلاس بندی، مدل حاشیه-پارامتری.

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. به‌عنوان زیرمجموعه‌ای 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 (پراواست و فاوست^۲، ۲۰۱۳)، زبان‌شناسی و پزشکی (پیریانسی و ال-دیرابی^۳، ۲۰۲۰) کاربرد دارد. یکی از مدل‌های کارا در جداسازی داده‌های دودویی در زمینه یادگیری ماشین، ماشین بردار پشتیبان^۴ (SVM) (کرتز و وپنک^۵، ۱۹۹۵؛ وپنک^۶، ۱۹۹۸) است. با وجود عملکرد مناسب SVM، هزینه محاسباتی آن برای حل مسئله‌های برنامه‌ریزی درجه دوم مقیاس-بزرگ بالا است. لذا به منظور

¹ Machine learning² Provost and Fawcett³ Pirayonesi and El-Diraby⁴ Support Vector Machine⁵ Cortes and Vapnik⁶ Vapnik



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، به‌عنوان مثال نصیری و میر^۲ (۲۰۲۰) مدل بهبودیافته 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. جیمز کاستانو و همکاران^۴ (۲۰۲۰) مدل بهبودیافته 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 به‌خصوص داده‌هایی که دارای اختلال ناهمگن هستند توسعه‌ای از SVM با نام ماشین بردار پشتیبان حاشیه-پارامتری^۶ (par v-SVM) است که توسط هاو^۷ (۲۰۱۰) بر اساس مدل v-SVM (اسکلپف و همکاران^۸، ۲۰۰۰) ارائه شد. یکی از اشکالات این مدل همانند مدل کلاسیک SVM، زمان بالای حل مسئله‌های برنامه‌ریزی درجه دوم است. به منظور رفع این مشکل، پنگ^۹ (۲۰۱۱) بر پایه par v-SVM و مدل TWSVM، مدل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری^{۱۰} (TPMSVM) را معرفی کرد که به طور غیرمستقیم 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 و همکاران^{۱۱} (۲۰۱۷) مدل لاگرانژ فازی TPMSVM را ارائه دادند که اثر نقاط پرت را با تخصیص دادن وزن به هر داده کاهش می‌دهد. راستوگی و همکاران^{۱۲} (۲۰۱۸) مدل استوار TPMSVM را معرفی کردند که به طور خودکار حاشیه-پارامتری غیرحساس را تنظیم می‌کند. وانگ و همکاران^{۱۳} (۲۰۱۸)، بهبودی از TPMSVM هموار را ارائه دادند که با تبدیل کردن دو مسئله برنامه‌ریزی 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 به دست می‌آورد.

یکی از عوامل اثرگذار در دقت کلاس بندی ابرصفحه‌های به دست 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 تضمین‌کننده 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 بر اساس روش‌های آماری مرسوم در یادگیری ماشین ارائه می‌شود.

۲- پیش‌زمینه

فرض کنیم نمونه‌های داده شده به صورت بردارهای سطری $x_i \in R^n$ ($i = 1, \dots, l$) 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 مثبت و منفی را به ترتیب به صورت I^+ و I^- نشان می‌دهیم. در ادامه مدل‌های TWSVM، par v-SVM و Twin par v-SVM به اختصار معرفی می‌شوند.

۲-۱- TWSVM

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 $A \in R^{m_1 \times n}$ 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 به دست می‌آیند:

¹ Jayadeva

² Nasiri and Mir

³ K-nearest-neighbor

⁴ Jimenez-Castano et al.

⁵ Kernel

⁶ Parametric-margin v-support vector machine

⁷ Hao

⁸ Schölkopf et al.

⁹ Peng

¹⁰ Twin parametric margin support vector machine

¹¹ Gupta et al.

¹² Rastogi et al.

¹³ Wang et al.



$$\min_{\omega_1, b_1, \xi_2} \frac{1}{2} \| A\omega_1 + e_1 b_1 \|^2 + c_1 e_2^T \xi_2$$

$$s.t. \quad -(B\omega_1 + e_2 b_1) + \xi_2 \geq e_2,$$

$$\xi_2 \geq 0.$$
(۱)

و

$$\min_{\omega_2, b_2, \xi_1} \frac{1}{2} \| B\omega_2 + e_2 b_2 \|^2 + c_2 e_1^T \xi_1$$

$$s.t. \quad A\omega_2 + e_1 b_1 + \xi_1 \geq e_1,$$

$$\xi_1 \geq 0,$$
(۲)

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 مدل‌های (۱) و (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 مدل‌های (۱) و (۲) سریع‌تر است، برای یافتن ابرصفحه‌ها از حل مسائل دوگان استفاده می‌شود. دوگان مدل‌های (۱) و (۲) به صورت زیر هستند (جایادوا، ۲۰۰۷):

$$\max_{\alpha} e_2^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha$$

$$s.t. \quad 0 \leq \alpha \leq c_1 e_2.$$
(۳)

و

$$\max_{\beta} e_1^T \beta - \frac{1}{2} \beta^T P (Q^T Q)^{-1} P^T \beta$$

$$s.t. \quad 0 \leq \beta \leq c_2 e_1,$$
(۴)

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

۲-۲- par v-SVM

دو مدل SVM کلاسیک و v-SVM بر پایه این فرض هستند که اختلال در داده‌ها به صورت یکنواخت هستند، اما در واقعیت این فرض همیشه برقرار نیست (اسکلیف^۱، ۲۰۰۰). برای حل این مشکل، هاو (۲۰۱۰) مدلی حاصل از تغییر v-SVM را ارائه کرد. در این مدل، حاشیه-پارامتری $g(x) = xz + d$ به جای حاشیه تابعی در مدل v-SVM 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:

$$1. \quad i \in I^+ \text{ برای } x_i \omega + b \geq x_i z + d$$

$$2. \quad i \in I^- \text{ برای } x_i \omega + b \leq -x_i z + d$$

برای پیدا کردن $f(x)$ و $g(x)$ مسئله زیر حل می‌شود:

$$\min_{\omega, b, z, d, \xi} \frac{1}{2} \|\omega\|^2 + c \left(v \left(\frac{1}{2} \|z\|^2 + d \right) + \frac{1}{m_1} e^T \xi \right)$$

$$s.t. \quad y_i (x_i \omega + b) \geq x_i z + d \xi_i, \quad i = 1, \dots, l$$

$$\xi \geq 0,$$
(۵)

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

¹ Schölkopf



$$\begin{aligned} \max_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{i=1}^l \alpha_i y_i x_i x_i^T y_i \alpha_i + \frac{1}{2cv} \sum_{i=1}^l \sum_{i=1}^l \alpha_i x_i x_i^T \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0, \\ & \sum_{i=1}^l \alpha_i \geq cv, \\ & 0 \leq \alpha_i \leq \frac{C}{m}, \quad i=1, \dots, l. \end{aligned} \tag{۶}$$

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

$$\omega^* = \sum_{i=1}^l y_i \alpha_i^* x_i, z^* = \frac{1}{cv} \sum_{i=1}^l x_i \alpha_i^*. \tag{۷}$$

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

$$b = -\frac{1}{2} (x_i \omega^* + x_j \omega^* - x_i z^* + x_j z^*), \tag{۸}$$

$$d = \frac{1}{2} (x_i \omega^* - x_j \omega^* - x_i z^* - x_j z^*),$$

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

TPMSVM-۲-۳

مدل $par\ v-SVM$ مانند مدل کلاسیک SVM دارای سرعت یادگیری پایینی است. برای حل این مشکل، پنگ (۲۰۱۱) بر اساس $v-par\ SVM$ ایده $TWSVM$ ، مدلی با نام $Twin\ par\ v-SVM$ 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 $Twin\ par\ v-SVM$ صورت می‌گیرد اگر و تنها اگر:

$$1. \quad i \in I^+ \text{ برای } x_i \omega_1 + b_1 \geq 0$$

$$2. \quad i \in I^- \text{ برای } x_i \omega_2 + b_2 \leq 0$$

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

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1} \quad & \frac{1}{2} \|\omega_1\|^2 + \frac{v_1}{m_2} e_2^T (B\omega_1 + e_2 b_1) + \frac{c_1}{m_1} e_1^T \xi_1 \\ \text{s.t.} \quad & A\omega_1 + e_1 b_1 \geq \xi_1, \\ & \xi_1 \geq 0. \end{aligned} \tag{۹}$$

و

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2} \quad & \frac{1}{2} \|\omega_2\|^2 + \frac{v_2}{m_1} e_1^T (A\omega_2 + e_1 b_2) + \frac{c_2}{m_2} e_2^T \xi_2 \\ \text{s.t.} \quad & B\omega_2 + e_2 b_2 \leq \xi_2, \\ & \xi_2 \geq 0, \end{aligned} \tag{۱۰}$$

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 مسئله‌های (۹) و (۱۰) عبارت اول

¹ Karush-Kuhn-Tucker (KKT)



در تابع هدف پیچیدگی ابرصفحه حاشیه‌ای را کنترل می‌کند. عبارت دوم در تابع هدف، مجموع مقدار تصویر نقاط کلاس منفی (مثبت) $f_1(x)$ و $f_2(x)$ 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

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

و

$$\begin{aligned} \max_{\beta} & -\frac{1}{2} \beta^T B^T B \beta + \frac{v_2}{m_1} e_2^T A B^T \beta \\ \text{s.t.} & \quad e_2^T \beta = v_2, \\ & \quad 0 \leq \beta \leq \frac{c_2}{m_2} e_1. \end{aligned} \quad (12)$$

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

$$\omega_1^* = A^T \alpha^* - \frac{v_1}{m_2} B^T e_2, \quad b_1^* = -\frac{1}{|N_+|} \sum_{i \in N_+} x_i \omega_i^*, \quad N_+ = \{i \mid 0 \leq \alpha_i \leq \frac{c_1}{m_1} e_1\}. \quad (13)$$

و

$$\omega_2^* = B^T \beta^* - \frac{v_2}{m_1} A^T e_1, \quad b_2^* = -\frac{1}{|N_-|} \sum_{i \in N_-} x_i \omega_i^*, \quad N_- = \{i \mid 0 \leq \beta_i \leq \frac{c_2}{m_2} e_2\}. \quad (14)$$

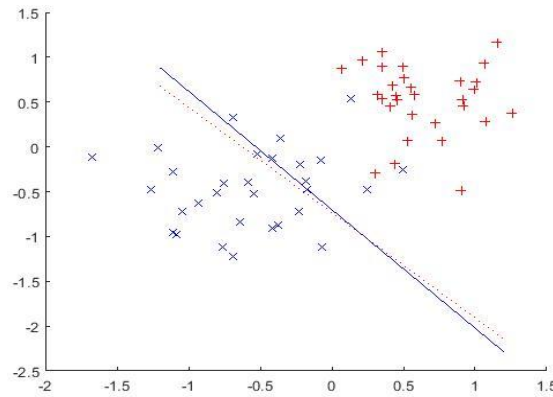
با استفاده از جواب‌های فوق، تابع تصمیم نیز به صورت زیر به دست می‌آید:

$$f(x) = \text{sign}(\hat{\omega}_1^* + \hat{\omega}_2^*)^T x + (\hat{b}_1^* + \hat{b}_2^*). \quad (15)$$

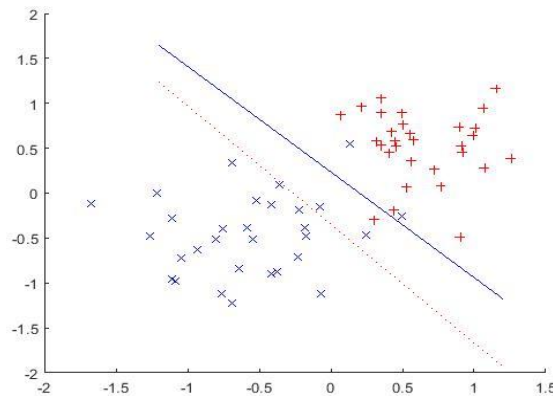
که در آن $(i=1,2) \hat{b}_i^* = b_i^* / \|b_i^*\|$ و $\hat{\omega}_i^* = \omega_i^* / \|\omega_i^*\|$ است.

۳- مدل بهبودیافته

در این بخش، مدل بهبودیافته $ITPMSVM$ (به منظور افزایش دقت کلاس بندی ارایه 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. به‌عنوان مثال، در شکل‌های 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 نیم فضای 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 نیم فضای به وجود آمده حاصل از ابرصفحه، این هدف از طریق بهینه کردن متغیر به‌طور مستقیم در تابع هدف انجام می‌شود. بدین صورت با قرار گرفتن تعداد بیشتری در طرف نیم فضای منفی حاصل از ابرصفحه کلاس دیگر و تا جای ممکن دورتر از آن، کلاس بندی مناسب‌تری تعیین می‌شود (شکل ۲). مدل بهبودیافته در دو حالت خطی و غیرخطی در ادامه بحث می‌شود.



شکل ۱- تفسیر هندسی مدل ماشین بردار پشتیبان دوگانه حاشیه-پارامتری.
Figure 1- Geometric interpretation of TPMSVM.



شکل ۲- تفسیر هندسی مدل بهبودیافته ماشین بردار پشتیبان دوگانه حاشیه-پارامتری.
Figure 2- Geometric interpretation of ITPMSVM.

۳-۱- حالت خطی ITPMSVM

مدل های بهبودیافته خطی روش پیشنهادی به صورت زیر هستند

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1, r_1} & \frac{1}{2} \|\omega_1\|^2 + v_1 r_1 + \frac{c_1}{m_1} \xi_1 \\ \text{s.t.} & A\omega_1 + e_1 b_1 \geq \xi_1, \\ & B\omega_1 + e_2 b_1 \leq -r_1 e_2, \\ & \xi_1 \geq 0. \end{aligned} \quad (16)$$

و

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2, r_2} & \frac{1}{2} \|\omega_2\|^2 + v_2 r_2 + \frac{c_2}{m_2} \xi_2 \\ \text{s.t.} & B\omega_2 + e_2 b_2 \leq \xi_2, \\ & A\omega_2 + e_1 b_2 \geq r_2 e_1, \\ & \xi_2 \geq 0, \end{aligned} \quad (17)$$

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 و متغیر r_1 (r_2) در تابع هدف نیز با بهینه شدن، نقاط کلاس منفی (مثبت) را تا جای ممکن از ابرصفحه حاشیه ای کلاس مثبت دور می کند. با توجه به سرعت حل بالای مدل های دوگان، همان بخش های قبل در ادامه مدل های دوگان مسائل فوق را ارائه می دهیم.

قضیه ۱- دوگان مدل (۱۶) به صورت زیر است:



$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T A^T A \alpha - 2\alpha^T A B^T \beta^T + \beta^T B B^T \beta \\ \text{s.t.} & \quad e_1^T \alpha = e_2^T \beta, \\ & \quad e_2^T \beta = v_1, \\ & \quad 0 \leq \alpha \leq \frac{c_1}{m_1} e_2. \end{aligned} \quad (18)$$

اثبات- تابع لاگرانژ مسئله (۱۷) به صورت زیر است

$$L(\omega_1, b_1, \xi_1, r, \alpha, \beta, \gamma) = \frac{1}{2} \|\omega_1\|^2 - v_1 r + \frac{c_1}{m_1} e_1^T \xi_1 - \alpha^T (A\omega_1 + e_1 b_1 \xi_1) + \beta^T (B\omega_1 + e_2 b_2 + r e_2) - \gamma \xi_1. \quad (19)$$

که در آن $\alpha = (\alpha_1, \dots, \alpha_{m_1})$ و $\beta = (\beta_1, \dots, \beta_{m_2})$ بردار ضرایب لاگرانژ هستند. لذا دوگان مسئله (۱۶) به صورت زیر به دست می آید:

$$\max_{\alpha, \beta, \gamma} \inf_{\omega_1, b_1, \xi_1, r} L(\omega_1, b_1, \xi_1, r, \alpha, \beta, \gamma). \quad (20)$$

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

$$\frac{\partial L}{\partial \omega_1} = \omega_1 - A^T \alpha + B^T \beta = 0, \quad (21)$$

$$\frac{\partial L}{\partial b_1} = -e_1^T \alpha + e_2^T \beta = 0, \quad (22)$$

$$\frac{\partial L}{\partial \xi_1} = \frac{c_1}{m_1} e_1 - \alpha - \gamma = 0, \quad (23)$$

$$\frac{\partial L}{\partial r_1} = -v_1 + e_2^T \beta = 0, \quad (24)$$

$$A\omega_1 + e_1 b_1 \geq -\xi_1, \quad (25)$$

$$B\omega_1 + e_2 b_2 \leq r_1 e_2, \quad (26)$$

$$\xi_1 \geq 0, \quad (27)$$

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

$$\beta \geq 0, \quad (29)$$

$$\gamma \geq 0. \quad (30)$$

از رابطه های (۲۳) و (۳۰) نیز نتیجه می شود

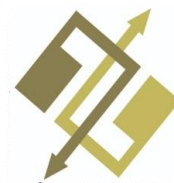
$$0 \leq \alpha \leq \frac{c_1}{m_1} e_1. \quad (31)$$

حال با جای گذاری روابط (۲۱) تا (۲۴) و رابطه (۳۱) در رابطه (۲۰)، رابطه (۱۸) به دست می آید.

قضیه ۲- دوگان مدل (۱۷) به صورت زیر است:

$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T B^T B \alpha - 2\alpha^T B A^T \beta^T + \beta^T A A^T \beta \\ \text{s.t.} & \quad e_2^T \alpha = e_1^T \beta, \\ & \quad e_1^T \beta = v_2, \\ & \quad 0 \leq \alpha \leq \frac{c_2}{m_2} e_2. \end{aligned} \quad (32)$$

اثبات- مشابه اثبات قضیه ۱ است، لذا از بیان جزییات صرف نظر می شود.



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

$$f_1(x) = \varphi(x)\omega_1 + b_1 = 0, \quad f_2(x) = \varphi(x)\omega_2 + b_2 = 0. \quad (33)$$

مشابه حالت خطی، دو ابرصفحه فوق با حل دو مسئله برنامه‌ریزی درجه دوم زیر به دست می‌آیند:

$$\begin{aligned} \min_{\omega_1, b_1, \xi_1, r_1} & \frac{1}{2} \|\omega_1\|^2 + v_1 r_1 + \frac{c_1}{m_1} \xi_1 \\ \text{s.t.} & \tilde{A}\omega_1 + e_1 b_1 \geq \xi_1, \\ & \tilde{B}\omega_1 + e_2 b_1 \leq -r_1 e_2, \\ & \xi_1 \geq 0. \end{aligned} \quad (34)$$

و

$$\begin{aligned} \min_{\omega_2, b_2, \xi_2, r_2} & \frac{1}{2} \|\omega_2\|^2 + v_2 r_2 + \frac{c_2}{m_2} \xi_2 \\ \text{s.t.} & \tilde{B}\omega_2 + e_2 b_2 \leq \xi_2, \\ & \tilde{A}\omega_2 + e_1 b_2 \geq r_2 e_1, \\ & \xi_2 \geq 0, \end{aligned} \quad (35)$$

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

قضیه ۳- دوگان مدل‌های (۳۳) و (۳۴) به صورت زیر هستند:

$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T K(A, A)\alpha - 2\alpha^T K(A, B)\beta^T + \beta^T K(B, B)\beta \\ \text{s.t.} & e_1^T \alpha = e_2^T \beta, \\ & e_2^T \beta = v_1, \\ & 0 \leq \alpha \leq \frac{c_1}{m_1} e_2. \end{aligned} \quad (36)$$

و

$$\begin{aligned} \max_{\alpha, \beta} & \alpha^T K(B, B)\alpha - 2\alpha^T K(B, A)\beta^T + \beta^T K(A, A)\beta \\ \text{s.t.} & e_2^T \alpha = e_1^T \beta, \\ & e_1^T \beta = v_2, \\ & 0 \leq \alpha \leq \frac{c_2}{m_2} e_2. \end{aligned} \quad (37)$$

اثبات- مشابه اثبات قضیه ۱ است.

¹ Mercer



در این بخش دقت کلاس‌بندی مدل پیشنهادی برای داده‌های مختلف که شامل داده‌های کتابخانه‌ای UCI برای جداسازی خطی و غیرخطی در مقایسه با مدل‌های TWSVM، TPMSVM تحلیل می‌شود. مشخصات داده‌های UCI استفاده شده برای مقایسه، در جدول ۱ ارایه و نتایج با استفاده از تست اصلاح شده فریدمن تحلیل شده است. پیاده‌سازی مدل‌ها در محیط برنامه‌نویسی متلب ۲۰۱۸ تحت سیستم عامل ویندوز ۷ و با پردازنده intel core i5 و چهار گیگابایت حافظه انجام شده است.

جدول ۱- مشخصات داده‌های کتابخانه‌ای UCI.

Table 1- Characteristics of UCI datasets benchmark.

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

۴-۱- 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:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (38)$$

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

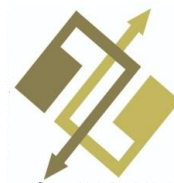
۴-۲- مقایسه دقت جداسازی مدل‌ها

نتایج حل مدل‌ها در جدول‌های ۲ و ۳ برای حالت خطی و غیرخطی برای داده‌های جدول ۱ ارایه شده است.

جدول ۲- دقت حاصل از اجرای مدل‌های کلاس‌بندی بر روی داده‌های UCI در حالت خطی.

Table 2- Accuracy results for linear classifier models on UCI datasets.

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



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 ارایه 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، به‌منظور 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 ۳- دقت حاصل از اجرای مدل‌های کلاس‌بندی بر روی داده‌های UCI در حالت غیرخطی.

Table 3- Accuracy results for non-linear classifier models on UCI datasets.

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

۴-۳- آزمایش‌های آماری

برای تحلیل نتایج عددی، از تست فریدمن^۱ (۱۹۴۰) استفاده می‌شود که این تست 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 معناداری دارند. این آزمون هر الگوریتم را برای هر داده رتبه‌بندی 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 th برای داده j th باشد. این تست میانگین رتبه برای هر کلاس‌بندی، یعنی $\bar{r}_i = \sum_{j=1}^N r_{ij}$ را باهم مقایسه می‌کند که N تعداد داده‌ها و فرمول آن به صورت زیر است:

$$X_F^2 = \frac{12N}{k(k+1)} \sum_{j=1}^k \left(\bar{r}_j - \frac{k+1}{2} \right)^2 \quad (39)$$

که در آن k تعداد الگوریتم است. ایمان و داوونپرت^۲ (۱۹۸۰) نشان دادند که آماره X_F^2 محافظه کارانه است و آماره بهتری که از رابطه (۳۹) بهره می‌گیرد ارایه کردند که به صورت زیر تعریف می‌شود:

$$F_f = \frac{X_F^2(N-1)}{N(k-1) - X_F^2} \quad (40)$$

که در آن F_f دارای توزیع F با درجه آزادی $((k-1), (k-1)(N-1))$ است. جدول‌های 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 دارای توزیع F با درجه آزادی $(2, 22)$ است. مقدار بحرانی F برای $\alpha = 0.05$ و $\alpha = 0.01$ به ترتیب 3.443 و 5.719 است. چون مقدار F_f بیشتر از مقدار بحرانی برای حالت غیرخطی است پس عملکرد سه الگوریتم مقایسه شده در حالت غیرخطی متفاوت است و از آنجایی که رتبه الگوریتم پیشنهادی کمتر از بقیه است پس عملکرد بهتری دارد.

¹ Friedman

² Iman and Davenport



جدول ۴- رتبه دقت مدل های کلاس بندی بر روی داده های UCI در حالت خطی.

Table 4- Accuracy rank of linear classifier models on UCI datasets.

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 مدل های کلاس بندی بر روی داده های UCI در حالت غیر خطی.

Table 5- Accuracy rank of non-linear classifier models on UCI datasets.

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

۵- نتیجه گیری

با توجه به این که از عوامل تعیین کننده میزان دقت کلاس بندی هر ابر صفحه در مدل *TPMSVM* تعداد داده های 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 می شود 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 ارایه گردید. مزیت مدل بهبود یافته نسبت به مدل اولیه در این است که ابر صفحه به دست آمده به گونه ای تعیین می شود که تعداد نقاط بیشتری از یک کلاس در طرف منفی نیم فضای حاصل از ابر صفحه کلاس دیگر قرار بگیرند و ابر صفحه های مناسب تری برای جداسازی داده ها به دست آید. آزمایش هایی انجام شده بر روی داده های *UCI* و مقایسه مدل ارایه شده با *TPMSVM* و *TWSVM*، نشان دهنده این است که در حالت خطی تنها مدلی که از لحاظ عملکرد به مدل ارایه شده نزدیک است مدل *TWSVM* می باشد؛ اما عملکرد مدل پیشنهادی برای حالت غیر خطی تقریباً در همه موارد بهتر از سایر مدل ها 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 مدل های مقایسه شده در دو حالت خطی و غیر خطی کمتر است.

اما در مسائل واقعی داده ها معمولاً با عدم قطعیت همراه هستند و جواب به دست آمده از حل مسئله بهینه سازی ممکن است نسبت به آشفتگی در داده ها حساس بوده و باعث خطا در جداسازی داده ها در ماشین بردار پشتیبان شود. برای رفع این مشکل، مدل های *SVM* با استفاده از قیود احتمالی (مالدونادو و همکاران^۱، ۲۰۱۶؛ ما و شن^۲، ۲۰۲۰) استوار سازی شده است، لذا در تحقیقات آتی می توان مدل

¹ Maldonado et al.

² Ma and Shen

پیشنهادی را با استفاده از قیود احتمالی استوارسازی کرد. همچنین به منظور کاهش پیچیدگی مدل برای داده‌های آموزشی بزرگ می‌توان ترکیب برنامه‌ریزی با اعداد صحیح برای انتخاب ویژگی‌های داده‌ها را با مدل ارایه شده در نظر گرفت (لی و همکاران، ۲۰۲۰).

تشکر و قدردانی

از نظرات و پیشنهادات ارزنده داوران محترم تشکر و قدردانی می‌شود.

منابع

۵۱۴



- Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3), 273-297.
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The annals of mathematical statistics*, 11(1), 86-92.
- Gupta, D., Borah, P., & Prasad, M. (2017, November). A fuzzy based Lagrangian twin parametric-margin support vector machine (FLTPMSVM). *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-7). IEEE.
- Hao, P. Y. (2010). New support vector algorithms with parametric insensitive/margin model. *Neural networks*, 23(1), 60-73.
- Iman, R. L., & Davenport, J. M. (1980). Approximations of the critical region of the fbietkan statistic. *Communications in statistics-theory and methods*, 9(6), 571-595.
- Jayadeva, H. M. (2007). *Studies on nitrogen losses, methane emission and productivity of rice under crop establishment techniques* (Doctoral dissertation, University of Agricultural Sciences). Retrieved from https://www.scitechnol.com/author-profile/H_M_JAYADEVA/
- Jimenez-Castano, C., Alvarez-Meza, A., & Orozco-Gutierrez, A. (2020). Enhanced automatic twin support vector machine for imbalanced data classification. *Pattern recognition*, 107, 107442. <https://doi.org/10.1016/j.patcog.2020.107442>
- Lee, I. G., Zhang, Q., Yoon, S. W., & Won, D. (2020). A mixed integer linear programming support vector machine for cost-effective feature selection. *Knowledge-based systems*, 203, 106145. <https://doi.org/10.1016/j.knosys.2020.106145>
- Ma, J., & Shen, J. (2020). RETRACTED: a novel twin minimax probability machine for classification and regression. *Knowledge-based systems*, 196, 105703. <https://doi.org/10.1016/j.knosys.2020.105703>
- Maldonado, S., López, J., & Carrasco, M. (2016). A second-order cone programming formulation for twin support vector machines. *Applied intelligence*, 45(2), 265-276.
- Mercer, J. (1909). Xvi. functions of positive and negative type, and their connection 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), 415-446.
- Nasiri, J. A., & Mir, A. M. (2020). An enhanced KNN-based twin support vector machine with stable learning rules. *Neural computing and applications*, 32(16), 12949-12969.
- Peng, X. (2011). TPMSVM: a novel twin parametric-margin support vector machine for pattern recognition. *Pattern recognition*, 44(10-11), 2678-2692.
- Piryonesi, S. M., & El-Diraby, T. E. (2020). Data analytics in asset management: cost-effective prediction of the pavement condition index. *Journal of infrastructure systems*, 26(1), 04019036. <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29IS.1943-555X.0000512>
- Provost, F., & Fawcett, T. (2013). *Data science for business: what you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.
- Rastogi, R., Saigal, P., & Chandra, S. (2018). Angle-based twin parametric-margin support vector machine for pattern classification. *Knowledge-based systems*, 139, 64-77.
- Rastogi, R., Sharma, S., & Chandra, S. (2018). Robust parametric twin support vector machine for pattern classification. *Neural processing letters*, 47(1), 293-323.
- Schölkopf, B. (2001). Statistical learning and kernel methods. In *Data fusion and perception* (pp. 3-24). Springer, Vienna.
- Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5), 1207-1245.
- Vapnik, V. (1998). The support vector method of function estimation. In *Nonlinear modeling* (pp. 55-85). Springer, Boston, MA.
- Wang, Z., Shao, Y. H., Bai, L., Li, C. N., Liu, L. M., & Deng, N. Y. (2018). Insensitive stochastic gradient twin support vector machines for large scale problems. *Information sciences*, 462, 114-131.

³ Lee et al.