



Presenting a Multi-Objective Mathematical Optimization Model for Classification in Data Mining

A. Kazemi* & S. Aboutaleb

*Abolfazl Kazemi, Assistant Professor, Faculty of Industrial and Mechanical Engineering, Qazvin Branch, Islamic Azad University,
Siamak Aboutaleb, M.Sc, Faculty of Industrial and Mechanical Engineering, Qazvin Branch, Islamic Azad University, siamak_a84@yahoo.com*

Keywords

Data Mining,
Optimization Model of Classification,
Multi-Objective Mathematical
Programming,
Support Vector Machine

ABSTRACT

In this paper we investigate the issues of data classification (as one of the branches of data mining science) in form of multi-objective mathematical programming model. The model that we present and investigate is a MODM problem. First time, based on support vector machine (SVM) idea (To maximize the margin of two groups), a multi-criteria mathematical programming model was proposed for data mining problems to classified instances in two separated group based on two goals of data discriminate (To maximize the distance between different groups and to minimize the misclassification of observations). Since then, some people have been working on developing the classification models based on mathematical programming methods. Simultaneously and also independently, individuals worked on support vector machine methods to improve them. Regarding to the same philosophy of these two groups of optimization methods, in this paper, to fill the gap between these two research paths, we use the updated and improved SVM methods to present a model for classification in data mining based on multi-objective programming.

© 2013 IUST Publication, IJIEPM. Vol. 23, No. 4, All Rights Reserved

*
Corresponding author. Abolfazl Kazemi
Email: abkaazemi@qiau.ac.ir



ارائه یک مدل بهینه‌سازی ریاضی چندهدفه برای طبقه‌بندی در داده‌کاوی

ابوالفضل کاظمی* و سیامک ابوطالب

چکیده:

در این مقاله به بررسی مسائل طبقه‌بندی داده‌ها (به عنوان یکی از شاخه‌های علم داده‌کاوی) در قالب مدل برنامه‌ریزی ریاضی چندهدفه می‌پردازیم. مدلی که ارائه و بررسی می‌گردد یک مسئله MODM می‌باشد. اولین بار بر پایه ایده ماشین بردار پشتیبان (SVM) (ماکزیمم کردن حاشیه دو گروه)، یک مدل برنامه‌ریزی ریاضی چندمعیاره برای مسائل داده‌کاوی بر پایه طبقه‌بندی مشاهدات به دو گروه مجزا مبتنی بر دو هدف تفکیک داده‌ها (ماکزیمم کردن فاصله بین گروه‌های مختلف و مینیمم کردن طبقه‌بندی نادرست داده‌های مورد مشاهده) معرفی شد و از آن پس تاکنون ضمن اینکه افراد زیادی روی گسترش مدل‌های طبقه‌بندی مبتنی بر روش‌های برنامه‌ریزی ریاضی کار کرده‌اند، همزمان و به صورت مستقل افرادی نیز روی بهبود روش‌های ماشین بردار پشتیبان مطالعه نموده‌اند. با توجه به فلسفه یکسان این دو دسته از روش‌های بهینه‌سازی، در این مقاله به منظور پر کردن شکاف بین این دو مسیر پژوهش، از روش‌های به روز و بهبود یافته SVM جهت ارائه مدلی به منظور طبقه‌بندی در داده‌کاوی مبتنی بر برنامه‌ریزی چندهدفه استفاده خواهد شد.

کلمات کلیدی

داده‌کاوی،
مدل بهینه‌سازی طبقه‌بندی،
برنامه‌ریزی ریاضی چندهدفه،
ماشین بردار پشتیبان.

۱. مقدمه

از دهه ۸۰ میلادی، با توسعه علوم و ارتباطات در جوامع بشری و انباشته شدن حجم عظیمی از داده‌های تجاری و علمی، نیاز به کسب دانش از داده‌های انباشته به صورت قابل ملاحظه‌ای افزایش یافت. این حجم عظیم داده‌ها که روز به روز به آنها اضافه می‌گردد، عموماً حاوی اطلاعات مهمی بوده که استخراج آنها از پایگاه‌های داده و پردازش آنها به آسانی امکان پذیر نیست. این نیاز روز افزون، منجر به رشد سریع دانش داده‌کاوی^۲ گردید. داده‌کاوی به مفهوم استخراج اطلاعات و دانش از پایگاه‌های داده در راستای کشف الگوها و مفاهیمی که به سادگی قابل درک

نیستند، می‌باشد. داده‌کاوی یکی از مراحل فرآیند کشف دانش در پایگاه داده‌ها^۳ (KDD) می‌باشد. یکی از بهترین تعاریف برای «کشف دانش در پایگاه داده» توسط Fayyad و همکاران ارائه شده است [۱]. بر اساس این تعریف KDD یک فرآیند تکرارپذیر و تعاملی و دارای چندین مرحله می‌باشد که «داده‌کاوی» قسمتی از این فرآیند می‌باشد. در تحقیقات صورت گرفته، داده‌کاوی بیشترین توجه را در بین مراحل مختلف KDD به خود معطوف کرده است و گاهاً در برخی از روشها، با قسمتی از مراحل قبیل و بعد از خود ادغام شده است. «کشف دانش در پایگاه داده» فرآیند غیربدیهی شناسایی الگوهای معتبر، جدید، بالقوه مفید و سرانجام قابل درک در داده‌ها می‌باشد که در حالت کلی دارای گامهای زیر است [۱]:

۱. درک دامنه، تعریف مسئله و جمع‌آوری داده‌ها
۲. پیش پردازش داده‌ها (آماده سازی داده‌ها جهت داده‌کاوی)
۳. داده‌کاوی (کشف الگوها از داده‌ها)

تاریخ وصول: ۹۰/۱/۱۷

تاریخ تصویب: ۹۰/۸/۷

*نویسنده مسئول مقاله: دکتر ابوالفضل کاظمی، استادیار دانشکده صنایع و مکانیک، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران
abkaazemi@qiau.ac.ir

سیامک ابوطالب، کارشناسی ارشد مهندسی صنایع، دانشکده صنایع و مکانیک، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران
siamak_a84@yahoo.com

² Data Mining

³ Knowledge Discovery in Databases

شمار می‌آید در دهه ۹۰ میلادی توسعه پیدا کرد. استفاده از روشهای برنامه‌ریزی چندهدفه (MOP) در طبقه‌بندی داده‌ها تقریباً از سال ۲۰۰۰ آغاز گردید و توسعه این روشها همچنان ادامه دارد. در کلیه این روشها، اهداف مدل به منظور تفکیک داده‌ها در گروه یا گروه‌های مجزا تعریف می‌شوند [۵].

در این مقاله، از برنامه‌ریزی ریاضی و روشهای مبتنی بر بهینه‌سازی، به منظور ارائه و بررسی مدل‌های یادگیری جهت طبقه‌بندی داده‌ها استفاده می‌کنیم. با استخراج نقاط مشترک و فلسفه یکسان تعدادی از مشهورترین مدل‌های بهینه‌سازی ریاضی در چند سال گذشته، مدل بهینه‌سازی چند هدفه جامعی پیشنهاد می‌گردد. نمایش هندسی مدل به منظور درک ساده‌تر از خصوصیات مدل ارائه شده خواهد بود. سپس نشان داده می‌شود که تعداد زیادی از مدل‌های مورد بررسی در گذشته و حال، زیر مجموعه و حالات خاصی از این مدل جامع پیشنهادی می‌باشند و نحوه تبدیل مدل جامع پیشنهادی به این روشها مورد بررسی قرار خواهد گرفت. از این طریق سعی می‌گردد تا شکاف بین مدل‌های جدید برنامه‌ریزی چندهدفه و روشهای قدرتمند و بهبود یافته SVM که به منظور طبقه‌بندی در داده‌کاوی ارائه می‌شوند را پر کرده و مطالعاتی که در راستای بهبود هر کدام از این روشها صورت می‌گیرد را قابل تعمیم به روشهای دیگر نمود. در این راستا و در ادامه مقاله، زیرمدل جدیدی بر پایه مدل جامع، ارائه خواهد شد. در ادامه و در بخش دوم، به بررسی ادبیات موضوع پرداخته خواهد شد و مطالعات صورت گرفته در زمینه تحقیق مرور می‌شود. در بخش سوم، مدل جامع پیشنهادی برای طبقه‌بندی معرفی خواهد شد و در بخش چهارم، نحوه تبدیل مدل جامع پیشنهادی به حالات خاص آن بررسی می‌شود. این حالات خاص شامل مدل‌های معرفی شده توسط سایر محققین و یک زیرمدل جدید پیشنهادی می‌باشد. در نهایت، در بخش آخر، نتایج حاصل از تحقیق جمع‌بندی شده و تحقیقات آتی مورد بررسی قرار می‌گیرد.

۲. مرور ادبیات

در سال ۱۹۳۶، Fisher اولین الگوریتم را برای شناسایی الگو ارائه نمود (الگوریتم تفکیک خطی) [۶]. Lerner و Vapnik در سال ۱۹۶۳، الگوریتم تعمیم یافته‌ای را برای ساخت ابرصفحه‌های تفکیک‌کننده با حاشیه بهینه معرفی کردند. الگوریتمی که به عنوان ماشین بردار پشتیبان ارائه شده است تعمیم غیرخطی این الگوریتم می‌باشد [۷]. در سال ۱۹۶۵، Mangasarian یک طبقه‌بندی‌کننده با حاشیه بزرگ را با استفاده از تکنیک‌های بهینه‌سازی به صورت مدل برنامه‌ریزی خطی تنظیم و ارائه کرد [۸]. بین سالهای ۱۹۸۰ تا ۱۹۹۰، Freed و Glover چند مدل

۴. پس پردازش الگوها (آماده سازی الگوها جهت استفاده از دانش کشف شده)

۵. استفاده از دانش کشف شده و پایش سیستم
یکی از کاربردهای مهم داده‌کاوی در طبقه‌بندی^۱ داده‌ها می‌باشد. در مسائل طبقه‌بندی، هر رکورد از مجموعه داده‌ها در قالب نقاطی در فضای ورودی یا فضای ویژگی توسط یک یا چند ابرصفحه^۲ از یکدیگر جدا می‌گردند. این ایده، منجر به طراحی الگوریتمی می‌شود که در آن دسته‌بندی‌کننده از طریق حل یک مدل برنامه‌ریزی ریاضی چند هدفه تولید می‌شود [۲]. الگوریتم‌های داده‌کاوی که برای ساخت مدل طبقه‌بندی‌کننده استفاده می‌شوند، در زمره الگوریتم‌های یادگیری قرار می‌گیرند. یادگیری ماشینی (ML)، به عنوان یکی از شاخه‌های وسیع و پرکاربرد هوش مصنوعی (AI)، با استخراج قوانین معنی‌دار از داده‌های خام ذخیره شده، مبنای علمی و فنی مناسبی را برای دانش داده‌کاوی ایجاد نموده است. روشهای یادگیری ماشینی شامل شیوه‌ها و الگوریتم‌هایی می‌باشند که بر اساس آنها رایانه‌ها (در کلی‌ترین مفهوم آن) به منظور کشف رفتار داده‌ها، توانایی تعلّم و یادگیری پیدا می‌کنند [۳].

در طبقه‌بندی از تکنیک‌های گسترده ای مانند درخت تصمیم، شبکه‌های عصبی، نزدیکترین k همسایه^۳، نیویز^۴، تحلیل تفکیک خطی و غیرخطی و تکنیک‌های بهینه‌سازی ریاضی استفاده می‌گردد [۳]. استفاده از روشهای مبتنی بر بهینه‌سازی، منجر به ارتقای مبنای تئوری و همچنین گسترش کاربرد عملی داده‌کاوی شده است. تغییر دادن پارامترهای یک طبقه‌بندی‌کننده، موجب دوران و حرکت انتقالی ابرصفحه می‌شود. با تغییر دادن موقعیت مرز تصمیم، کیفیت طبقه‌بندی‌کننده دچار تغییر و خطای طبقه‌بندی کمینه می‌شود. کمینه سازی این خطا از طریق بهینه‌سازی پارامترهای طبقه‌بندی‌کننده، و به کمک یادگیری میسر می‌شود. اگر طبقه‌بندی‌کننده، یک تابع خطی از پارامترهایش بوده و مشتق پذیر باشد، می‌توان به یک فرمول از نوع بسته رسید. ممکن است که راه حلی با چنین ویژگی، یک کمینه سراسری را برای شاخص عملکرد به دست دهد. اگر مساله در حقیقت غیر خطی باشد، استفاده از تکنیک‌های بهینه‌سازی نظیر تکنیک‌های مبتنی بر گرادینت، بهینه‌سازی تکاملی، و غیره اجتناب ناپذیر خواهد بود. می‌توان معروف‌ترین الگوریتم مبتنی بر بهینه‌سازی ریاضی ارائه شده تاکنون را ماشین بردار پشتیبان^۵ (SVM) دانست [۴]. این تکنیک که یکی از روشهای یادگیری به

¹ Classification

² Hyperplane

³ k-Nearest Neighbor

⁴ Naive Bayes

⁵ Support Vector Machine

Shi و همکارانش در سال ۲۰۰۵ روش برنامه‌ریزی غیر خطی چند معیاره چندگروهی (MC-MCQP) را برای مسائل تصمیم‌گیری در طبقه‌بندی داده‌کاوی ارائه نمودند [۲۸]. در سال ۲۰۰۶، Kou و همکارانش یک روش برنامه‌ریزی درجه دوم محدب چند معیاره (MCCQP) معرفی نمودند و از آن در تجزیه و تحلیل ریسک کارتهای اعتباری استفاده کرده و نتایج مقایسه آن با چند روش دیگر را نیز ارائه نمودند [۲۹]. در سال ۲۰۰۷، Shi و Zhang یک چارچوب جامع بهینه‌سازی برای طبقه‌بندی ارائه کرده و چند مدل مشهور بهینه‌سازی طبقه‌بندی (مدلهای LP و MCLP) را به عنوان حالات خاصی از مدل جامع ارائه دادند. آنها در ادامه چند مدل جدید بهینه‌سازی طبقه‌بندی را مبتنی بر مدل جامع توسعه دادند (از جمله مدل‌های MCVQP، MCCQP و MCIQP) [۳۰].

Li و همکارانش در سال ۲۰۰۸ روش برنامه‌ریزی خطی چند معیاره جریمه بسته شده (PMCLP) را در راستای بهبود روش MCLP پیشنهاد نمودند و کارایی این مدل را در مقایسه با دو مدل توسعه یافته دیگر متد MCLP، با استفاده از نمونه‌های آزمایشی بررسی کردند [۲۸]. در سال ۲۰۰۹، Kou و همکارانش یک مدل برنامه‌ریزی ریاضی درجه دو چند هدفه را برای طبقه‌بندی چند گروهی ارائه کردند. (MC-MCQP یا MC2MP) سپس به جای جستجوی جواب بهینه برای حل یک مسئله برنامه‌ریزی ریاضی محدب، نشان داده شد که محاسبه پاسخ بهینه مدل ارائه شده تنها نیاز به محاسبات ماتریسی دارد [۳۱]. Ben-Hur و Weston در سال ۲۰۱۰ مروری بر الگوریتم‌های ماشین بردار پشتیبان و انتخاب پارامترها داشتند [۳۲]. در سال ۲۰۱۱، Zhang و همکاران، مدل‌های طبقه‌بندی MCLP را با سیستم‌های خیره ادغام کردند تا نقاط ضعف هر کدام از این روشها را بپوشانند [۳۳].

استفاده از روشهای مبتنی بر بهینه‌سازی، منجر به ارتقای مبنای تئوری و همچنین گسترش کاربرد عملی داده‌کاوی (به خصوص در زمینه طبقه‌بندی داده‌ها) شده است. در قالب یک دسته‌بندی کلی، تاکنون دو دسته از روشهای طبقه‌بندی مبتنی بر بهینه‌سازی مورد مطالعه قرار گرفته اند: دسته اول، مطالعات اولیه‌ای است که در دهه ۶۰ توسط Vapnik و همکارانش و همچنین Mangasarian و همکارانش صورت گرفت و سپس در دهه ۹۰ در قالب ماشین بردار پشتیبان (SVM) به بلوغ رسید و سپس به مرور به تکامل دست یافت. دسته دوم شامل مطالعاتی است که اولین بار توسط Freed و Glover در دهه ۸۰ ارائه شد و از سال ۲۰۰۰ توسط Shi و همکارانش در قالب روشهای برنامه‌ریزی ریاضی چندهدفه خطی و غیرخطی بسط داده شد. در طول این مدت، تاکنون افراد مختلفی روی بهبود این روشها

برنامه‌ریزی خطی را به منظور حل مسائل تفکیک‌کننده با نمونه کوچک داده‌ها ارائه نمودند [۹]. ماشین بردار پشتیبان به شکل نزدیک به روش فعلی، برای اولین بار در قالب یک مقاله در سال ۱۹۹۲ توسط Boser و همکاران معرفی شد. آنها همچنین در این مقاله راهی را برای ساخت طبقه‌بندی‌کننده‌های غیرخطی با حداکثر حاشیه با استفاده از کاربرد Kernel Trick در نگاشت غیرخطی ابرصفحه‌های بهینه به فضای ویژگی (با ابعاد بیشتر) ارائه کردند. الگوریتم PCG-Chunking جهت حل مدل ارائه گردید [۱۰]. در سال ۱۹۹۵، Cortes و Vapnik، ایده حداکثر حاشیه اصلاح شده را مطرح نمودند. در این روش، خطای طبقه‌بندی نادرست در مدل SVM در نظر گرفته شده است و الگوریتم حداکثر حاشیه با تعریف متغیر کمبود در مدل بهینه‌سازی، به مسائلی که بصورت خطی جدا پذیر نیستند، تعمیم داده شد (C-SVC) [۱۱]. در سال ۱۹۹۷، Osuna و همکاران، رسماً اولین الگوریتم تجزیه برای آموزش مسائل با داده‌هایی در ابعاد بزرگ را برای SVM ارائه نمودند. آنها همچنین روشی را جهت برخورد با کلاسهای نامتعادل ارائه نمودند [۱۲]. در سال ۱۹۹۹، Mangasarian مدل نرم اول SVM را ارائه کرد [۱۳]. در سال ۱۹۹۹، Suykens و Vandewalle مدل LS-SVC را ارائه کردند [۱۴]. در سال ۲۰۰۰، Schölkopf و همکاران مدل v-SVC را ارائه کردند [۱۵]. Mangasarian و همکاران در سال ۲۰۰۱ تعدادی الگوریتم برای SVM ارائه کردند. از جمله می‌توان به ASVM، LSVM، RSVM، PSVM اشاره کرد. آنها همچنین از روش Newton Armijo برای حل تعدادی از مدل‌های خود استفاده کردند [۱۶-۲۰]. در سال ۲۰۰۱، Shi و همکاران از یک رویکرد مبتنی بر برنامه‌ریزی چندمعیاره به منظور معرفی مدل MCLP استفاده کردند و به این ترتیب استفاده از مدل‌های بهینه‌سازی ریاضی را در مسائل داده‌کاوی گسترش دادند [۲۱]. در سال ۲۰۰۲، Bugera و همکاران یک تابع مطلوبیت درجه دوم محدب را برای SVM معرفی کردند و از آن برای مدل‌سازی مسئله طبقه‌بندی حساب کارتهای اعتباری استفاده کردند [۲۲]. Peng و همکارانش در سال ۲۰۰۲ بهبود یافته روش برنامه‌ریزی خطی Shi را برای طبقه‌بندی چند گروهی پیشنهاد دادند (MC-MCLP) [۲۳]. Hsu و همکاران در سال ۲۰۰۳ یک رویکرد مبتنی بر فرآیند برای استفاده از SVM ارائه کردند [۲۴]. در سال ۲۰۰۳، Bi مدل C-SVM را در قالب یک مسئله برنامه‌ریزی چندهدفه ارائه نمود و از روش ϵ -Constraint برای حل آن استفاده کرد [۲۵]. در سال ۲۰۰۳، Perez-Cruz و همکاران مدل v-SVC را به مدل Ev-SVC توسعه دادند [۲۶]. Asada و همکاران در سال ۲۰۰۴ از رویکرد برنامه‌ریزی آرمانی (GP) در طبقه‌بندی توسط SVM استفاده کردند [۲۷].

$X = (X_1, \dots, X_m)^T$	ماتریس ضرایب
$X_i = (x_{i,1}, \dots, x_{i,l})$	بردار ضرایب (بردار قائم بر ابرصفحه جداکننده l ام)
$x_{j,l} \in \mathbb{R}$	مقدار ضریب خصیصه j ام نسبت به ابرصفحه جداکننده l ام
$A_i \cdot X_i$	امتیاز طبقه‌بندی رکورد i ام نسبت به ابرصفحه جداکننده l ام
$B = (b_1, \dots, b_m)$	بردار مرزی بایاس
$b_i \in \mathbb{R}$	اسکالر مرزی بایاس جدا کننده کلاس‌های l ام و $l+1$ ام
$\alpha = (\alpha_1, \dots, \alpha_n)^T$	ماتریس انحراف
$\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,m})$	فاصله امتیاز رکورد i ام با طبقه‌بندی نادرست یا خطای حاشیه از ابرصفحه مرزی تنظیم شده
$\alpha_{i,l} \geq 0$	فاصله امتیاز رکورد i ام با طبقه‌بندی نادرست یا خطای حاشیه از ابرصفحه مرزی تنظیم شده جداکننده l ام
$\beta = (\beta_1, \dots, \beta_n)^T$	ماتریس انطباق
$\beta_i = (\beta_{i,1}, \dots, \beta_{i,m})$	فاصله امتیاز رکورد i ام با طبقه‌بندی صحیح یا صحت حاشیه از ابرصفحه مرزی تنظیم شده
$\beta_{i,l} \geq 0$	فاصله امتیاز رکورد i ام با طبقه‌بندی صحیح یا صحت حاشیه از ابرصفحه مرزی تنظیم شده جداکننده l ام

۳-۲. فرموله کردن مسئله و محدودیت‌های مدل جامع

پیشنهادی

با داشتن مجموعه داده‌های آموزشی با n رکورد و r خصیصه و همچنین با دانستن این موضوع که هر رکورد (A_i) از این مجموعه متعلق به کدام کلاس از پیش تعریف شده می‌باشد ($\{G_1, \dots, G_{m+1}\}$)، مسئله طبقه‌بندی داده‌ها در قالب مدل ریاضی معرفی می‌گردد. به منظور پیدا کردن تابع تصمیم، اسکالره‌های مرزی بایاس (b_i) و بردار ضرایب (X_i) در ابرصفحه‌های طبقه‌بندی‌کننده ($A_i \cdot X_i = b_i$) با تصدیق در نامساوی‌های زیر به ازای $1 \leq i \leq n$ شناسایی می‌شوند:

$$A_i \cdot X_1 \leq b_1 \quad \forall A_i \in G_1 \\ A_i \cdot X_{k-1} > b_{k-1}, \quad A_i \cdot X_k \leq b_k \quad (2 \leq k \leq m) \quad (1)$$

$$m) \quad \forall A_i \in G_k \\ A_i \cdot X_m > b_m \\ \in G_{m+1}$$

نامساوی‌های فوق را می‌توان به ازای $1 \leq i \leq n$ و $1 \leq l \leq m$ به شکل زیر تبدیل کرد:

$$A_i \cdot X_l \leq b_l \quad \forall A_i \in G_1 \cup \dots \cup G_l \\ A_i \cdot X_l > b_l \quad \forall A_i \in G_{l+1} \cup \dots \cup G_{m+1} \quad (2)$$

مطالعه نموده‌اند. مطالعات صورت گرفته در دو زمینه «ارائه یا توسعه مدل‌های بهینه‌سازی» و «ارائه یا توسعه الگوریتم حل مدل‌ها» بوده است. در این میان، نکته قابل توجه، فلسفه یکسان هر دو دسته از روش‌های طبقه‌بندی مبتنی بر بهینه‌سازی می‌باشد.

۳. مدل جامع بهینه‌سازی پیشنهادی

در این بخش، پس از تعریف نمادهای مشترک، محدودیت‌ها و توابع هدف مدل جامع، طبقه‌بندی‌کننده را به صورت هندسی و همچنین در قالب برنامه‌ریزی ریاضی، مدل‌سازی خواهیم کرد. مدل جامع طبقه‌بندی، به گونه‌ای طراحی شده است که مدل‌های اولیه و بهبود یافته SVM و همچنین مدل‌های برنامه‌ریزی ریاضی یک و چندهدفه خطی و غیرخطی که تاکنون ارائه شده‌اند را به عنوان حالات خاصی از خود دربرگیرد و همچنین، ایجاد حالات خاص جدید نیز امکان‌پذیر باشد.

۳-۱. نمادگذاری

به منظور ایجاد قابلیت مقایسه بین مدل‌های مختلف و بررسی ساده‌تر آنها، نمادهای مشترکی را تعریف نموده و از قالب این نمادگذاری در طول مقاله استفاده خواهد شد.

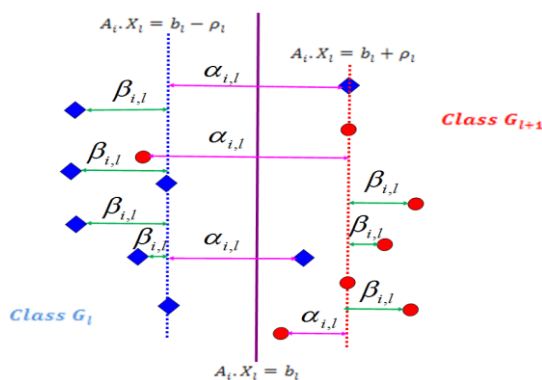
r	تعداد خصیصه‌ها
$a = (a_1, \dots, a_r)$	بردار خصیصه‌ها
a_j	خصیصه j ام
$1 \leq j \leq r$	بطوری که:
n	تعداد رکوردها
$A = (A_1, \dots, A_n)^T$	ماتریس رکوردها
$A_i = (a_{i,1}, \dots, a_{i,r})$	رکورد i ام
$1 \leq i \leq n$	بطوری که:
$a_{i,j} \in \mathbb{R}$	مقداری که رکورد i ام به ازای خصیصه j ام اختیار کرده است
m	تعداد ابرصفحه‌های جدا کننده (که همواره یکی کمتر از تعداد کلاس‌ها می‌باشد)
$y = (y_1, \dots, y_m)$	بردار برچسب‌ها (مشخص‌کننده وضعیت عضویت-درکلاس‌های $\{G_1, \dots, G_{m+1}\}$)
y_l	برچسب l ام
$1 \leq l \leq m$	بطوری که:
$Y = (Y_1, \dots, Y_n)^T$	ماتریس برچسب‌ها
$Y_i = (y_{i,1}, \dots, y_{i,m})$	برچسب حقیقی رکورد i ام
$y_{i,l} \in \{-1, +1\}$	مقداری که برچسب حقیقی رکورد i ام نسبت به ابرصفحه جداکننده l ام اختیار کرده است
$\text{if } y_{i,l} = -1 \text{ then } A_i \in G_1 \cup \dots \cup G_l$ $\text{if } y_{i,l} = +1 \text{ then } A_i \in G_{l+1} \cup \dots \cup G_{m+1}$	بطوری که:

فراهم می‌کنیم. به منظور تکمیل محدودیت‌های تفکیک داده‌ها، می‌بایست تعدادی از متغیرها معرفی شوند. فرض می‌کنیم $\alpha_{i,l} \geq 0$ به مفهوم فاصله از $A_i \cdot X_l$ تا مرز تنظیم‌شده طبقه‌بندی‌کننده l ام است در صورتی که A_i به صورت نادرست در گروه دیگری طبقه‌بندی شده باشد و یا در حاشیه مثبت قرار گرفته باشد. در حالت اول به $\alpha_{i,l}$ خطای طبقه‌بندی نادرست و در حالت دوم به آن خطای حاشیه می‌گوییم. به عبارت دیگر، این فاصله میزان انحراف یک نمونه را از وضعیت ایده‌آل تفکیک پذیری می‌سنجد. بنابراین با در نظر گرفتن این مقدار خطا، نامساوی‌های فوق، به ازای $1 \leq l \leq m$ و $1 \leq i \leq n$ به صورت زیر درمی‌آیند:

$$\begin{aligned} A_i \cdot X_l - b_l &\leq -\rho_l + \alpha_{i,l} \quad \forall A_i \in G_1 \cup \dots \cup G_l \quad (5) \\ A_i \cdot X_l - b_l &> +\rho_l - \alpha_{i,l} \quad \forall A_i \\ &\in G_{l+1} \cup \dots \cup G_{m+1} \end{aligned}$$

فرض می‌کنیم $\beta_{i,l} \geq 0$ به مفهوم فاصله از $A_i \cdot X_l$ تا مرز تنظیم‌شده طبقه‌بندی‌کننده l ام است در صورتی که A_i به صورت صحیح طبقه‌بندی شده باشد و یا در حاشیه منفی قرار گرفته باشد (صحت حاشیه). به عبارت دیگر، این فاصله میزان انطباق یک نمونه را با وضعیت ایده‌آل تفکیک پذیری می‌سنجد. با اضافه کردن $\beta_{i,l}$ ، محدودیت‌ها از حالت نامساوی به حالت تساوی تبدیل می‌گردند به ازای $1 \leq l \leq m$ و $1 \leq i \leq n$ به صورت زیر درمی‌آیند:

$$\begin{aligned} \forall A_i \in G_1 \cup \dots \cup G_l \quad (6) \\ A_i \cdot X_l - b_l &= -\rho_l + \alpha_{i,l} - \beta_{i,l} \\ A_i \cdot X_l - b_l &= +\rho_l - \alpha_{i,l} + \beta_{i,l} \quad \forall A_i \\ &\in G_{l+1} \cup \dots \cup G_{m+1} \end{aligned}$$



شکل ۲. نمایش هندسی $\alpha_{i,l}$ و $\beta_{i,l}$ در طبقه‌بندی‌کننده l ام با حاشیه انعطاف‌پذیر

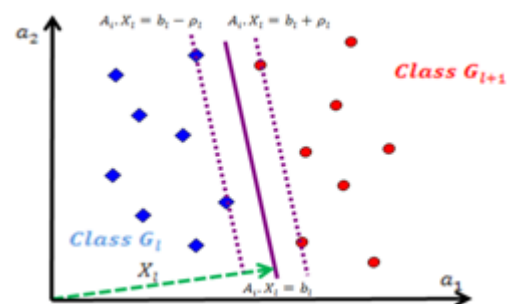
با دانستن این که A_i متعلق به کدامیک از کلاس‌های از پیش تعریف شده می‌باشد ($\{G_1, \dots, G_{m+1}\}$), برچسب کلاس A_i

شایان ذکر است، $A_i \cdot X_l$ به عنوان امتیاز طبقه‌بندی l امین رکورد داده‌ها نسبت به ابرصفحه جداکننده l ام در نظر گرفته می‌شود که ترکیب آن با b_l ، تابع تصمیم را می‌سازد. علامت « \cdot » نشانگر ضرب نقطه‌ای دو بردار می‌باشد:

$$A_i \cdot X_l = A_i X_l^T = \sum_{j=1}^r a_{i,j} x_{j,l} \quad (3)$$

ابرفرصه‌های طبقه‌بندی‌کننده‌ای که بتوانند نقاط را از هم جدا کنند، منحصر به فرد نمی‌باشند. بر اساس تئوری VC¹ [۳۴]، به منظور افزایش توانایی تعمیم، ابرصفحه‌ای انتخاب می‌شود که حاشیه را به مقدار بهینه خود برساند. حاشیه از دیدگاه هندسی عبارت است از فاصله موجود بین ابرصفحه طبقه‌بندی‌کننده و نزدیکترین نمونه آموزشی. به منظور ساخت حاشیه بهینه، دو ابرصفحه مرزی تنظیم شده، موازی با ابرصفحه طبقه‌بندی‌کننده l ام رسم کرده ($A_i \cdot X_l - b_l = \pm \rho_l$)، آن دو را به یک اندازه آنقدر از هم دور می‌کنیم که به داده‌ها برخورد کنند. ابرصفحه طبقه‌بندی‌کننده‌ای که بیشترین فاصله را از ابرصفحات مرزی داشته باشد، بهترین طبقه‌بندی‌کننده خواهد بود. به منظور جلوگیری از قرار گرفتن نقاط داده در داخل حاشیه، به ازای $1 \leq l \leq m$ و $1 \leq i \leq n$ محدودیت‌های زیر را اضافه می‌کنیم:

$$\begin{aligned} A_i \cdot X_l - b_l &\leq -\rho_l \quad \forall A_i \in G_1 \cup \dots \cup G_l \quad (4) \\ A_i \cdot X_l - b_l &> +\rho_l \quad \forall A_i \in G_{l+1} \cup \dots \cup G_{m+1} \end{aligned}$$



شکل ۱. ابرصفحه طبقه‌بندی‌کننده با حاشیه بهینه

زمانی که تفکیک خطی نمونه‌ها بطور کامل شدنی نباشد و تعدادی از نقاط متعلق به کلاس‌های مختلف با یکدیگر مخلوط شده باشند (در اکثر مسائل واقعی)، این امکان وجود دارد که فضای حل موجه مسئله با محدودیت‌های فوق، تهی باشد. بنابراین، با کاستن از شدت نامساوی‌های فوق و ایجاد حاشیه انعطاف‌پذیر^۲ (نرم)، امکان وجود داده‌های حاوی اختلال را نیز

¹ Vapnik Chervonenkis Theory

² Soft Margin

مدل‌های طبقه‌بندی مبتنی بر بهینه‌سازی، مقدار حاشیه $\rho_l = 0$ در نظر گرفته شده است که در نتیجه این تابع در هدف مدل انعکاس نخواهد داشت. اکثر روشهای برنامه‌ریزی ریاضی چندهدفه جزو این دسته از مدل‌ها می‌باشند.

ب) بهینه‌سازی جریمه تجمیعی خطای طبقه‌بندی نادرست و خطای حاشیه

دومین تابع مورد بررسی که یکی از نقاط مشترک هر دو دسته از روشهای طبقه‌بندی مبتنی بر بهینه‌سازی می‌باشد، بیشترین وزن را در بین سایر توابع مسئله بهینه‌سازی دارا است. خطای حاشیه، در مدل‌های ماکزیمم حاشیه مثبت اتفاق می‌افتد. در این حالت، مشاهده در فضای حاشیه قرار گرفته است که می‌بایست از این حالت نیز اجتناب کرد.

به منظور کنترل میزان انحراف مشاهدات از وضعیت ایده‌آل تفکیک پذیری آنها، مقدار تجمیعی خطای طبقه‌بندی نادرست را با استفاده از تابع $\|\alpha\|_p^p$ مینیمم می‌کنیم. منظور از $\|\alpha\|_p^p$ نرم p ام ماتریس α می‌باشد که به توان p رسیده است. از آنجایی که $\alpha_{il} \geq 0$ می‌باشد، تابع فوق به شکل زیر در تابع هدف ظاهر می‌شود:

$$\text{Minimize: } \sum_i \sum_l \alpha_{i,l}^p \quad (11)$$

در مدل‌های طبقه‌بندی مبتنی بر بهینه‌سازی، از نرم‌های مختلفی برای تابع خطا استفاده شده است. تنها در صورتی می‌توان از این تابع صرف نظر کرد ($\alpha_{j,l} = 0$) که در مسئله طبقه‌بندی تمامی نقاط کاملاً به صورت خطی تفکیک‌پذیر باشند. در غیر این صورت، فضای حل موجه مسئله، تهی شده و حل مسئله بهینه‌سازی امکان‌پذیر نمی‌باشد.

ضریبی که به این تابع در هدف مدل اختصاص می‌یابد، جریمه خطا بوده و از پارامترهای با اهمیتی می‌باشد که بر ایجاد تعادل بین پیچیدگی و دقت طبقه‌بندی تاثیر بسزایی دارد. هرچه این ضریب بزرگتر باشد، حساسیت بیشتری نسبت به خطاهایی که اجازه می‌دهیم اتفاق بیافتد، در نظر گرفته می‌شود. در حقیقت این ضریب، یک مصالحه بین پهنای حاشیه و جداسازی اشتباه است. به این ضریب، پارامتر تنظیم‌کننده نیز گفته می‌شود که می‌بایست توسط DM انتخاب شود.

ج) بهینه‌سازی پاداش تجمیعی موفقیت در طبقه‌بندی صحیح و صحت حاشیه

سومین تابع هدف، مختص روش‌های برنامه‌ریزی ریاضی بوده و در مسائل کلاسیک SVM مشاهده نمی‌شود. این تابع، پاداش تجمیعی موفقیت در طبقه‌بندی صحیح و صحت حاشیه را ماکزیمم می‌کند.

نسبت به ابرصفحه جداکننده l ام در قالب $Y_{i,l} \in \{-1, +1\}$ و به صورت زیر قابل تعریف است:

$$\begin{aligned} \text{if } y_{i,l} = -1 & \text{ then } A_i \in G_1 \cup \dots \cup G_l \quad (7) \\ \text{if } y_{i,l} = +1 & \text{ then } A_i \in G_{l+1} \cup \dots \cup G_{m+1} \end{aligned}$$

بنابراین با داشتن رکوردهای مجموعه آموزشی (A_i) و برچسب‌های آنها (Y_i) می‌توان مساوی‌های فوق را در قالب یک محدودیت بازنویسی کرد:

$$y_{i,l}(A_i \cdot X_l - b_l) = \rho_l - \alpha_{i,l} + \beta_{i,l} \quad (8)$$

به منظور تکمیل محدودیت‌های مدل، می‌بایست موارد زیر را نیز در نظر گرفت:

$$\begin{aligned} \alpha_{i,l}, \beta_{i,l} & \geq 0 \\ 1 & \leq i \leq n \\ 1 & \leq l \leq m \\ x_j, b_l & \text{ are unrestricted} \end{aligned} \quad (9)$$

۳-۳. توابع هدف مدل جامع پیشنهادی

در روشهای طبقه‌بندی مبتنی بر بهینه‌سازی، توابع هدف مدل‌های برنامه‌ریزی ریاضی از ترکیب اهداف زیر حاصل می‌گردند:

الف) بهینه‌سازی حاشیه بین ابرصفحات مرزی تنظیم شده موازی با ابرصفحات طبقه‌بندی‌کننده

اولین تابع مورد بررسی، ریشه در تئوری VC و روش‌های مبتنی بر SVM دارد. با توجه به اینکه بردار ضرایب (X_l) ، بردار قائم بر ابرصفحات طبقه‌بندی‌کننده می‌باشد، پس بر ابرصفحات مرزی تنظیم شده موازی نیز عمود است، لذا فاصله بین این دو ابرصفحه که همان پهنای حاشیه می‌باشد مضرری از X_l خواهد بود. با استفاده از محاسبات هندسی ساده پهنای حاشیه برای ابرصفحه جداکننده l ام برابر با $\frac{2\rho_l}{\|x_l\|_2}$ است. منظور از $\|x_l\|_2^2$ نرم 2 ام تابع X_l می‌باشد که به توان 2 رسیده است.

$$\|X_l\|_2^2 = \sum_{j=1}^r |x_{j,l}|^2 \quad (10)$$

در مدل‌های طبقه‌بندی مبتنی بر بهینه‌سازی، علاوه بر استفاده از نرم‌های مختلف، برخوردهای مختلفی نیز با تابع بهینه‌ساز حاشیه شده است:

الف-۱) ماکزیمم کردن حاشیه مثبت (PM-Max): در این حالت فرض بر این است که $\rho_l \geq 0$ می‌باشد.

الف-۲) مینیمم کردن حاشیه منفی (NM-Min): در این حالت فرض بر این است که $\rho_l \leq 0$ می‌باشد. ضمناً در تعدادی از

- روش برنامه‌ریزی آرمانی (در مسئله تفکیک خطی در [۳۵] و برای مسئله SVM در [۲۷])
- روش قید \mathcal{E} (در مدل C-SVM در [۲۵])
- روش وزن‌دهی (در حالات خاص مدل جامع پیشنهادی بررسی می‌گردند)

از بین روش‌های مذکور، متداول‌ترین روش، متد وزن‌دهی می‌باشد. با استفاده از این روش، مدل چند هدفه با استفاده از اختصاص ضریب وزنی مثبت به توابع هدف، به یک مسئله یک هدفه بهینه‌سازی مجموع وزنی توابع، تبدیل می‌گردد. این روش به DM اجازه می‌دهد تا بر اساس اولویت‌ها و پس از هر تکرار، اوزان را اصلاح نماید. این روش در مدل‌های برنامه‌ریزی خطی و محدب، کارایی مناسبی دارد و نقطه ضعف آن در فضاهای غیرمحدب می‌باشد. ضمناً توزیع یکنواخت اوزان در هر تکرار، لزوماً منجر به ارائه مجموعه پاسخ‌های یکنواخت نمی‌گردد.

۴. حالات خاص مدل جامع پیشنهادی و نحوه تبدیل

استفاده از نُرم‌های مختلف (p, q, r, s, t) بر اهداف مدل طبقه‌بندی تأثیرگذار است و منجر به ایجاد مسائل بهینه‌سازی خطی و غیرخطی (درجه دوم محدب، درجه دوم مقعر و درجه دوم نامعین) می‌شود.

بستگی به اینکه مدل بهینه‌سازی به چه شکلی ساخته شود، روش حل مدل پیشنهاد می‌گردد. بعضاً به منظور حل یک مدل خاص، چند روش حل نیز ارائه شده است. حل مسئله اصلی یا دوگان و استفاده از روش‌های مستقیم یا متدهای تجزیه^۱، در کنار ساختار حل، در شکل‌گیری روش‌های حل مدل تأثیرگذار است. برخی از این متدها جواب بهینه سراسری و برخی دیگر پاسخ بهینه محلی مدل را پیدا می‌کنند. این حالات خاص و نحوه تبدیل مدل جامع پیشنهادی به آنها، به تفکیک مدل‌های دو کلاسی و چند کلاسی، در این بخش بررسی خواهند شد و در انتها، زیرمدل جدید پیشنهادی ارائه می‌گردد.

۴-۱. زیرمدل‌های دو کلاسی

در مدل‌های دو کلاسی که حالت خاصی از مدل جامع ارائه شده می‌باشند، تنها دو کلاس (G_1 و G_2) و یک طبقه‌بندی‌کننده و در نتیجه یک ابرصفحه جداکننده ساخته می‌شود ($m = 1$). در این حالت اندیس l کاربرد نداشته و در نمادگذاری نمایش داده نخواهد شد. ضمناً بر اساس تعریف قبلی و به عنوان یک حالت خاص، برچسب کلاس A_i نسبت به تنها ابرصفحه جداکننده به صورت زیر قابل تعریف است:

صحت حاشیه، در مدل‌های مینیمم حاشیه منفی اتفاق می‌افتد. در این حالت، مشاهده در فضای حاشیه منفی قرار گرفته است که می‌بایست از این حالت نیز اجتناب کرد. به منظور کنترل میزان انطباق مشاهدات با وضعیت ایده‌آل تفکیک پذیری آنها، مقدار تجمیعی موفقیت در طبقه‌بندی صحیح را با استفاده از تابع $\|\beta\|_q^q$ ماکزیمم می‌کنیم. منظور از $\|\beta\|_q^q$ نُرم q ام ماتریس β می‌باشد که به توان q رسیده است. از آنجایی که $\beta_{il} \geq 0$ می‌باشد، تابع فوق به شکل زیر در تابع هدف ظاهر می‌شود:

$$\text{Maximize: } \sum_i \sum_l \beta_{i,l}^q \quad (12)$$

در مدل‌های طبقه‌بندی مبتنی بر بهینه‌سازی، از نُرم‌های مختلفی برای تابع خطا استفاده شده است. ضریبی که به این تابع در هدف مدل اختصاص می‌یابد، پاداش طبقه‌بندی بوده و مقدار کوچکتری از جریمه خطا دارد که می‌بایست توسط DM انتخاب شود.

در مسائل طبقه‌بندی چندگروهی، نقش این تابع در برخی مدل‌ها پررنگ‌تر می‌شود و بعضاً نقش یک محدودیت، جهت کنترل موقعیت گروه‌های مجاور و یا تضمین کننده ساخت تعداد کافی گروه‌ها با تفکیک پذیری مناسب را ایفا می‌کند. این نقش معمولاً با مینیمم کردن اختلاف $\beta_{i,l}$ از مقادیر تعریف شده‌ای (f_l) صورت می‌گیرد. در این حالت، تابع سوم در قالب دو قسمت در تابع هدف ظاهر می‌شود. به طوریکه، قسمت اول این تابع در محدوده ابرصفحات اول و آخر و قسمت دوم آن در محدوده بین ابرصفحات، فعال می‌گردد:

$$\text{Maximize: } \sum_i \sum_l \beta_{i,l}^q - \sum_i \sum_l |\beta_{i,l} - f_l|^t \quad (13)$$

استفاده از ترکیبی از اهداف ذکر شده، منجر به ایجاد مدل‌های مختلفی می‌گردد. این ترکیب اغلب به یکی از سه حالت زیر مشاهده می‌شود:

۱. ترکیبی از هدف اول و دوم (در مدل‌های SVM رایج است).
۲. ترکیبی از هدف دوم و سوم (در مدل‌های برنامه‌ریزی ریاضی رایج است).
۳. ترکیبی از اهداف اول، دوم و سوم (در مدل‌های توسعه یافته اخیر استفاده می‌شود).

کارایی نهایی طبقه‌بندی‌کننده، به بهینه‌سازی همزمان توابع هدف مورد استفاده در مدل بستگی دارد. وجود دو یا سه تابع هدف در مدل، منجر به استفاده از روش‌های برنامه‌ریزی ریاضی چندهدفه (MOP) می‌گردد. به منظور حل این مدل چندهدفه، از روش‌های مختلفی استفاده شده است:

- روش شاخص جهانی (L_p) یا راه حل سازشی (در مدل‌های MCLP و MC-MCLP در [۲۱] و [۲۳])
- روش تابع مطلوبیت (معرفی یک تابع مطلوبیت درجه دوم محدب برای مسئله SVM در [۲۲])

¹ Decomposition Methods

$$\text{Minimize } \frac{1}{2} \|X\|_2^2 + W_\alpha \sum_{i=1}^n \alpha_i$$

S.T.:

$$y_i(A_i \cdot X - b) \geq 1 - \alpha_i$$

$$\alpha_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۱-۳. مدل 1-Norm-SVC - مدل نرم اول طبقه‌بندی کننده بردار پشتیبان

این مدل اولین بار در سال ۱۹۹۹ توسط Mangasarian ارائه شد [۱۳]. تفاوت این مدل با مدل C-SVC در نرم تابع هدف اول می‌باشد. در این مدل از نرم‌های $s = 1$ و $p = 1$ استفاده شده است. همانند دو مدل قبلی، تابع هدف اول در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $p = 1$ به صورت ثابت در نظر گرفته می‌شود و تابع هدف دوم با روش وزن دهی و با ضریب وزنی به شکل $W_\alpha = \frac{\gamma}{2}$ به تابع هدف اول اضافه شده است

$$\text{Minimize: } \frac{1}{2} \|X\|_1 + W_\alpha \sum_{i=1}^n \alpha_i$$

S.T.:

$$W_\alpha = \frac{\gamma}{2}$$

$$y_i(A_i \cdot X - b) \geq 1 - \alpha_i$$

$$\alpha_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۱-۴. مدل LS-SVC - مدل کوچک‌ترین مربعات طبقه‌بندی کننده بردار پشتیبان

این مدل اولین بار در سال ۱۹۹۹ توسط Suykens و Vandewalle ارائه شد [۱۴]. تفاوت این مدل با مدل C-SVC در نرم تابع هدف دوم می‌باشد. در این مدل از نرم‌های $s = 1$ و $p = 2$ استفاده شده است. همانند مدل‌های قبلی، تابع هدف اول در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $p = 1$ به صورت ثابت در نظر گرفته می‌شود و تابع هدف دوم با روش وزن دهی و با ضریب وزنی به شکل $W_\alpha = \frac{\gamma}{2}$ به تابع هدف اول اضافه شده است.

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2 + W_\alpha \sum_{i=1}^n \alpha_i^2$$

S.T.:

$$W_\alpha = \frac{\gamma}{2}$$

$$\forall A_i \in G_1 \quad A_i \cdot X \leq b \quad y_i = -1$$

$$\forall A_i \in G_2 \quad A_i \cdot X > b \quad y_i = +1$$

بر همین اساس، $A_i \cdot X$ به عنوان امتیاز طبقه‌بندی i امین رکورد داده‌ها نسبت به ابرصفحه جداکننده در نظر گرفته می‌شود که ترکیب آن با b ، تابع تصمیم را می‌سازد.

۴-۱-۱. زیرمدل‌های مبتنی بر SVM

مدل‌های مبتنی بر SVM، اغلب با ترکیبی از توابع اول و دوم ساخته می‌شوند. با توجه به اهمیت اساسی بحث حاشیه در این مدل‌ها، در کلیه آنها تابع $\|X\|_s^s$ مشاهده می‌شود.

۴-۱-۱-۱. مدل SVC - مدل پایه‌ای و مقدماتی طبقه‌بندی کننده بردار پشتیبان با حاشیه سخت (انعطاف ناپذیر)

این مدل اولین بار در سال ۱۹۹۲ توسط Boser و همکاران ارائه شد [۱۰] و فقط شامل هدف اول می‌باشد که به صورت ماکزیمم کردن حاشیه مثبت ظاهر می‌گردد. بنابراین، این مدل، تنها قادر به حل مسائلی می‌باشد که کاملاً به صورت خطی تفکیک‌پذیر باشند. در این مدل $p = 1$ به صورت ثابت در نظر گرفته می‌شود و $\frac{1}{2} \|X\|_s^s$ مینیمم می‌گردد. در این مدل از نرم $s = 2$ استفاده شده است.

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2$$

S.T.:

$$y_i(A_i \cdot X - b) \geq 1$$

$$1 \leq i \leq n$$

۴-۱-۱-۲. مدل C-SVC - مدل کلاسیک طبقه‌بندی کننده بردار پشتیبان با حاشیه نرم (انعطاف پذیر)

این مدل اولین بار در سال ۱۹۹۵ توسط Cortes و Vapnik ارائه شد [۱۱] و با اضافه کردن هدف دوم به مدل SVC، منجر به ایجاد حاشیه انعطاف پذیر شده است. همانند مدل SVC، تابع هدف اول در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $p = 1$ به صورت ثابت در نظر گرفته می‌شود. تابع هدف دوم با روش وزن دهی و با ضریب وزنی $W_\alpha > 0$ به تابع هدف اول اضافه شده است. در این مدل از نرم‌های $s = 2$ و $p = 1$ استفاده شده است. این مدل مبنای توسعه‌های آتی روش‌های مبتنی بر SVM می‌باشد.

$$\text{Minimize: } -W_p \rho + W_\alpha \sum_{i=1}^n \alpha_i$$

$$S.T.:$$

$$(20)$$

$$W_\alpha = \frac{1}{n}$$

$$y_i(A_i \cdot X - b) \geq \rho - \alpha_i$$

$$\|X\|_2^2 = 2$$

$$\rho \leq 0$$

$$\alpha_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۱-۷. مدل MOP-SVC - قالب برنامه‌ریزی چند

هدفه^۱ طبقه‌بندی‌کننده بردار پشتیبان

این مدل اولین بار در سال ۲۰۰۴ توسط Asada و همکاران ارائه شد [۲۷]. به منظور توسعه مدل‌های مبتنی بر SVM، تابع هدف سوم، معرفی و به توابع اول و دوم اضافه گردیده است. تابع هدف سوم که در قالب ماکزیمم سازی معرفی شده است، به صورت مینیمم نمودن منفی آن تابع، در هدف مدل ظاهر می‌شود. در این مدل از نرُم‌های $s = 2$ ، $p = 1$ و $q = 1$ استفاده شده است.

تابع هدف اول، همانند مدل کلاسیک، در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $\rho = 1$ به صورت ثابت در نظر گرفته می‌شود. با روش وزن دهی، تابع هدف دوم با ضریب وزنی $W_\alpha > 0$ و تابع هدف سوم با ضریب وزنی $W_\beta > 0$ به تابع هدف اول اضافه شده است.

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2 + W_\alpha \sum_{i=1}^n \alpha_i - W_\beta \sum_{i=1}^n \beta_i$$

$$S.T.:$$

$$(21)$$

$$y_i(A_i \cdot X - b) = 1 - \alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۲. زیرمدل‌های مبتنی بر برنامه‌ریزی ریاضی

چندهدفه

مدل‌های مبتنی بر برنامه‌ریزی ریاضی چندهدفه، اغلب با ترکیبی از توابع دوم و سوم ساخته می‌شوند.

$$y_i(A_i \cdot X - b) \geq 1 - \alpha_i$$

$$\alpha_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۱-۵. مدل v-SVC - نوع جدید طبقه‌بندی‌کننده بردار

پشتیبان

این مدل اولین بار در سال ۲۰۰۰ توسط Schölkopf و همکاران ارائه شد [۱۵]. در این مدل همانند مدل C-SVC از نرُم‌های $s = 2$ و $p = 1$ استفاده شده است. همانند مدل‌های قبلی، تابع هدف اول در قالب ماکزیمم کردن حاشیه مثبت ظاهر می‌گردد اما تفاوت این مدل با مدل‌های قبلی در ثابت فرض نکردن ρ می‌باشد. در این مدل، همزمان با مینیمم کردن $\frac{1}{2} \|X\|_2^2$ ، مقدار ρ نیز ماکزیمم می‌شود یا به عبارت دیگر، مقدار $-\rho$ مینیمم می‌شود و با ضریب وزنی W_p به تابع هدف اضافه می‌گردد. همچنین تابع هدف دوم نیز با ضریب وزنی به شکل $W_\alpha = \frac{1}{n}$ به تابع هدف اول اضافه شده است. (ن: تعداد رکوردها).

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2 - W_p \rho + W_\alpha \sum_{i=1}^n \alpha_i$$

$$S.T.:$$

$$(19)$$

$$W_\alpha = \frac{1}{n}$$

$$y_i(A_i \cdot X - b) \geq \rho - \alpha_i$$

$$\rho \geq 0$$

$$\alpha_i \geq 0$$

$$1 \leq i \leq n$$

۴-۱-۱-۶. مدل Ev-SVC - نوع جدید توسعه یافته

طبقه‌بندی‌کننده بردار پشتیبان

این مدل اولین بار در سال ۲۰۰۳ توسط Perez-Cruz و همکاران ارائه شد [۲۶]. در این مدل همانند مدل‌های C-SVC و v-SVC از نرُم‌های $s = 2$ و $p = 1$ استفاده شده است. بر خلاف مدل‌های مورد بررسی، در این مدل تابع هدف اول در قالب مینیمم کردن حاشیه منفی ظاهر می‌گردد.

در این مدل، $\|X\|_2^2 = 2$ به صورت ثابت در نظر گرفته می‌شود و مقدار $-\rho$ با ضریب وزنی W_p در تابع هدف، مینیمم می‌گردد. همچنین، تابع هدف دوم با روش وزن دهی و با ضریب وزنی به شکل $W_\alpha = \frac{1}{n}$ به تابع هدف اول اضافه شده است. (ن: تعداد رکوردها)

¹ MultiObjective Programming

در این مدل از نُرم‌های $p = 1$ و $q = 2$ استفاده شده است که منجر به ایجاد یک مدل بهینه‌سازی درجه دوم مقعر شده است. همانند مدل‌های قبلی، تابع دوم با ضریب وزنی $W_\alpha > 0$ و تابع سوم با ضریب وزنی $W_\beta > 0$ در تابع هدف مدل ظاهر شده‌اند.

$$\text{Minimize: } W_\alpha \sum_{i=1}^n \alpha_i - W_\beta \sum_{i=1}^n \beta_i^2$$

$$\text{S.T.:} \quad (24)$$

$$y_i(A_i \cdot X - b) = -\alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۲-۱-۴. مدل MCIQP - مدل برنامه‌ریزی درجه دوم نامعین چند هدفه^۴

این مدل اولین بار در سال ۲۰۰۷ توسط Shi و Zhang ارائه شد [۳۰]. تفاوت این مدل با مدل خطی در نُرم توابع هدف دوم و سوم می‌باشد. در این مدل از نُرم‌های $p = 2$ و $q = 2$ استفاده شده است که منجر به ایجاد یک مدل بهینه‌سازی درجه دوم نامعین شده است. همانند مدل‌های قبلی، تابع دوم با ضریب وزنی $W_\alpha > 0$ و تابع سوم با ضریب وزنی $W_\beta > 0$ در تابع هدف مدل ظاهر شده‌اند.

$$\text{Minimize: } W_\alpha \sum_{i=1}^n \alpha_i^2 - W_\beta \sum_{i=1}^n \beta_i^2$$

$$\text{S.T.:} \quad (25)$$

$$y_i(A_i \cdot X - b) = -\alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۲-۱-۵. مدل MCMP - مدل برنامه‌ریزی ریاضی چند هدفه^۵

این مدل اولین بار در سال ۲۰۰۶ توسط Kou و همکاران ارائه شد [۲۹]. به منظور توسعه مدل‌های مبتنی بر برنامه‌ریزی ریاضی چند هدفه، تابع هدف اول، معرفی و به توابع دوم و سوم اضافه گردیده است. تابع هدف اول، در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $\rho = 1$ به صورت ثابت در نظر گرفته می‌شود و $\frac{1}{2} \|X\|_2^2$ مینیمم می‌گردد.

۴-۲-۱-۱. مدل MCLP - مدل برنامه‌ریزی خطی چند هدفه^۱

این مدل اولین بار در قالب مدل LP در سال ۱۹۸۱ توسط Freed و Glover ارائه [۹] و سپس در سال ۲۰۰۱ توسط Shi و همکاران در قالب مدل MCLP توسعه داده شد [۲۱]. تابع دوم با ضریب وزنی $W_\alpha > 0$ و تابع سوم با ضریب وزنی $W_\beta > 0$ در تابع هدف مدل ظاهر شده‌اند. در این مدل از نُرم‌های $p = 1$ و $q = 1$ استفاده شده است. به صورت مینیمم نمودن منفی آن تابع، در هدف مدل ظاهر می‌شود.

$$\text{Minimize: } W_\alpha \sum_{i=1}^n \alpha_i - W_\beta \sum_{i=1}^n \beta_i$$

$$\text{S.T.:} \quad (22)$$

$$y_i(A_i \cdot X - b) = -\alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۲-۱-۲. مدل MCVQP - مدل برنامه‌ریزی درجه دوم محدب چند هدفه^۲

این مدل اولین بار در سال ۲۰۰۷ توسط Shi و Zhang ارائه شد [۳۰]. تفاوت این مدل با مدل خطی در نُرم تابع هدف دوم می‌باشد. در این مدل از نُرم‌های $p = 2$ و $q = 1$ استفاده شده است که منجر به ایجاد یک مدل بهینه‌سازی درجه دوم محدب شده است. همانند مدل خطی، تابع دوم با ضریب وزنی $W_\alpha > 0$ و تابع سوم با ضریب وزنی $W_\beta > 0$ در تابع هدف مدل ظاهر شده‌اند.

$$\text{Minimize: } W_\alpha \sum_{i=1}^n \alpha_i^2 - W_\beta \sum_{i=1}^n \beta_i$$

$$\text{S.T.:} \quad (23)$$

$$y_i(A_i \cdot X - b) = -\alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۲-۱-۳. مدل MCCQP - مدل برنامه‌ریزی درجه دوم مقعر چند هدفه^۳

این مدل اولین بار در سال ۲۰۰۷ توسط Shi و Zhang ارائه شد [۳۰]. تفاوت این مدل با مدل خطی در نُرم تابع هدف سوم می‌باشد.

¹ Multiple Criteria Linear Programming

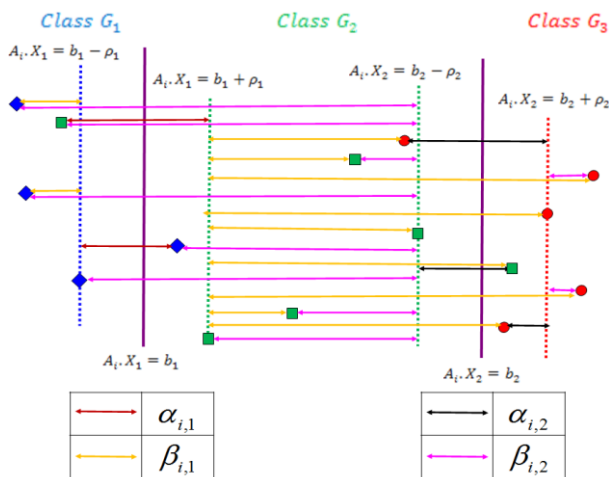
² Multiple Criteria Convex Quadratic Programming

³ Multiple Criteria Concave Quadratic Programming

⁴ Multiple Criteria Indefinite Quadratic Programming

⁵ Multiple Criteria Mathematical Programming

برنامه‌ریزی ریاضی چند هدفه نیز به منظور حل این گونه از مسائل توسعه یافته‌اند.



شکل ۳. نمایش هندسی مدل جامع طبقه‌بندی در حالت سه کلاسی

مدل‌های چند کلاسی مبتنی بر برنامه‌ریزی ریاضی چند هدفه که در ادامه مورد بررسی قرار می‌گیرند، حالت خاصی از مدل جامع ارائه شده می‌باشند.

برخلاف مدل‌های دو کلاسی، اندیس l در این مدل‌ها فعال می‌گردد. در این مسائل $m + 1$ کلاس (G_1, \dots, G_{m+1}) و در نتیجه m طبقه‌بندی‌کننده (ابرفصله جداکننده) وجود دارد.

در دو مدل اول مورد بررسی فرض بر این است که برای کلیه طبقه‌بندی‌کننده‌ها یک بردار ضرایب وجود دارد $(X_l = X)$ که بر کلیه ابرصفحات جداکننده عمود می‌باشد. بر همین اساس، همچنان $A_i \cdot X$ به عنوان امتیاز طبقه‌بندی i امین رکورد داده‌ها نسبت به ابرصفحات جداکننده در نظر گرفته می‌شود که ترکیب آن با مقادیر مختلف بایاس b_i ، توابع تصمیم را می‌سازد. ضمناً بر اساس تعریف قبلی، برچسب کلاس A_i نسبت به ابرصفحات جداکننده به صورت زیر تعریف می‌شود:

$$\forall A_i \in G_1 \cup \dots \cup G_l \quad A_i \cdot X \leq b_i \quad y_{i,l} = -1 \quad (27)$$

$$\forall A_i \in G_{l+1} \cup \dots \cup G_{m+1} \quad A_i \cdot X > b_i \quad y_{i,l} = +1$$

این مدل اولین بار در سال ۲۰۰۵ توسط Shi و همکاران ارائه شد [۲۸]. تابع دوم با ضریب وزنی $W_\alpha > 0$ و تابع سوم با ضریب وزنی $W_\beta > 0$ در تابع هدف مدل ظاهر شده‌اند. در این مدل از نرم‌های $p = 2$ ، $q = 2$ و $t = 2$ استفاده شده است.

در این مدل از نرم‌های $s = 2$ ، $p = 2$ و $q = 1$ استفاده شده است که منجر به ایجاد یک مدل بهینه‌سازی درجه دوم شده است. با روش وزن دهی، تابع هدف دوم با ضریب وزنی $W_\alpha > 0$ و تابع هدف سوم با ضریب وزنی $W_\beta > 0$ به تابع هدف اول اضافه شده است.

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2 + W_\alpha \sum_{i=1}^n \alpha_i^2 - W_\beta \sum_{i=1}^n \beta_i \quad (26)$$

S.T.:

$$y_i(A_i \cdot X - b) = 1 - \alpha_i + \beta_i$$

$$\alpha_i, \beta_i \geq 0$$

$$1 \leq i \leq n$$

۴-۲. زیرمدل‌های چند کلاسی

عمومی‌ترین روش‌های طبقه‌بندی مبتنی بر بهینه‌سازی، در بر گیرنده مسائل دو کلاسی بوده که به ایجاد برچسب‌های -1 و $+1$ می‌انجامد. این در حالی است که غالباً در محیط‌های واقعی، طبقه‌بندی بر روی چند کلاس صورت می‌گیرد که با توسعه مدل‌های دو کلاسی، به چند کلاسی قابل حل می‌باشند.

یکی از قدیمی‌ترین و ساده‌ترین روش‌های پیاده‌سازی طبقه‌بندی چند کلاسی، روش one against all می‌باشد که برای هر کلاس، به تعداد کلاس‌ها $(m + 1)$ طبقه‌بندی‌کننده دو کلاسی ایجاد می‌کند.

طبقه‌بندی‌کننده k ام، کلاس k ام را از سایر کلاس‌ها تفکیک می‌کند و با ادامه این روند $(k = 1, \dots, m + 1)$ ، طبقه‌بندی سایر کلاس‌ها نیز صورت می‌گیرد.

از روش‌های دیگر طبقه‌بندی چند کلاسی، می‌توان به روش one against one اشاره نمود. در این روش برای $m + 1$ کلاس، $m(m + 1)/2$ طبقه‌بندی‌کننده دو کلاسی ایجاد می‌شود، به طوری که هر ترکیب دو کلاسی از داده‌ها مورد تفکیک قرار می‌گیرند. ترکیب خروجی این طبقه‌بندی‌کننده‌ها، کلاس هر رکورد را تعیین می‌کند.

روش پیشنهادی در مدل جامع، فقط نیازمند ایجاد m طبقه‌بندی‌کننده دو کلاسی به ازای $m + 1$ کلاس می‌باشد. در این روش، طبقه‌بندی‌کننده l ام، اجتماع کلاس‌های $1, \dots, l$ را از اجتماع کلاس‌های $l + 1, \dots, m + 1$ تفکیک می‌نماید. این در حالی است که کلیه توابع تصمیم با حل تنها یک مسئله بهینه‌سازی چندهدفه حاصل می‌شوند.

همانطور که حل مسائل طبقه‌بندی چند کلاسی در مدل‌های مبتنی بر SVM مورد توجه قرار گرفته است، مدل‌های مبتنی بر

$$\text{Minimize: } \frac{1}{2} \|X\|_2^2 + W_\alpha \sum_{i=1}^n \sum_{l=1}^m \alpha_{i,l}^2 - W_\beta \left(\sum_{i=1}^n \sum_{l=1,m} \beta_{i,l} - \sum_{i=1}^n \sum_{l=2}^m (\beta_{i,l} - \frac{b_l - b_{l-1}}{2})^2 \right)$$

S.T. : (۲۹)

$$y_{i,l}(A_i \cdot X - b_l) = 1 - \alpha_{i,l} + \beta_{i,l}$$

$$\alpha_{i,l}, \beta_{i,l} \geq 0$$

$$1 \leq i \leq n$$

$$1 \leq l \leq m$$

۳-۲-۴. مدل پیشنهادی MC-MO-CQP - مدل

برنامه‌ریزی درجه دوم محدب چند هدفه چند کلاسی^۱

بر خلاف مدل‌های MC-MCQP و MC-MCMP، با فعال شدن اندیس l ، برای هر طبقه‌بندی‌کننده، یک بردار ضرایب وجود دارد (X_l) که بر ابرصفحه جداکننده l ام عمود می‌باشد. بر این اساس، $A_i \cdot X_l$ به عنوان امتیاز طبقه بندی i امین رکورد داده‌ها نسبت به ابرصفحه جداکننده l ام در نظر گرفته می‌شود که ترکیب آن با b_l ، تابع تصمیم را می‌سازد. به ازای $1 \leq i \leq n$ و $1 \leq l \leq m$ ، برچسب کلاس A_i نسبت به ابرصفحات جداکننده به صورت زیر تعریف می‌شود:

$$\forall A_i \in G_1 \cup \dots \cup G_l \quad A_i \cdot X_l \leq b_l \quad y_{i,l} = -1$$

(۳۰)

$$\forall A_i \in G_{l+1} \cup \dots \cup G_{m+1} \quad A_i \cdot X_l > b_l \quad y_{i,l} = +1$$

در مدل پیشنهادی، تابع هدف اول در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $\rho_l = 1$ به صورت ثابت در نظر گرفته می‌شود و $\frac{1}{2} \|X\|_s^s$ با نُرم $s = 2$ مینیمم می‌گردد. در تابع هدف دوم، مقدار تجمیعی خطای طبقه‌بندی نادرست را با استفاده از $\|\alpha\|$ با نُرم $p = 2$ مینیمم نموده و در قسمت اول تابع هدف سوم، مقدار تجمیعی موفقیت در طبقه‌بندی صحیح را نسبت به ابرصفحات اول و آخر ($l = 1, m$)، با استفاده از $\|\beta\|$ با نُرم $q = 1$ ماکزیمم می‌کنیم.

تابع هدف سوم که در قالب ماکزیمم سازی معرفی شده است، به صورت مینیمم نمودن منفی آن تابع، در هدف مدل ظاهر می‌شود. در قسمت اول این تابع، فواصل نسبت به ابرصفحات اول و آخر محاسبه می‌گردد.

قسمت دوم تابع هدف سوم، نقش یک محدودیت، جهت کنترل تفکیک پذیری مناسب گروه‌های مجاور را ایفا می‌کند. این نقش با مینیمم کردن اختلاف $\beta_{i,l}$ از وسط دو ابرصفحه جداکننده مجاور ($f_l = \frac{b_l - b_{l-1}}{2}$) اجرا می‌گردد.

$$\text{Minimize: } W_\alpha \sum_{i=1}^n \sum_{l=1}^m \alpha_{i,l}^2 - W_\beta \left(\sum_{i=1}^n \sum_{l=1,m} \beta_{i,l}^2 - \sum_{i=1}^n \sum_{l=2}^m (\beta_{i,l} - \frac{b_l - b_{l-1}}{2})^2 \right)$$

S.T. : (۲۸)

$$y_{i,l}(A_i \cdot X - b_l) = -\alpha_{i,l} + \beta_{i,l}$$

$$\alpha_{i,l}, \beta_{i,l} \geq 0$$

$$1 \leq i \leq n$$

$$1 \leq l \leq m$$

این مدل اولین بار در سال ۲۰۰۹ توسط Kou و همکاران ارائه شد [۳۱]. تفاوت اصلی این مدل با مدل MC-MCQP در تعریف و اضافه شدن تابع هدف اول به توابع دوم و سوم می‌باشد. تابع هدف اول، در قالب ماکزیمم کردن حاشیه مثبت ظاهر شده و $\rho_l = 1$ به صورت ثابت در نظر گرفته می‌شود و $\frac{1}{2} \|X\|_s^s$ مینیمم می‌گردد.

ضمناً نُرم قسمت اول تابع هدف سوم نیز نسبت به مدل قبلی تغییر کرده است. در این مدل از نُرم‌های $s = 2$ ، $p = 2$ ، $q = 1$ و $t = 2$ استفاده شده است که منجر به ایجاد یک مدل بهینه‌سازی درجه دوم محدب شده است.

با روش وزن دهی، تابع هدف دوم با ضریب وزنی $W_\alpha > 0$ و تابع هدف سوم با ضریب وزنی $W_\beta > 0$ به تابع هدف اول اضافه شده است.

همانند مدل قبلی، قسمت دوم تابع هدف سوم، با مینیمم کردن اختلاف $\beta_{i,l}$ از وسط دو ابرصفحه جداکننده مجاور، تفکیک پذیری مناسب گروه‌های مجاور را کنترل می‌کند.

¹ Multi Class Multiple Objective Convex Quadratic Programming

طبقه‌بندی روی مجموعه آزمایشی) می‌باشد. با پیاده‌سازی مدل پیشنهادی روی مثال عددی، با استفاده از کدنویسی در محیط نرم‌افزار Lingo11 نتایج زیر حاصل گردیده است:

جدول ۱. نتایج دقت طبقه‌بندی مدل پیشنهادی (روی مثال عددی)

مجموعه آزمایشی	مجموعه آموزشی	دقت طبقه بندی (%)
۹۷,۶	۹۹,۳	کلی
۹۵,۸	۹۹,۲	کلاس L
۱۰۰,۰	۱۰۰,۰	کلاس B
۹۹,۰	۹۹,۲	کلاس R

در این مقاله، پنج مورد از معروف‌ترین الگوریتم‌های طبقه‌بندی در داده‌کاوی انتخاب شده تا از آنها به منظور مقایسه با مدل پیشنهادی استفاده گردد:

- نسخه ۸ درخت تصمیم C4.5 (J 4.8)
- مدل احتمالی استاندارد نیو بیز (NB)
- مدل استاندارد نزدیک‌ترین k همسایه (KNN)
- شبکه عصبی به روش انتشار بازگشتی^۳ (NN)
- ماشین بردار پشتیبان به روش بهینه‌سازی حداقل متوالی^۴ (SMO-SVC)

پس از پیاده‌سازی هر یک از الگوریتم‌های فوق در محیط نرم افزار Weka، دقت طبقه‌بندی کل و هر کلاس را روی مجموعه آزمایشی مثال فوق، محاسبه کرده و نتایج را در جدول ۲ نشان می‌دهیم.

جدول ۲. مقایسه نتایج دقت پیش‌بینی الگوریتم‌های مختلف طبقه‌بندی (مجموعه آزمایشی مثال عددی)

طبقه‌بندی کننده	دقت طبقه‌بندی هر کلاس (%)		
	R	B	L
J 4.8	٪۸۹,۶	٪۱۲,۵	٪۸۳,۳
NB	٪۱۰۰,۰	٪۰,۰	٪۹۹,۰
KNN	٪۹۹,۰	٪۶,۳	٪۹۳,۸
NN	٪۹۲,۷	٪۶۲,۵	٪۹۹,۰
SMO-SVC	٪۹۹,۰	٪۸۱,۳	٪۹۷,۹
MC-MO-CQP	٪۹۹,۰	٪۱۰۰,۰	٪۹۵,۸

³ Backpropagation

⁴ Sequential Minimal Optimization

قسمت دوم تابع هدف سوم، با مینیمم کردن اختلاف $\beta_{i,l}$ از $\beta_{i,l-1}$ ، تفکیک‌پذیری مناسب گروه‌های مجاور را کنترل می‌کند ($f_l = \beta_{i,l} - \beta_{i,l-1}$) بطوریکه $l \geq 2$ و به این منظور از نرم $t = 2$ استفاده می‌گردد. با روش وزن دهی، تابع هدف دوم با ضریب وزنی $W_\alpha > 0$ و تابع هدف سوم با ضریب وزنی $W_\beta > 0$ به تابع هدف اول اضافه شده است. تابع هدف سوم در قالب مینیمم نمودن منفی آن، در هدف مدل ظاهر می‌شود.

$$\text{Minimize: } \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^r x_{j,i}^2 + W_\alpha \sum_{i=1}^n \sum_{l=1}^m \alpha_{i,l}^2 - W_\beta \left(\sum_{i=1}^n \sum_{l=1,m} \beta_{i,l} - \sum_{i=1}^n \sum_{l=2}^m (\beta_{i,l} - \beta_{i,l-1})^2 \right)$$

S.T. : (۳۱)

$$y_{i,l}(A_{i,l} X_l - b_l) = 1 - \alpha_{i,l} + \beta_{i,l}$$

$$\alpha_{i,l}, \beta_{i,l} \geq 0$$

$$1 \leq i \leq n$$

$$1 \leq l \leq m$$

با توجه به نرم‌های مورد استفاده، این مدل به یک مسئله برنامه‌ریزی غیرخطی مقید چندهدفه (مدل بهینه‌سازی درجه دوم محدب) تبدیل می‌گردد که پیدا کردن پاسخ بهینه سراسری آن با استفاده از روال‌های شناخته‌شده و نرم‌افزارهای موجود امکان‌پذیر است. این مدل، قادر به حل مسائل طبقه‌بندی چندکلاسی و در حالت خاص آن، مسائل دوکلاسی می‌باشد.

۵. پیاده‌سازی مدل پیشنهادی روی مثال عددی و

مقایسه با سایر الگوریتم‌ها

در این قسمت، مدل پیشنهادی بر روی مجموعه داده‌های استاندارد پیاده می‌شود. به این منظور، مجموعه داده‌های وزن و فاصله مقیاس تعادلی^۱، از مخزن یادگیری ماشینی دانشگاه UCI^۲ استخراج شده است [۳۶].

این داده‌ها به صورت از پیش طبقه‌بندی شده هستند و کلاس متناظر با هر رکورد مشخص شده است. در این مجموعه، ۶۲۵ نمونه موجود با ۴ مشخصه در سه کلاس (L,B,R) توزیع شده‌اند. معیار مورد استفاده در این مقاله که به منظور اعتبارسنجی مدل پیشنهادی مورد استفاده قرار می‌گیرد، دقت پیش‌بینی (دقت

¹ Balance Scale Weight & Distance Database

² University of California, Irvine Machine Learning Repository

در ادامه، با استفاده از ترکیب تجربیات حاصل از توسعه روش‌های مبتنی بر ماشین بردار پشتیبان و روش‌های مبتنی بر برنامه‌ریزی ریاضی چند هدفه، یک زیرمدل پیشنهادی جدید (MC-MO-CQP) بر پایه مدل جامع و به عنوان حالت خاص جدیدی از آن ایجاد کردیم. این زیرمدل که یک مسئله برنامه‌ریزی درجه دوم محدب مقید چندهدفه را تشکیل می‌دهد، قادر به حل مسائل طبقه‌بندی چندکلاسی می‌باشد. از جمله ویژگی‌های این زیرمدل، حصول پاسخ بهینه سراسری می‌باشد. در پایان، مدل پیشنهادی و چند الگوریتم متداول طبقه‌بندی، روی مجموعه داده‌های معتبر نمونه پیاده گردیدند تا میزان کارایی مدل، با استفاده از بانک داده‌های نمونه در مقایسه با نتایج چند الگوریتم دیگر طبقه بندی مورد ارزیابی و اعتبارسنجی قرار گیرد. معیار مورد استفاده در این مقاله که به منظور مقایسه روش‌های اشاره شده مورد استفاده قرار گرفت، دقت پیش‌بینی بوده است. نتایج پیاده‌سازی الگوریتم و محاسبه دقت طبقه‌بندی کل و هر کلاس روی مجموعه آزمایشی مثال عددی اثبات می‌کند که مدل پیشنهادی، در مقایسه با سایر متدها، تفکیک مناسبی از وضعیت داده‌های جدید حاصل می‌کند. از جمله مزایای مدل طبقه‌بندی پیشنهادی می‌توان به موارد زیر اشاره کرد:

- حصول دقت بالای پیش‌بینی برچسب کلاس داده‌های جدید
- حصول پاسخ بهینه سراسری از حل مدل بهینه‌سازی
- استقرار الگوریتم بر پایه اصول نظری منسجم و برخورداری از ویژگی‌های هندسی
- قابلیت طبقه‌بندی مسائل چند کلاسی

با توجه متداول شدن روش‌های مبتنی بر بهینه‌سازی و به‌کارگیری آنها در بیشتر ابزارهای داده‌کاوی، در این مقاله با برقراری ارتباط بین مدل‌های مختلف طبقه‌بندی مبتنی بر بهینه‌سازی، مسیر هموارتری به منظور گسترش مدل‌های جدید طبقه‌بندی گشوده شد.

با توجه به مطالعات صورت گرفته، امکان ایجاد زیرمدل‌های کارتر بر مبنای مدل جامع طبقه‌بندی فراهم گردیده است. به عنوان پیشنهاداتی برای تحقیقات آتی می‌توان به مطالعه بر روی نحوه انتخاب موثر ضرایب وزنی توابع هدف مدل و توسعه زیرمدل پیشنهادی به نحوی که قادر به حل مسائل طبقه‌بندی مجموعه داده‌های با پیچیدگی غیرخطی نیز باشد، اشاره کرد. در نهایت، لازم به ذکر است که پیاده‌سازی مدل‌های مختلف روی مسائل متفاوت دنیای واقعی، قابلیت تفکیک‌پذیری داده‌ها را توسط این الگوریتم‌ها، در حوزه‌های مختلف مشخص می‌کند.

همان‌طور که در جدول ۲ مشخص می‌باشد، پیاده‌سازی مدل طبقه‌بندی‌کننده پیشنهادی روی مثال عددی، نسبت به سایر الگوریتم‌های مورد بررسی، دقت کلی بالاتری را ارائه نموده است. الگوریتم SMO-SVC که از خانواده روش‌های مبتنی بر بهینه‌سازی می‌باشد، پس از الگوریتم پیشنهادی، بالاترین دقت را داشته است. تنها الگوریتم NN از نظر دقت پیش‌بینی با روش‌های مبتنی بر بهینه‌سازی قابل رقابت می‌باشد.

۶. نتیجه‌گیری و پیشنهاد تحقیقات آتی

در این مقاله به بررسی مسائل طبقه‌بندی داده‌ها در قالب مدل برنامه‌ریزی ریاضی چندهدفه پرداختیم. در قالب یک دسته‌بندی کلی، تاکنون دو دسته از روش‌های طبقه‌بندی مبتنی بر بهینه‌سازی مورد مطالعه قرار گرفته‌اند: دسته اول، مطالعات اولیه‌ای است که در دهه ۶۰ توسط Vapnik و همکارانش و همچنین Mangasarian و همکارانش صورت گرفت و سپس در دهه ۹۰ در قالب ماشین بردار پشتیبان (SVM) به بلوغ رسید و سپس به مرور به تکامل دست یافت. دسته دوم شامل مطالعاتی است که اولین بار توسط Glover و Freed در دهه ۸۰ ارائه شد و از سال ۲۰۰۰ توسط Shi و همکارانش در قالب روش‌های برنامه‌ریزی ریاضی چندهدفه خطی و غیرخطی بسط داده شد. از زمان ارائه اولین مدل‌های برنامه‌ریزی برای مسائل طبقه‌بندی، افراد زیادی به صورت مستقل و همزمان روی گسترش مدل‌های طبقه‌بندی مبتنی بر روش‌های برنامه‌ریزی ریاضی و روش‌های ماشین بردار پشتیبان مطالعه نموده‌اند. در این میان، نکته قابل توجه، فلسفه یکسان هر دو دسته از روش‌های طبقه‌بندی مبتنی بر بهینه‌سازی می‌باشد. در این مقاله به منظور پر کردن شکاف بین این دو مسیر پژوهش، مدل جامعی به منظور طبقه‌بندی در داده‌کاوی مبتنی بر برنامه‌ریزی چندهدفه ارائه گردید که مدل‌های اولیه و بهبود یافته SVM و همچنین مدل‌های برنامه‌ریزی ریاضی یک و چندهدفه ارائه شده، حالات خاصی از این مدل جامع پیشنهادی باشند. ارائه این مدل جامع برنامه‌ریزی چندهدفه (MOP)، مطالعاتی که به صورت مستقل، در راستای بهبود هر کدام از این روش‌ها صورت گرفته است را قابل تعمیم به روش‌های دیگر گردید. غالباً در محیط‌های واقعی، طبقه‌بندی بر روی چند کلاس صورت می‌گیرد که با توسعه مدل‌های دو کلاسی، به چند کلاسی قابل حل می‌باشند. در روش پیشنهادی در مدل جامع، با استفاده از رویکرد حل یکجا و همزمان مسائل طبقه‌بندی چند گروهی، کلیه توابع تصمیم با حل تنها یک مسئله بهینه‌سازی چندهدفه حاصل می‌شوند. تبدیل مدل جامع پیشنهادی به چندین زیرمدل دو یا چند کلاسی مبتنی بر ماشین بردار پشتیبان و مبتنی بر برنامه‌ریزی ریاضی چند هدفه مورد بررسی قرار گرفت.

منابع

- [16] Lee, Y., Mangasarian, O., "RSVM: Reduced Support Vector Machines", Presented at SIAM International Conference on Data Mining, Chicago, USA, 2001.
- [17] Lee, Y.J., Mangasarian, O.L., "SSVM: A Smooth Support Vector Machine for Classification", Computational Optimization and Applications, Vol. 20, No. 1, 2001, pp. 5-22.
- [18] Mangasarian, O., Musicant, D., "Lagrangian Support Vector Machine", Journal of Machine Learning Research, Vol. 1, 2001, pp. 161-177.
- [19] Mangasarian, O., Musicant, D., "Active Support Vector Machine Classification", Advances in Neural Information Processing Systems, Vol. 13, 2001, pp. 577-583.
- [20] Mangasarian, O., Wild, E., "Proximal Support Vector Machine Classifiers", Presented at International Conference on Knowledge Discovery & Data Mining (KDD), San Francisco, USA, 2001.
- [21] Shi, Y., et al., "Data Mining in Credit Card Portfolio Management: a Multiple Criteria Decision Making Approach", Presented at International Conference on Multiple Criteria Decision Making (MCDM), Ankara, Turkey, 2001.
- [22] Bugera, V., Konno, H., Uryasev, S., "Credit Cards Scoring with Quadratic Utility Functions", Journal of Multi-Criteria Decision Analysis, Vol. 11, No. 4-5, 2002, pp. 197-211.
- [23] Peng, Y., Shi, Y., Xu, W., "Classification for Three-Group of Credit Cardholders' Behavior Via a Multiple Criteria Approach", Advanced Modeling and Optimization, Vol. 4, 2002, pp. 39-56.
- [24] Hsu, C., Chang, C., Lin, C., "A Practical Guide to Support Vector Classification", Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, 2003.
- [25] Bi, J., "Multi-Objective Programming in SVMs", Presented at International Conference on Machine Learning (ICML), Washington DC, USA, 2003.
- [26] Perez-Cruz, F., et al., "Extension of the nu-SVM Range for Classification", Nato Science Series Sub Series, Computer And Systems Sciences, Vol. 190, 2003, pp. 179-196.
- [27] Asada, T., et al., "Pattern Classification by Goal Programming and Support Vector Machines", Computational Management Science, Vol. 1, No. 3, 2004, pp. 211-230.
- [28] Shi, Y., et al., "Classifying Credit Card Accounts for Business Intelligence and Decision Making: A
- [1] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data", Communications of the ACM, Vol. 39, No. 11, 1996, pp. 27-34.
- [2] Han, J., Kamber, M., *Data Mining: Concepts and Techniques* (2 ed.), Morgan Kaufmann, 2006.
- [3] Witten, I., Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques* (2 ed.), Morgan Kaufmann Pub, 2005.
- [4] Olson, D., Delen, D., *Advanced Data Mining Techniques*, Springer Verlag, 2008.
- [5] Zhang, J., Shi, Y., Zhang, P., "Several Multi-Criteria Programming Methods for Classification", Computers & Operations Research, Vol. 36, No. 3, 2009, pp. 823-836.
- [6] Fisher, R., "The use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, Vol. 7, No. 2, 1936, pp. 179-188.
- [7] Vapnik, V., Lerner, A., "Pattern Recognition using Generalized Portrait Method", Automation and Remote Control, Vol. 24, No. 6, 1963, pp. 774-780.
- [8] Mangasarian, O., "Linear and Nonlinear Separation of Patterns by Linear Programming", Operations Research, Vol. 13, No. 3, 1965, pp. 444-452.
- [9] Freed, N., Glover, F., "A Linear Programming Approach to the Discriminant Problem", Decision Sciences, Vol. 12, No. 1, 1981, pp. 68-79.
- [10] Boser, B., Guyon, I., Vapnik, V., "A Training Algorithm for Optimal Margin Classifiers", Presented at Annual Conference on Learning Theory (COLT), New York, USA, 1992.
- [11] Cortes, C., Vapnik, V., "Support-Vector Networks", Machine Learning, Vol. 20, No. 3, 1995, pp. 273-297.
- [12] Osuna, E., Freund, R., Girosi, F., "Support Vector Machines: Training and Applications", Massachusetts Institute of Technology, 1997.
- [13] Mangasarian, O., "Generalized Support Vector Machines", Advances in Neural Information Processing Systems, Vol. 10, 1999, pp. 135-146.
- [14] Suykens, J., Vandewalle, J., "Least Squares Support Vector Machine Classifiers", Neural processing letters, Vol. 9, No. 3, 1999, pp. 293-300.
- [15] Schölkopf, B., et al., "New Support Vector Algorithms", Neural Computation, Vol. 12, No. 5, 2000, pp. 1207-1245.

Multiple-Criteria Quadratic Programming Approach", International Journal of Information Technology & Decision Making (IJITDM), Vol. 4, No. 04, 2005, pp. 581-599.

- [29] Kou, G., et al., "A New Multi-Criteria Convex Quadratic Programming model for Credit Analysis", Presented at International Conference on Computational Science (ICCS), Reading, UK, 2006.
- [30] Zhang, J., Shi, Y., "A Framework of Optimization Method for Classification", Presented at International Symposium on Optimization and Systems Biology (OSB), Beijing, China, 2007.
- [31] Kou, G., et al., "Multiple Criteria Mathematical Programming for Multi-Class Classification and Application in Network Intrusion Detection", Information Sciences, Vol. 179, No. 4, 2009, pp. 371-381.
- [32] Ben-Hur, A., Weston, J., "A User's Guide to Support Vector Machines", Methods in Molecular Biology, Vol. 609, 2010, pp. 223-239.
- [33] Zhang, D., Tian, Y., Shi, Y., "A Group of Knowledge-Incorporated Multiple Criteria Linear Programming Classifiers", Journal of Computational and Applied Mathematics, In Press, 2011.
- [34] Vapnik, V., Chervonenkis, A., *On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities*", Theory of Probability and its Applications, Vol. 16, No. 2, 1971, pp. 264-280.
- [35] Freed, N., Glover, F., "Simple But Powerful Goal Programming Models for Discriminant Problems", European Journal of Operational Research, Vol. 7, No. 1, 1981, pp. 44-60.
- [36] *Machine Learning Repository, Center for Machine Learning and Intelligent Systems*, University of California, Irvine, Available from: <http://archive.ics.uci.edu/ml/index.html>.

