طبقه‌بندی بااحتیاط داده‌های ابرمستطیلی، ابردایروی و ابربیضوی با حداکثر حاشیه متقارن نسبت به لبه داده‌ها

یحیی فرقانی، میثاق سادات حجازی و هادی صدوقی یزدی

چکیده: مدل طبقه‌بندی مقاوم، یک مدل غیر استاندارد برای یادگیری طبقه‌بند بر اساس یک مجموعه داده توأم با عدم قطعیت است. به هر مدل طبقه‌بندی که در مجموعه جواب‌های ممکن آن، جواب بی‌معنی وجود داشته باشد، مدل بی‌احتیاط گفته می‌شود. جواب بهینه یک مدل طبقه‌بندی مقاوم بی‌احتیاط به ازای یک مجموعه داده آموزشی، ممکن است ابرصفحه نباشد که در این صورت امکان طبقه‌بندی داده‌ها در مرحله آزمون میسر نخواهد بود. در این مقاله مدل‌های طبقه‌بند مقاوم بی‌احتیاط معرفی و مشکلات آنها بررسی شده و سپس با تغییر تابع ضرر در طبقه‌بند مقاوم، مدل طبقه‌بندی مقاوم بااحتیاط برای ممانعت از بی‌احتیاطی معرفی می‌شود. مدل بااحتیاط پیشنهادی، استاندارد شده و راهکارهایی برای کاهش زمان آموزش و زمان آزمون آن ارائه می‌گردد. در آزمایشات از مدل طبقه‌بند مقاوم بااحتیاط پیشنهادی در مقایسه با چند مدل مقاوم بی‌احتیاط، برای طبقه‌بندی مجموعه داده‌های آموزشی ناقص و مجموعه داده‌های آموزشی قطعی کامل استفاده شد. نتایج به دست آمده نشان داد که در مجموعه داده‌های ناقص، مدل پیشنهادی زمان آموزش و زمان آزمون و نرخ خطای کمتری نسبت به مدل‌های بی‌احتیاط داشت. همچنین در مجموعه داده‌های کامل قطعی، مدل پیشنهادی زمان آموزش و زمان آزمون کمتری نسبت به مدل‌های بی‌احتیاط داشت. نتایج به دست آمده کارایی افزودن احتیاط به طبقه‌بند مقاوم را تأیید نمود.

کلیدواژه: **داده توأم با عدم قطعیت، زمان آزمون، زمان آموزش، طبقه‌بند مقاوم بااحتیاط.**

۱- مقدمه

داده‌ها در دنیای واقعی، توأم با نوعی عدم قطعیت هستند. عدم قطعیت در داده‌ها ممکن است ناشی از نویز باشد.

نویز در مراحل پیش‌پردازش، رقومی‌سازی، انتقال یا در نمایش داده‌ها به وجود می‌آید. مدل‌های اولیه طبقه‌بندی، غالباً عدم قطعیت در داده‌ها را نادیده می‌گیرند ولی از آنجا که عدم قطعیت در داده‌ها، خود نوعی دانش است که می‌تواند در کارایی طبقه‌بند مؤثر باشد، مدل‌های جدیدتر و پیشرفته‌تر طبقه‌بندی، از این دانش چشم‌پوشی نمی‌کنند. یکی از این روش‌ها مدل طبقه‌بندی مقاوم است که با عناوین دیگری چون طبقه‌بندی بر اساس بدترین شرایط، طبقه‌بندی مبتنی بر دانش و طبقه‌بندی نامتغیر نیز شناخته می‌شود.

این مقاله در تاریخ ۲ آبان ماه ۱۳۹۷ دریافت و در تاریخ ۹ خرداد ماه ۱۳۹۸ بازنگری شد.

یحیی فرقانی، گروه کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: yahyafor2000@yahoo.com).

میثاق سادات حجازی، گروه هوش مصنوعی، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: misagh.hejazi69@mail.um.ac.ir).

هادی صدوقی یزدی (نویسنده مسئول)، گروه هوش مصنوعی، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: h-sadoghi@um.ac.ir).

۱-۱ مدل‌های طبقه‌بندی استاندارد

مدل‌های محدبی که الگوریتم‌های سریع‌تر و پایاتری نسبت به سایر مدل‌های محدب برای آنها وجود دارد، مدل‌های استاندارد برنامه‌ریزی محدب یا به طور خلاصه مدل استاندارد نامیده می‌شوند [۱]. از جمله این مدل‌ها، مدل برنامه‌ریزی مخروط مرتبه دوم (SOCP)[1] است که بسط برنامه‌ریزی مربعی و برنامه‌ریزی خطی بوده و در آن ترکیب وابسته متغیرها به صورت قید داخل یک مخروط مرتبه دوم تعریف می‌شوند. این نوع برنامه‌ریزی در حل مسایل هندسی و همچنین در برنامه‌ریزی‌های خطی که داده‌ها با خطا همراه بوده و به طور دقیق مشخص نیستند، کاربرد دارد. خواص مدل‌های دسته SOCP عبارتند از: ۱) قابل تبدیل به فرم استاندارد برنامه‌ریزی درجه دوم، ۲) عدم نیمه‌معین مثبت‌بودن، ۳) دارای فضای محدب و ۴) کارایی حل با استفاده از الگوریتم‌های نقطه داخلی.

۱-۲ مدل طبقه‌بندی داده‌های قطعی

ماشین بردار پشتیبان SVM[2] یک مدل طبقه‌بندی داده‌های قطعی است که برای یادگیری طبقه‌بند بر اساس داده‌های قطعی مورد استفاده قرار می‌گیرد و شکل کلی آن در (۱) نشان داده شده است

$$\min_{w,b} f(w,b)+\mathrm{C}\sum_{i=1}^{n}\ell(w,b;\varphi(x_{(i)})) \qquad (1)$$

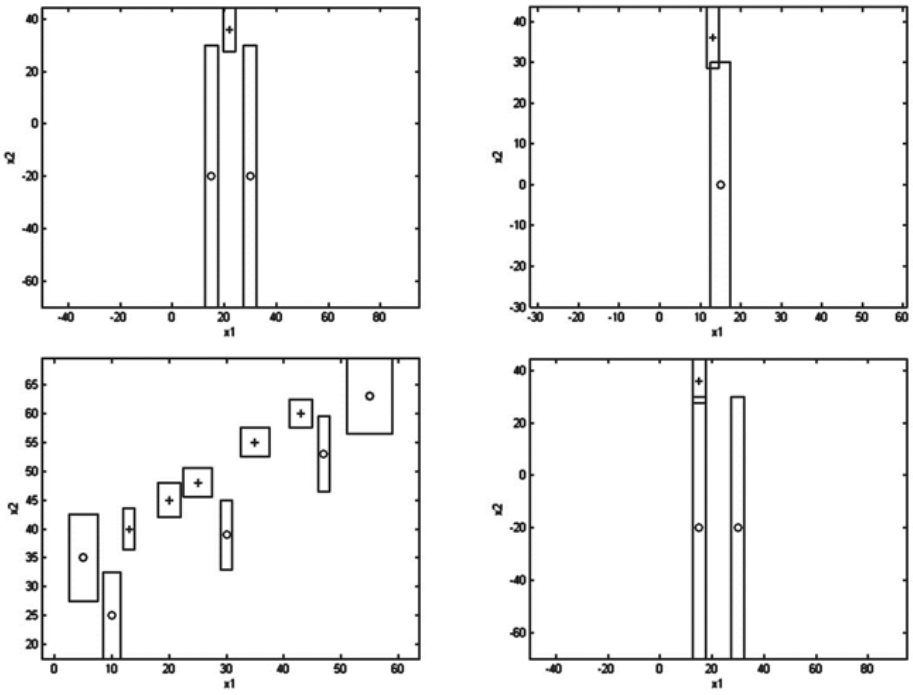
که w و b به ترتیب بردار وزن و بایاس طبقه‌بند، $f(w,b)$ ترم تنظیم[3]، $\ell(w,b;\varphi(x_{(i)}))$ تابع ضرر[4]، $\varphi(x_{(i)})$ تابع تبدیل از فضای ورودی به فضای ویژگی با بعد بالا[5] و C ضریب تأثیر جریمه[6] است.

۱-۳ مدل طبقه‌بندی مقاوم

تعریف ۱-۱: طبقه‌بند مقاوم، طبقه‌بندی است که بر اساس بدترین شرایط هر داده توأم با عدم قطعیت شکل گرفته است. اگر مجموعه داده‌های آموزشی، توأم با عدم قطعیت باشند یعنی $x_{(i)}\in u_{(i)}$ که $u_{(i)}$ یک مجموعه است، آن گاه نوع مقاوم مدل طبقه‌بندی SVM به صورت (۲) تعریف می‌شود

$$\min_{w,b}\ \max_{\substack{x_{(i)}\in u_{(i)}\\ i=1,2,\ldots,n}} f(w,b)+\mathrm{C}\sum_{i=1}^{n}\ell(w,b;\varphi(x_{(i)})) \qquad (2)$$

1. Second Order Cone Programming
2. Support Vector Machine
3. Regularization
4. Loss Function
5. High Dimensional Feature Space
6. Penalty



شکل ۱: چند مجموعه آموزشی چاله برای مدل RSVM بی‌احتیاط در فضای ورودی. هر مجموعه عدم قطعیت با یک مستطیل نشان داده شده است. مجموعه عدم قطعیت مربوط به طبقه ۱+ و ۱- به ترتیب با علامت + و o مشخص شده است.

این مدل، ماشین بردار پشتیبان مقاوم و به اختصار RSVM[1] نامیده می‌شود و مدلی است غیر استاندارد و در حقیقت نوعی مدل SVM فازی است. سه چالش اصلی پیش روی مدل‌های طبقه‌بندی مقاوم از جمله مدل RSVM وجود دارد که عبارتند از بالابودن زمان آموزش، بالابودن زمان آزمون (هر تلاش برای کاهش زمان آزمون، غالباً با افزایش زمان آموزش مدل همراه است) و وجود جواب بی‌معنی در مجموعه جواب‌های ممکن.

۱-٤ مدل طبقه‌بندی مقاوم بی‌احتیاط

تعریف ۱-۲: یک ابرصفحه در فضای R^m، مجموعه‌ای است از نقاطی به فرم $\{x|w^Tx+b=0\}$ که در آن بردار وزن w یک بردار غیر صفر در R^m و b یک اسکالر است. یک ابرصفحه دو زیرفضای بسته $\{x|w^Tx+b\geq 0\}$ و $\{x|w^Tx+b\leq 0\}$ می‌سازد [۲]. در مدل RSVM به خاطر همین توانایی ابرصفحه، یعنی توانایی تقسیم فضای ورودی یا فضای ویژگی به دو زیرفضا، از ابرصفحه برای طبقه‌بندی دوکلاسی داده‌ها استفاده می‌شود. مشکل اصلی مدل‌های مقاوم این است که در مجموعه جواب‌های ممکن مدل‌های مقاومی که تا پیش از این ارائه شده‌اند از جمله مدل RSVM، جوابی با بردار وزن صفر نیز وجود دارد که جوابی بی‌معنی است. چنین جوابی اصلاً ابرصفحه‌ای تشکیل نمی‌دهد و توانایی تقسیم فضای ورودی یا فضای ویژگی به دو زیرفضا را ندارد. طبقه‌بندی داده‌ها در مرحله آزمون با استفاده از این جواب، میسر نخواهد بود چرا که در چنین شرایطی، برچسب هر داده دلخواه x، مقداری ثابت و برابر با $sign(w^Tx+b)=sign(b)$ خواهد شد. بروز این مشکل قبلاً در نوعی مدل مبتنی بر SVM نیز گزارش شده است [۳].

تعریف ۱-۳: جوابی با بردار وزن برابر با بردار صفر، جوابی بی‌معنی برای یک مدل طبقه‌بندی نامیده می‌شود.

تعریف ۱-٤: به یک مجموعه داده‌های آموزشی که به ازای آن، جواب بهینه یک مدل طبقه‌بندی، جواب بی‌معنی باشد، مجموعه آموزشی «چاله» برای آن مدل طبقه‌بندی گوییم.

تعریف ۱-۵: به هر مدل طبقه‌بندی که در مجموعه جواب‌های ممکن آن، جواب بی‌معنی وجود داشته باشد یا به عبارت دیگر، به هر مدل طبقه‌بندی که حداقل یک مجموعه آموزشی چاله برای آن وجود داشته باشد، به دلیل احتمال گیرافتادن مدل در چاله و رسیدن به جواب بی‌معنی، مدل بی‌احتیاط[2] گفته می‌شود. جواب بهینه مدل RSVM، وابسته به آرایش داده‌های آموزشی، ممکن است جوابی بی‌معنی باشد لذا مدل RSVM یک مدل بی‌احتیاط است. شکل ۱ تعدادی مجموعه آموزشی چاله برای مدل بی‌احتیاط RSVM در فضای ورودی را نشان می‌دهد. به عبارت دیگر، مدل بی‌احتیاط RSVM در فضای ورودی نمی‌تواند هیچ یک از مجموعه‌های عدم قطعیت نشان داده شده در این شکل را طبقه‌بندی کند.

مدل مقاوم RSVM به ازای انواعی از مجموعه داده‌های توأم با عدم قطعیت، به مدل استاندارد تبدیل شده است. همچنین به ازای انواعی از مجموعه داده‌های توأم با عدم قطعیت با مدل استاندارد تقریب زده شده است که به آن مدل RSVM تقریبی ARSVM-HC[3] گفته می‌شود. در مورد مدل تقریبی ARSVM-HC بایستی در نظر داشت که حتی اگر در شرایطی خاص مثلاً در فضای ویژگی، به ازای مجموعه عدم قطعیت ابردایروی در فضای ورودی، جواب بهینه مدل RSVM جواب معنی‌داری باشد، با این حال ممکن است که جواب بهینه مدل، جوابی بی‌معنی باشد. لذا بنا به تعریفی که در ادامه می‌آید مدل ARSVM-HC یک مدل خیلی بی‌احتیاط است.

تعریف ۱-٦: یک مدل طبقه‌بندی بی‌احتیاط و تقریبی از آن مدل را در نظر بگیرید. به این مدل تقریبی، مدل خیلی بی‌احتیاط می‌گوییم اگر یکی از دو شرط زیر را داشته باشد:

1. Robust SVM

2. Incautious

3. Approximate RSVM-HC

الف) جواب بهینه مدل طبقه‌بندی بی‌احتیاط به ازای یک مجموعه آموزشی چاله، جواب بی‌معنی باشد آن گاه جواب بهینه مدل طبقه‌بندی تقریبی نیز جواب بی‌معنی باشد.

ب) حداقل به ازای یک مجموعه آموزشی، جواب بهینه مدل طبقه‌بندی تقریبی، جواب بی‌معنی و جواب بهینه مدل طبقه‌بندی بی‌احتیاط، جواب معنی‌دار باشد.

۱-۵ هدف و سازمان مقاله

هدف این مقاله، رفع چالش وجود جواب بی‌معنی در مجموعه جواب‌های ممکن و بهبود چالش‌های زیادبودن زمان آموزش و زمان آزمون مدل طبقه‌بندی مقاوم RSVM است. ساختار مقاله از پنج بخش اصلی تشکیل می‌شود. در بخش اول مفاهیم اساسی و تعاریف مدل‌های طبقه‌بندی مبتنی بر SVM، مقاوم و غیر مقاوم بودن، بی‌احتیاطی و استاندارد و غیر استاندارد بودن مدل‌ها تشریح می‌شود. در بخش دوم مروری بر نتایج برخی از پژوهش‌های انجام‌شده در زمینه رفع مشکلات مدل‌های طبقه‌بندی مبتنی بر SVM انجام می‌شود. بخش سوم به تشریح کامل روش پیشنهادی اختصاص دارد. در این بخش در ابتدا چالش‌های مدل‌های طبقه‌بندی مقاوم مبتنی بر SVM تشریح می‌شود و سپس ضمن ارائه روش پیشنهادی و تشریح گام‌های آن برای رفع چالش‌های مذکور، مفهوم جدید «احتیاط» و مدل جدید RSVM بااحتیاط معرفی و مزایای آن بیان می‌شود. در ادامه این بخش، روش‌های استانداردسازی و کاهش زمان آموزش و زمان آزمون مدل پیشنهادی آن مورد بررسی قرار گرفته و در انتها چهار مدل طبقه‌بندی مقاوم بااحتیاط استانداردشده جدید از مدل پیشنهادی استخراج و معرفی می‌گردد. بخش چهارم شامل آزمایش‌ها و نتایج بوده و در آن از مدل‌های پیشنهادی برای طبقه‌بندی داده‌های آموزشی ناقص و داده‌های آموزشی قطعی کامل استفاده شده و تأثیرات آن در کاهش نرخ خطای طبقه‌بندی به ازای داده‌های آموزشی ناقص و کاهش زمان آموزش به ازای داده‌های آموزشی قطعی کامل بررسی می‌گردد. در بخش پنجم با استفاده از یافته‌های آزمایش‌ها به نتیجه‌گیری پرداخته می‌شود.

۲- بررسی کارهای مرتبط

کامبار [۴] و سیمارد [۵]، مجموعه داده‌های توأم با عدم قطعیت را از طریق انتخاب برخی از اعضای مجموعه‌های عدم قطعیت به عنوان نماینده مجموعه، به مجموعه داده‌های آموزشی قطعی تبدیل کرده و از این طریق مدل مقاوم را به مدل طبقه‌بندی بر اساس داده‌های قطعی تبدیل نمودند و برای کاهش پیچیدگی، نمونه‌های مجازی را فقط از روی بردارهای پشتیبان با اهمیت بیشتر ایجاد نمودند. لوسلی و همکاران [۶] یک الگوریتم یادگیری برخط برای کاهش پیچیدگی و افزایش چشم‌گیر سرعت یادگیری مدل مذکور ارائه دادند. فانگ و همکاران [۷] مجموعه‌های عدم قطعیت را مجموعه‌های چندوجهی فرض کرده و سپس مدل طبقه‌بندی SVM مقاوم در فضای ورودی را به کمک لِم فارکاس نامتقارن به یک مدل استاندارد خطی تبدیل نمودند. جیاکومار و همکاران [۸] با استفاده از تکنیک مشابهی، مدل RSVM در فضای ورودی به ازای مجموعه‌های عدم قطعیت ابربیضوی را به یک مدل استاندارد برنامه‌ریزی نیمه‌معین تبدیل کردند. آنها در سال ۲۰۱۲ با استفاده از تکنیک مشابهی، مدل RSVM در فضای ورودی به ازای مجموعه‌های عدم قطعیت ابرمستطیلی را به یک مدل استاندارد برنامه‌ریزی درجه دوم تبدیل نمودند. کاریزوسا و همکاران [۹] مدل RSVM در فضای ورودی را به ازای مجموعه‌های عدم قطعیت ابرمستطیلی و به کمک عملیات پایه روی داده‌های مقدار بازه‌ای به یک مدل استاندارد برنامه‌ریزی خطی تبدیل کردند. ترافالیس و گیلبرت [۱۰] و [۱۱] مدل RSVM در فضای ورودی به ازای مجموعه‌های عدم قطعیت ابردایروی را به کمک نامساوی هلدر به یک مدل استاندارد SOCP تبدیل کردند. سپس مجموعه عدم قطعیت ابردایروی فضای ورودی را پس از تبدیل با تابع تبدیل کرنل گوسی، با یک مجموعه ابردایروی در فضای ویژگی تقریب زدند و به کمک این تقریب، مدل غیر استاندارد SVM مقاوم در فضای ویژگی را به یک مدل استاندارد SOCP تبدیل کردند و این مدل تقریبی را ARSVM-HC نامیدند که HC مخفف Hyper-circle می‌باشد. ژو [۱۲] و ژو و همکاران [۱۳] به اثبات این موضوع که مجموعه عدم قطعیت ابردایروی در فضای ورودی پس از تبدیل با تابع تبدیل کرنل گوسی به یک مجموعه ابردایروی در فضای ویژگی تبدیل می‌شود پرداختند. فرقانی و صدوقی [۱۴] این اثبات را رد کرده و نشان دادند که در واقع، مجموعه عدم قطعیت ابردایروی در فضای ورودی، پس از تبدیل با تابع تبدیل کرنل گوسی، زیرمجموعه‌ای از یک مجموعه ابردایروی در فضای ویژگی است نه برابر با یک مجموعه ابردایروی در فضای ویژگی. شی و همکاران [۱۵] مدل ARSVM-HC را برای حالت چندکلاسی توسعه دادند و برای تضمین تنک‌شدن بردار وزن طبقه‌بند، دوگان مدل را نیز به دست آوردند. سپس ارتباط بین متغیرهای مدل اولیه و مدل دوگان را که برای طبقه‌بندی در زمان آزمون لازم است به دست آوردند. ژو و همکاران [۱۶] و ژو و مانور [۱۷] و [۱۸] شکل جدیدی از عدم قطعیت را که در آن، شکل هیچ مجموعه عدم قطعیتی به صورت مستقل تعریف نمی‌شود بلکه شکل حاصل از نوعی تجمیع روی مجموعه‌های عدم قطعیت تعریف می‌شود، معرفی کرده و مدل RSVM در فضای ورودی را به یک مدل استاندارد SOCP تبدیل کردند. آن و سون [۱۹] به عنوان یک تقریب، با فرض نوشتن مجموعه‌های عدم قطعیت به صورت تابعی خطی از داده‌های آموزشی قطعی در فضای ویژگی به عنوان یک تقریب، مدل RSVM در فضای ویژگی را به یک مدل استاندارد برنامه‌ریزی درجه دوم تبدیل کردند. لیو و همکاران [۲۰] راه حلی تکراری برای حل مدل RSVM در فضای ویژگی در حالت یک‌کلاسی و به ازای مجموعه‌های عدم قطعیت ابردایروی ارائه نمودند که در آن هر بار یک دسته از متغیرها ثابت فرض شده و با حل یک مسئله بهینه‌سازی، دسته دیگر متغیرها به روش تقریبی به دست می‌آید و این روند تا برقراری شرط پایان، تکرار می‌شود. پوتین و ریچارد [۲۱] راه حلی تکراری برای حل مدل RSVM در فضای ویژگی به ازای مجموعه‌های عدم قطعیت ابربیضوی ارائه نمودند که در آن، ابتدا مدل SVM به ازای یک مجموعه آموزشی قطعی تشکیل‌شده از یک عضو از هر مجموعه عدم قطعیت اجرا می‌شود. سپس در هر مرحله با بسط مرتبه اول تیلور تابع کرنل، بدترین نقطه از هر مجموعه نسبت به طبقه‌بند به دست آمده و به مجموعه آموزشی قطعی اضافه می‌گردد و این روند تا زمانی تکرار می‌شود که بدترین نقطه، نقطه جدیدی نباشد. لی و همکاران [۲۲] ابتدا مدل SVM را به ازای مجموعه داده‌های آموزشی قطعی حل کردند و طبقه‌بندی را انجام دادند. سپس هر مجموعه عدم قطعیت و برچسب آن را به عنوان یک قانون غیر فازی تلقی نموده و طبقه‌بند نهایی را با ترکیب طبقه‌بند به دست آمده از مدل SVM و قوانین غیر فازی یادشده به دست آوردند. تئو و همکاران [۲۳] نشان دادند که مدل RSVM را می‌توان یک مدل محدب برنامه‌ریزی نیمه‌معین در نظر گرفت که البته حل دقیق آن ممکن نیست. لذا برای حل آن دو الگوریتم تقریبی پیشنهاد دادند. جی و همکاران [۲۴] مدل RSVM

در فضای ورودی و به ازای یک مجموعه داده آموزشی فازی که هر ویژگی از هر داده آن، یک عدد فازی مثلثی است را به مدل ساده‌تری تبدیل کردند. آنان ادعا کردند که مدل ساده‌شده، مدل برنامه‌ریزی درجه دوم محدبی است در حالی که تابع هدف این مدل از درجه چهار می‌باشد. فرقانی و صدوقی [۱۴] به کمک یک مثال نقض، محدب‌بودن مدل یادشده را رد کردند. یانگ و دانگ [۲۵] به منظور افزایش مقاومت در برابر نویز، با افزودن یک تابع ضرر کران‌دار مقاوم[1] به نام L_t-loss به مدل ماشین بردار پشتیبان با تابع ضرر پین‌بال PINSVM[2]، مدل مقاوم جدید TPINSVM را معرفی نمودند که تابع ضرر pinball را کوتاه‌تر، قانون «بِیز» را تأمین و تنک‌بودن را تضمین می‌کند. گائو و وانگ [۲۶] برای رفع مشکل کارایی و داده‌های نویزی مدل ماشین بردار پشتیبان دوتایی (TSVM)[3] در کاربردهای واقعی، مدل ماشین بردار پشتیبان دوتایی سریع و مقاوم FR-TSVM[4] را ارائه نمودند. چن و همکاران [۲۷] یک چارچوب برای مقاوم‌سازی SVM به منظور ارتقای سرعت و دقت پردازش تصویر چهره (FER)[5] و اشیا ارائه دادند که مبتنی بر سه کلید برای افزایش مقاومت بود. کاتسوماتا و تاکدا [۲۸] با استفاده از مدل ماشین بردار پشتیبان حساس به هزینه (CSSVM)[6] و کشف رابطه بین SVMهای تنظیم‌شده و SVMهای مقاوم، مدل جدید ماشین بردار پشتیبان حساس به هزینه مقاوم (RCSSVM)[7] را ارائه نمودند که در آن اندازه مجموعه‌های کراندار برای هر داده متفاوت است. تزلیپس و همکاران [۲۹] یک مدل طبقه‌بند با حداکثر حاشیه برای داده‌های توأم با عدم قطعیت با توزیع گاوسی (SVM-GSU)[8] ارائه نمودند که این مدل پیشنهادی عدم قطعیت ورودی‌ها را به شکل مؤثری در نظر گرفته و عملکرد طبقه‌بندی بهتر یا شناسایی بهتری نسبت به SVM استاندارد و مدل‌های مشتقه از آن دارد. فرقانی و صدوقی [۳۰] یک شبکه عصبی حداقل- حداکثر فازی با حاشیه متقارن جدید به نام (FMNWSM)[9] ارائه ارائه کردند که برای داده‌های آموزش و آزمون با توزیع احتمال یکسان، خطای طبقه‌بندی کمتری نسبت به شبکه‌های عصبی حداقل- حداکثر فازی سنتی دارد. همچنین یک روش سریع برای حل مدل SVM برای کاربرد در تنظیم پارامترهای قوانین فازی «اگر- آن گاه» ارائه کردند که از زمان آموزش و آزمون خیلی کمتر و نرخ خطای طبقه‌بندی بهتری نسبت به مدل‌های SVM و RSVM و مدل‌های فازی مبتنی بر SVM برخوردار بود. آنان در پژوهشی دیگر برای غلبه بر دشواری حل مدل ماشین بردار پشتیبان مقاوم RoSVM[10] برای داده‌های بیضوی، فرم اولیه با یک مدل برنامه‌ریزی محدب مخروط درجه دوم SOCP تقریب زدند که فرم اولیه تقریبی RoSVM نامیده می‌شود. جینگ و یانگ [۳۱] با فرض شناخته شده بودن میانگین و کوواریانس هر کلاس از طبقه‌بندی دوکلاسه و چندکلاسه، با مقاوم‌سازی مدل یادگیری ماشین حداکثری (ELM)[11]، یک چارچوب مقاوم (RELM)[12] را برای طبقه‌بندی داده‌های گم‌شده ارائه کرده و با استفاده از نامساوی Chebyshev-Cantelli آن را به صورت یک مدل برنامه‌ریزی مخروط درجه دوم با جواب بهینه سراسری بازنویسی کردند. عادلی و همکاران [۳۲] برای مقاوم‌سازی روش‌های متمایزکننده در برابر داده‌های پرت و نویز در ویژگی‌ها، با استفاده از داده‌های آموزشی برچسب‌دار و داده‌های آزمون بدون برچسب، یک طبقه‌بند متمایزکننده مقاوم نیمه نظارت‌شده مبتنی بر فرمول حداقل مربعات تحلیل گسسته خطی، برای یافتن هم‌زمان داده‌های پرت و ویژگی‌های نویزی پیشنهاد دادند.

۳- روش پیشنهادی

در یک جمع‌بندی می‌توان گفت که کلیه مدل‌های طبقه‌بندی مقاوم ارائه‌شده تا کنون، مدل‌های طبقه‌بندی بی‌احتیاط هستند. در این مقاله برای رفع این مشکل، مدل طبقه‌بندی مقاوم بااحتیاط[13] معرفی می‌گردد و سپس به مدل استاندارد SOCP تبدیل می‌شود. به طور خلاصه مراحل روش پیشنهادی به شرح ذیل است:

گام اول) معرفی مدل RSVM بااحتیاط: نشان داده خواهد شد که برخلاف مدل طبقه‌بندی مقاوم بی‌احتیاط، جواب بهینه مدل طبقه‌بندی مقاوم بااحتیاط بر اساس تعریف ۱-۷ (که در قسمت بعد ارائه خواهد شد) و قید $w\neq 0$ در تابع ضرر جدید و نبودن هیچ بردار با وزن صفر، ضرورتاً یک جواب معنی‌دار است.

گام دوم) استانداردسازی مدل RSVM بااحتیاط: مدل RSVM بااحتیاط در فضای ویژگی به ازای مجموعه‌های عدم قطعیت ابردایروی و ابربیضوی به مدل استاندارد SOCP تبدیل می‌شود ولی با وجود این، زمان آموزش همچنان یک مشکل این مدل استاندارد و به طور کلی یک مشکل مدل‌های مبتنی بر SVM است. در این گام دو مدل RSVM بااحتیاط استاندارد جدید معرفی می‌گردد.

گام سوم) معرفی مدل طبقه‌بندی بااحتیاط تکه‌ای برای غلبه بر مشکل بالابودن زمان آموزش: برای این منظور در مدل RSVM بااحتیاط استانداردشده جدید، مدل طبقه‌بندی تکه‌ای پیشنهاد می‌شود که هر تکه از طبقه‌بند آموزش‌یافته توسط این مدل، حاصل اجرای یک مدل طبقه‌بندی مقاوم بااحتیاط به ازای داده‌های آموزشی واقع در بخشی از فضاست. زمان آموزش مدل تکه‌ای تقریباً رابطه معکوس و زمان آزمون مدل تکه‌ای تقریباً رابطه مستقیمی با تعداد تقسیمات فضا دارد. در این گام دو مدل RSVM بااحتیاط استانداردشده تکه‌ای جدید، با قوانین فازی مرتبه صفر و مرتبه یک معرفی می‌گردد.

گام چهارم) تبدیل مدل RSVM بااحتیاط استانداردشده تکه‌ای جدید به مدل نزدیک‌ترین همسایه: در این گام ثابت می‌شود که در حالت خاص، این مدل تکه‌ای به یک مدل نزدیک‌ترین همسایه تبدیل می‌شود که فاقد زمان آموزش است. سپس نشان داده می‌شود که جواب مدل نزدیک‌ترین همسایه به ازای انواع خاصی از داده‌های توأم با عدم قطعیت به روش تحلیلی به دست می‌آید. مدل نزدیک‌ترین همسایه در طبقه‌بندی مجموعه‌های عدم قطعیت از کلاس‌های متفاوت که دارای عضو مشترک هستند، دچار مشکل شده و برای طبقه‌بندی نقاطی از فضا که به بیش از یک کلاس نسبت داده شده‌اند از استراتژی طبقه‌بندی تصادفی استفاده می‌کند. در حالی که این نقاط منتسب به چندطبقه، همانند نقاطی از فضا هستند که به هیچ طبقه‌ای نسبت داده نشده‌اند و لذا باید مثلاً از یک

1. Robust Bounded Loss Function
2. Support Vector Machine with Pinball Loss
3. Twin Support Vector Machine
4. Fast and Robust TSVM
5. Facial Expression Recognition
6. Cost Sensitive SVM
7. Regularized Cost Sensitive SVM
8. SVM with Gaussian Sample Uncertainty
9. Fuzzy Min-Max Neural Network with Symmetric Margin
10. Robust SVM
11. Extreme Learning Machine
12. Robust Extreme Learning Machine
13. Cautious

استراتژی طبقه‌بندی مبتنی بر یادگیری برای طبقه‌بندی این نقاط منتسب به چند طبقه استفاده شود. برای رفع این مشکل مدل نزدیک‌ترین همسایه مرتبه یک پیشنهاد می‌شود که به جای استفاده از استراتژی طبقه‌بندی تصادفی برای طبقه‌بندی نقاط منتسب به چند طبقه، از استراتژی طبقه‌بندی RSVM بااحتیاط استفاده می‌کند. سپس نشان داده می‌شود که جواب مدل نزدیک‌ترین همسایه مرتبه یک به ازای انواع خاصی از داده‌های توأم با عدم قطعیت به روش تحلیلی به دست می‌آید.

نوآوری‌های روش پیشنهادی

الف) استفاده از مدل طبقه‌بندی مقاوم بااحتیاط جدید CRSVM برای حذف جواب بی‌معنی از مجموعه جواب‌ها

ب) استفاده از مدل طبقه‌بندی مقاوم بااحتیاط جدید CRSVM به عنوان اجزای مدل طبقه‌بندی تکه‌ای برای کاهش زمان آموزش

ج) بیان ارتباط بین مدل تکه‌ای پیشنهادی و مدل نزدیک‌ترین همسایه

د) به دست آوردن جواب بهینه مدل نزدیک‌ترین همسایه به ازای انواع خاصی از مجموعه‌های توأم با عدم قطعیت به روش تحلیلی

ه) ارائه مدل جدید نزدیک‌ترین همسایه مرتبه یک که برخلاف مدل نزدیک‌ترین همسایه معمولی، برای طبقه‌بندی داده‌های منتسب به چند طبقه، به جای استراتژی طبقه‌بندی تصادفی از مدل CRSVM استفاده می‌کند.

و) به دست آوردن جواب مدل نزدیک‌ترین همسایه مرتبه یک به ازای انواع خاصی از مجموعه‌های توأم با عدم قطعیت به روش تحلیلی

۳-۱ گام اول: معرفی مدل طبقه‌بند مقاوم بااحتیاط (CRSVM)

تعریف ۱-۷: یک مدل طبقه‌بندی، مدل بااحتیاط خوانده می‌شود اگر جواب بهینه آن مدل به ازای هر مجموعه داده‌های آموزشی ممکن، جوابی معنی‌دار باشد.

مدل طبقه‌بندی مقاوم بااحتیاط، همان مدل RSVM با تابع ضرر جدید است که آن را به اختصار CRSVM[1] می‌نامیم. برای تبدیل RSVM به CRSVM دو راه وجود دارد. مدل RSVM (۳) را در نظر بگیرید

$$\min_{w,b,\xi}\left(\not{g}(w,b)+C\sum_{i=1}^{n}\max_{x^{(\varphi)}\in u_{(i)}^{(\varphi)}}\ell(w,b;x^{(\varphi)},y_{(i)})\right) \quad (۳)$$

راه اول برای این که جواب مدل (۳) معنی‌دار باشد کافی است قید $w\neq 0$ به طور مستقیم به آن اضافه شود[2]. در این صورت، مدل (۴) حاصل می‌شود که بنا به تعریف ۱-۷، یک مدل CRSVM محسوب می‌شود

$$\begin{aligned}&\min_{w,b,\xi}\left(\not{g}(w,b)+C\sum_{i=1}^{n}\max_{x^{(\varphi)}\in u_{(i)}^{(\varphi)}}\ell(w,b;x^{(\varphi)},y_{(i)})\right) \quad (۴)\\ &s.t, w\neq 0\end{aligned}$$

مدل (۴)، یک مدل مقاوم بااحتیاط هست ولی یک مدل محدب استاندارد نیست، لذا به دست آوردن جواب بهینه آن بسیار زمان‌بر است.

راه دوم برای تبدیل مدل (۳) به مدل CRSVM این است که قید $w\neq 0$ به طور ضمنی به آن اضافه شود. مثلاً می‌توان تابع ضرر جدیدی به صورت (۵) تعریف کرد

$$\tilde{\ell}(w,b;x^{(\varphi)},y_{(i)})=\frac{\ell(w,b;x^{(\varphi)},y_{(i)})}{\|w\|_1} \quad (۵)$$

در این صورت، مدل (۳) به صورت (۶) درمی‌آید

$$\min_{w,b,\xi}\left(\not{g}(w,b)+C\sum_{i=1}^{n}\max_{x^{(\varphi)}\in u_{(i)}^{(\varphi)}}\tilde{\ell}(w,b;x^{(\varphi)},y_{(i)})\right) \quad (۶)$$

بدیهی است که جواب با وزن صفر یعنی "$w=0$" در مجموعه جواب‌های مدل (۵) وجود ندارد چرا که تابع $\tilde{\ell}(w,b;x^{(\varphi)},y_{(i)})$ به ازای "$w=0$" تعریف نشده است. لذا بنا به تعریف، مدل (۵) نیز یک مدل CRSVM محسوب می‌شود.

شکل کلی تابع ضرر $\tilde{\ell}(w,b;x^{(\varphi)},y_{(i)})$ را می‌توان به صورت (۷) نوشت

$$\tilde{\ell}(w,b;x^{(\varphi)},y_{(i)})=\frac{\ell(w,b;x^{(\varphi)},y_{(i)})}{\not{g}(w,b)} \quad (۷)$$

که در آن $\not{g}(w,b)$ تابعی مثبت است به طوری که اگر $w\neq 0$ باشد آن گاه $\not{g}(w,b)$ نیز برابر با صفر است. به عبارت دیگر، شروط لازم تابع $\not{g}(w,b)$ عبارتند از:

شرط اول: شرط مثبت‌بودن تابع $\not{g}(w,b)$،

شرط دوم: شرط صفرشدن تابع $\not{g}(w,b)$ در جواب بی‌معنی "$w=0$".

مثبت‌بودن $\not{g}(w,b)$ برای حفظ علامت تابع ضرر لازم است (مقدار تابع ضرر، همواره باید مقداری نامنفی باشد). شرط دومی که برای تابع $\not{g}(w,b)$ ذکر شد جواب بی‌معنی "$w=0$" را از مجموعه جواب‌های ممکن مدل (۵) حذف می‌کند. متأسفانه مدل RSVM بااحتیاط (۵) نیز مدل استانداردی نیست و لذا به دست آوردن جواب بهینه آن بسیار زمان‌بر است.

۳-۲ گام دوم: استانداردسازی مدل CRSVM

دو روش برای استانداردسازی مدل طبقه‌بندی مقاوم RSVM مورد استفاده قرار می‌گیرد، یکی طبقه‌بندی مقاوم در فضای ورودی و دیگری طبقه‌بندی مقاوم در فضای ویژگی. می‌توان مدل RSVM بااحتیاط (۶) را به صورت (۸) نوشت

$$\begin{aligned}&\min_{\theta,w,b,\xi}\not{g}(w,b)+C\sum_{i=1}^{n}\xi_{(i)}\\ &s.t,\begin{cases}\left\|\begin{matrix}2\theta_{(i)}\\ \xi_{(i)}-\not{g}(w,b)\end{matrix}\right\|_2\le\xi_{(i)}+\not{g}(w,b), i=1,2,\dots,n\\ \xi_{(i)},\theta_{(i)}\ge 0, i=1,2,\dots,n\\ \theta_{(i)}\ge 1-\min_{x^{(\varphi)}\in u_{(i)}^{(\varphi)}}(y_{(i)}(w^Tx^{(\varphi)}+b)),\\ \quad i=1,2,\dots,n\end{cases}\end{aligned} \quad (۸)$$

هنوز آخرین دسته از قیود مدل (۸)، قیود مخروطی درجه دوم نیستند ولی به ازای انواع خاصی از مجموعه‌های عدم قطعیت، به قیود مخروطی درجه دوم تبدیل می‌شوند. در ضمن، فرض شد $\not{g}(w,b)$ یک چندجمله‌ای از درجه یک است که دو شرط لازم $\not{g}(w,b)$ را ارضا می‌کند، از همین رو لازم است این تابع دقیقاً مشخص گردد. این کار در دو حالت مختلف انجام می‌شود: حالت اول در فضای ورودی یعنی به ازای $\varphi(x)=x$ و

1. Cautious RSVM

۲. در [۳] نوعی مدل مبتنی بر SVM ارائه شده که همان مشکل بی‌معنی‌شدن جواب جواب بهینه را دارد (البته این مدل، یک مدل مقاوم نیست). در این مرجع، قید $\|w\|_2\ge 1$ به مدل اضافه شده است. افزودن این قید، اولاً موجب نامحدب‌شدن مدل می‌شود که زمان حل مدل را شدیداً افزایش می‌دهد. ثانیاً مدل را از یک مدل مبتنی بر SVM خارج می‌سازد. لذا دیگر تئوری‌های اثبات‌شده SVM همچون تئوری تعمیم‌پذیری در مورد آن قابل استفاده نیست.

حالت دوم در فضای ویژگی. در این مقاله، مدل‌های استاندارد CRSVM پیشنهادی از استانداردسازی در فضای ویژگی به دست می‌آیند، لذا از استانداردسازی در فضای ورودی صرف نظر شده است.

استانداردسازی مدل CRSVM در فضای ویژگی

در این حالت، نوع عدم قطعیت داده‌ها در فضای ویژگی معلوم فرض شده است. برای استانداردسازی مدل RSVM (۳) و استفاده از حقه کرنل، نخست بسط بردار وزن در این مدل را به دست می‌آوریم و برای این کار از قضیه ۱ استفاده می‌کنیم.

قضیه ۱: اگر $\boldsymbol{\mathcal{U}}_{(i)}^{(\varphi)}$ها مجموعه‌هایی با تعداد متناهی عضو باشند و $\not\in(w,b)=(1/2)\|w\|_2^2$، آن گاه در نقطه بهینه مدل (۳) بردار وزن به صورت (۹) خواهد بود

$$w=\sum_{j=1}^{n}\sum_{x^{(\varphi)}\in\boldsymbol{\mathcal{U}}_{(j)}^{(\varphi)}} a_{(j,x^{(\varphi)})}y_{(j)}x^{(\varphi)} \qquad (۹)$$

که در آن $\alpha_{(j,x^{(\varphi)})}\geq 0$ می‌باشد. البته تعداد اعضای مجموعه‌های عدم قطعیت، معمولاً بسیار زیاد یا بی‌نهایت است، لذا استفاده از قضیه ۱ برای بسط بردار وزن کارامد یا ممکن نیست. در ادامه نشان داده می‌شود که در صورت استفاده از انواع خاصی از مجموعه‌های عدم قطعیت ابردایروی و ابربیضوی در فضای ویژگی، بردار وزن را می‌توان به صورت ترکیب خطی تعداد اندکی از اعضای مجموعه‌های عدم قطعیت نوشت.

استانداردسازی مدل CRSVM در فضای ویژگی به ازای مجموعه‌های عدم قطعیت ابردایروی- (CRSVM-HC)

قضیه ۲: اگر $\boldsymbol{\mathcal{U}}_{(i)}^{(\varphi)}$ ابردایره‌ای به مرکز $m_{(i)}^{(\varphi)}$ و شعاع $\gamma_{(i)}r_{(i)}^{(\varphi)}$ باشد، آن گاه به شرط آن که $\not\in(w,b)=(1/2)\|w\|_2^2$ باشد، در نقطه بهینه مدل (۸) بردار وزن به صورت (۱۰) خواهد بود

$$w=\sum_{j=1}^{n}a_{(j)}y_{(j)}m_{(j)}^{(\varphi)} \quad , \quad \forall j: a_{(j)}\geq 0 \qquad (۱۰)$$

قضیه ۳ (تبدیل مدلی با تابع هدف درجه دوم به یک مدل استاندارد SOCP): مدل (۱۱) را در نظر بگیرید

$$\begin{aligned}&\min_x \|x\|_2^2\\ &s.t, \{x\in\mathscr{C},\end{aligned} \qquad (۱۱)$$

که در آن $\mathscr{C}$ یک مخروط درجه دوم است. این مدل با مدل SOCP (۱۲) معادل است

$$\min_{g,x} g \qquad s.t,\begin{cases}\left\|\begin{matrix}2\\ g-1\end{matrix}\right\|_2\leq g+1\\ x\in\mathscr{C}\end{cases} \qquad (۱۲)$$

بنا به قضایای ۲ و ۳ مدل غیر استاندارد RSVM بااحتیاط (۸) به مدل استاندارد SOCP (۱۳) تبدیل می‌شود

مدل (۱۳) را CRSVM-HC می‌نامیم. پس از حل مدل (۱۳)، داده آزمون x به طبقه $sign(f(x))$ نسبت داده می‌شود

$$f(x)=w^T\varphi(x)+b=\alpha^T\mathbb{K}(x)+b \qquad (۱۴)$$

که در آن

$$\min_{g,a,b,\xi,\theta} \frac{1}{2}g+C\sum_{i=1}^{n}\xi_{(i)}$$

$$\begin{cases}\left\|\begin{matrix}2\theta_{(i)}\\ \xi_{(i)}-\sum_{j=1}^{n}a_{(j)}\end{matrix}\right\|_2\leq \xi_{(i)}+\sum_{j=1}^{n}a_{(j)} \ , \ i=1,2,\dots,n\\ \left\|\begin{matrix}2\mathbb{K}^{\frac{1}{2}}a\\ g-1\end{matrix}\right\|_2\leq g+1\\ a_{(j)},\theta_{(i)},\xi_{(i)}\geq 0 \ , \ i=1,2,\dots,n\\ \theta_{(i)}\geq 1-y_{(i)}(a^T m_{(i)}^{(\mathbb{K})}+b)+\gamma_{(j)}r_{(j)}^{(\varphi)}t \ , \ i=1,2,\dots,n\\ \left\|\mathbb{K}^{\frac{1}{2}}a\right\|_2\leq t\end{cases} \qquad (۱۳)$$

$$\mathbb{k}(x)^T=(y_{(1)}\mathbb{k}_{(1)}(x),y_{(2)}\mathbb{k}_{(2)}(x),\dots,y_{(n)}\mathbb{k}_{(n)}(x)) \qquad (۱۵)$$

$$\mathbb{k}_{(i)}(x)=\frac{1}{\aleph_{(i)}}\sum_{j=1}^{\aleph_{(i)}}k(x,\mathfrak{X}_{(i,j)}) \qquad (۱۶)$$

لذا برای حداقل‌شدن زمان آزمون باید بردار α تنک باشد. برای این منظور می‌توان از تابع هموارسازی $\|\alpha\|_1$ به جای تابع هموارسازی $1/2\left\|\mathbb{K}^{1/2}\alpha\right\|_2^2$ استفاده نمود.

استانداردسازی مدل CRSVM در فضای ویژگی به ازای مجموعه‌های عدم قطعیت ابربیضوی- (CRSVM-HE)

قضیه ٤: اگر $\boldsymbol{\mathcal{U}}_{(i)}^{(\varphi)}$ یک ابربیضی به مرکز $m_{(i)}^{(\varphi)}=(1/\aleph_{(i)})\sum_{j=1}^{\aleph_{(i)}}\varphi(\mathfrak{X}_{(i,j)})$ و کواریانس $\hat{C}_{(i)}^{(\varphi)}=y_{(i)}^2C_{(i)}^{(\varphi)}$ باشد که در آن

$$C_{(i)}^{(\varphi)}=\frac{1}{\aleph_{(i)}}\sum_{j=1}^{\aleph_{(i)}}(\varphi(\mathfrak{X}_{(i,j)})-m_{(i)}^{(\varphi)})(\varphi(\mathfrak{X}_{(i,j)})-m_{(i)}^{(\varphi)})^T \qquad (۱۷)$$

آن گاه به شرط آن که $\not\in(w,b)=(1/2)\|w\|_2^2$، در نقطه بهینه مدل غیر استاندارد RSVM بااحتیاط (۷)، بردار وزن به صورت (۱۸) خواهد بود

$$w=\sum_{i=1}^{n}\sum_{j=1}^{\aleph_{(i)}}\alpha_{(i,j)}\varphi(\mathfrak{X}_{(i,j)}) \ , \ \forall i,j:\alpha_{(i,j)}\in R \qquad (۱۸)$$

پس از انجام مراحلی و بنا به قضیه ۳، مدل غیر استاندارد RSVM بااحتیاط (۸) به مدل استاندارد SOCP (۱۹) تبدیل می‌شود که آن را مدل CRSVM-HE می‌نامیم

$$\min_{\hat{a},b,\xi,\theta} \frac{1}{2}g+C\sum_{i=1}^{n}\xi_{(i)}$$

$$s.t\begin{cases}\left\|\begin{matrix}2\theta_i\\ \xi_{(i)}-\sum_{i=1}^{N}\hat{\alpha}_{(i)}\end{matrix}\right\|_2\leq \xi_{(i)}+\sum_{i=1}^{n}\hat{\alpha}_{(i)} \ , \ i=1,2,\dots,n\\ \left\|\begin{matrix}2k^{\frac{1}{2}}\hat{\alpha}\\ g-1\end{matrix}\right\|\leq g+1\\ \sum_{i=1}^{N}\hat{\alpha}_{(i)}\geq 0\\ \theta_i,\xi_{(i)}\geq 0 \ , \ i=1,2,\dots,n\\ \theta_i\geq 1-y_{(i)}(\hat{\alpha}^T m_{(i)}^{(k)}+b)+y_{(i)}\left\|C_{(i)}^{(k)^{\frac{1}{2}}}\hat{\alpha}\right\|_2 \ ,\\ \quad i=1,2,\dots,n\end{cases} \qquad (۱۹)$$

۳-۳ گام سوم: ارائه مدل پیشنهادی CRSVM تکه‌ای- (مدل‌های PiCRSVM0 و PiCRSVM1)

اگرچه در بخش قبلی، مدل CRSVM در فضای ویژگی به ازای انواعی از مجموعه‌های عدم قطعیت به مدل استاندارد SOCP تبدیل شد، ولی تعداد متغیرها و قیدهای مدل‌های حاصل زیاد است که این امر منجر به افزایش زمان آموزش می‌شود. یک روش مرسوم برای طبقه‌بندی غیر خطی و کاهش زمان آموزش یک مدل طبقه‌بندی، استفاده از یک مدل تکه‌ای می‌باشد. در این مقاله برای طبقه‌بندی غیر خطی و کاهش زمان آموزش مدل CRSVM، مدل CRSVM تکه‌ای معرفی می‌شود. در مدل CRSVM تکه‌ای پیشنهادی، مشابه با دیگر مدل‌های تکه‌ای، نخست کل فضای مسئله به چند بخش تقسیم می‌شود. سپس با حل مدل CRSVM به ازای داده‌های هر بخش از زیرفضا[1]، یک طبقه‌بند و یک قانون فازی برای هر بخش از زیرفضا حاصل می‌شود. در پایان به کمک یک روش استدلال فازی، نتایج تجمیع شده و یک طبقه‌بند غیر خطی حاصل می‌شود.

مرحله آموزش مدل CRSVM تکه‌ای

حالت دوکلاسی مدل پیشنهادی دارای دو گام زیر در مرحله آموزش می‌باشد:

الف) بخش‌بندی[2] فضای ورودی به چند زیرفضای کوچک‌تر (مثلاً به روش ارائه‌شده در [۳۳][3]).

ب) اجرای دستورات زیر به ازای iامین زیرفضا:

ب-۱) در صورتی که زیرفضا شامل داده‌هایی متعلق به هر دو طبقه می‌باشد: اجرای مدل CRSVM به ازای داده‌های این زیرفضا و سپس تشکیل قانون فازی مرتبه یک[4] زیر:

- اگر x در مجموعه فازی $A_{(i)}$ هست آن گاه

$$f_{(i)}(x) = w_{(i)}^T \varphi(x) + b_{(i)} \qquad (۲۰)$$

که در آن A_i مجموعه‌ای فازی است که میزان تعلق به iامین زیرفضا را مشخص می‌کند و $w_{(i)}$ و $b_{(i)}$ بردار وزن و بایاس ابرصفحه طبقه‌بند حاصل از اجرای مدل CRSVM به ازای داده‌های iامین زیرفضاست.

ب-۲) در صورتی که زیرفضا فقط شامل داده‌های متعلق به طبقه $y_{A_{(i)}}$ است، در این حالت دو وضعیت می‌تواند وجود داشته باشد:

ب-۲-۱) تشکیل قانون فازی مرتبه صفر[5] زیر:

- اگر x در مجموعه فازی $A_{(i)}$ هست آن گاه

$$f_{(i)}(x) = y_{A_{(i)}} \qquad (۲۱)$$

در صورت استفاده از این بند در مدل CRSVM تکه‌ای، یعنی تشکیل قوانین فازی مرتبه صفر در مرحله ب-۲، مدل حاصل را PiCRSVM0[6] می‌نامیم.

ب-۲-۲) یا تشکیل قانون فازی مرتبه یک زیر:

- اگر x در مجموعه فازی $A_{(i)}$ هست آن گاه

$$f_{(i)}(x) = w_{(i)}^T \varphi(x) + b_{(i)} \qquad (۲۲)$$

که در آن $w_{(i)}$ و $b_{(i)}$ بردار وزن و بایاس ابرصفحه طبقه‌بند حاصل از اجرای مدل CRSVM به ازای داده‌های iامین زیرفضا و نزدیک‌ترین داده توأم با عدم قطعیت از طبقه $-y_{A_{(i)}}$ است. در صورت استفاده از این بند در مدل CRSVM تکه‌ای، یعنی تشکیل قوانین فازی مرتبه یک در مرحله ب-۲ مدل، مدل حاصل را PiCRSVM1[7] می‌نامیم.

مرحله آزمون مدل CRSVM تکه‌ای

در مرحله آزمون در حالت دوکلاسی، داده آزمون x به طبقه $sign(f(x))$ نسبت داده می‌شود که $f(x)$ به کمک قوانین فازی ایجادشده در مرحله آموزش و یک روش استدلال فازی مثل روش «انتخاب قانون غالب بر همه»[8] به صورت (۲۳) به دست می‌آید

$$f(x) = f_{(i)}(x) \qquad (۲۳)$$

به شرطی که (۲۴) برقرار باشد

$$\mu_{A_{(i)}}(x) = \max\{\mu_{A_{(j)}}(x) \,|\, j = ۱,۲,\ldots,l\} \qquad (۲۴)$$

ایجاد قوانین فازی مرتبه صفر به ازای هر بخش

اگر هر بخش فقط یک مجموعه عدم قطعیت را دربرگیرد، فقط قسمت ب-۲ مدل RSVM بااحتیاط تکه‌ای اجرا خواهد شد. حال اگر به ازای هر بخش و مطابق با توضیح بند ب-۲-۱، یک قانون فازی مرتبه صفر ایجاد شود، همه قوانین فازی ایجادشده از نوع قوانین فازی مرتبه صفر خواهند بود و طبقه‌بند حاصل، فقط وابسته به توابع عضویت استفاده‌شده در هر قانون فازی می‌شود چرا که از مدل CRSVM برای ایجاد قانون فازی به ازای هیچ بخشی از فضا استفاده نشده است. در چنین شرایطی، مدل CRSVM تکه‌ای تبدیل به یک مدل فازی مرتبه صفر می‌شود.

۳-۴ گام چهارم: تبدیل مدل CRSVM تکه‌ای به مدل نزدیک‌ترین همسایه

مدل تکه‌ای در حالت خاص، به مدل نزدیک‌ترین همسایه مرتبه صفر تبدیل می‌شود که فاقد زمان آموزش است. سپس جواب مدل نزدیک‌ترین همسایه مرتبه صفر به ازای انواع خاصی از داده‌های توأم با عدم قطعیت به روش تحلیلی تعیین می‌شود. تابع تصمیم مدل اولین نزدیک‌ترین همسایه به ازای داده‌های آموزشی قطعی و داده آزمون x به صورت (۲۵) می‌شود

$$f(x) = y_{(i)} \qquad (۲۵)$$

به شرطی که (۲۶) برقرار باشد

$$\min_{z \in u_{(i)}} d(x,z) \le \min\{\min_{z \in u_{(j)}} d(x,z) \,|\, j \ne i\} \qquad (۲۶)$$

که در آن $d(x,x_{(i)})$ یک معیار اندازه‌گیری فاصله است. پس از آن، این مدل، مدل نزدیک‌ترین همسایه مرتبه صفر خوانده می‌شود چرا که تابع تصمیم آن، حاوی یک چندجمله‌ای درجه صفر است.

تابع تصمیم مدل اولین نزدیک‌ترین همسایه مرتبه صفر به ازای داده‌های آموزشی توأم با عدم قطعیت و داده آزمون x به صورت همان (۲۵) خواهد بود به شرطی که (۲۶) برقرار باشد.

1. Sub-Space
2. Partitioning
3. زمان اجرای این الگوریتم بخش‌بندی خیلی کم و مرتبه زمانی آن $O(nP)$ می‌باشد که P تعداد تقسیمات فضاست.
4. First-Order
5. Zero-Order
6. Piece-Wise CRSVM with Zero-Order Fuzzy Rules
7. Piece-Wise CRSVM with One-Order Fuzzy Rules
8. Take-Winner-All

قضیه ۵ ارتباط بین مدل نزدیک‌ترین همسایه مرتبه صفر و مدل CRSVM تکه‌ای را در شرایطی بیان می‌کند که هر بخش حاصل از بخش‌بندی فضا، فقط یک مجموعه عدم قطعیت را دربرمی‌گیرد. همه قوانین فازی ایجادشده از نوع قوانین فازی مرتبه صفر هستند و مدل CRSVM تکه‌ای تبدیل به یک مدل فازی مرتبه صفر شده است.

قضیه ۵: قوانین فازی مرتبه صفر زیر (که در مدل تکه‌ای پیشنهادی تولید می‌شود) را در نظر بگیرید:

«اگر x در مجموعه فازی $A_{(i)}$ هست آن گاه $f(x)=y_{(i)}$ »

$i=1,2,\ldots,n$

اگر از روش «انتخاب قانون غالب بر همه» برای استدلال فازی استفاده شود و تابع عضویت مجموعه فازی $A_{(i)}$ به صورت زیر باشد، آن گاه خروجی مدل فازی مرتبه صفر حاصل با خروجی مدل نزدیک‌ترین همسایه مرتبه صفر به صورت (۲۷) خواهد بود

$$\mu_{(i)}(x)=\begin{cases}1 & \min\limits_{z\in u_{(i)}} d(x,z)\le \min\{\min\limits_{z\in u_{(j)}} d(x,z)\,|\,j\neq i\}\\ 0 & \text{otherwise}\end{cases} \qquad (۲۷)$$

محاسبه $\min\limits_{z\in u_{(i)}} d(x,z)$ در مدل فازی مرتبه صفر یا مدل نزدیک‌ترین همسایه مرتبه صفر، وابسته به نوع مجموعه‌های عدم قطعیت، می‌تواند بسیار زمان‌بر باشد (به خصوص اگر $u_{(i)}$ نامحدب باشد). در ادامه به کمک نکته زیر نشان داده می‌شود در صورتی که مجموعه‌های عدم قطعیت، ابرمستطیلی یا ابردایروی باشند و در ضمن از نوع خاص معیار فاصله استفاده شود، جواب مسئله $\min\limits_{z\in u_{(i)}} d(x,z)$ به روش تحلیلی به دست می‌آید.

نکته: فرض کنید مجموعه‌های عدم قطعیت متعلق به طبقات متفاوت، فاقد عضو مشترک باشند. اگر $x\in u_{(i)}$، $\min\limits_{z\in u_{(i)}} d(x,z)=0$ و $\min\{\min\limits_{z\in u_{(j)}} d(x,z)\,|\,j\neq i\}>0$، در این صورت با توجه به تابع تصمیم مدل نزدیک‌ترین همسایه مرتبه صفر، برچسب x برابر با $y_{(i)}$ می‌شود. در ضمن اگر به جای $\min\limits_{z\in u_{(i)}} d(x,z)$ مقداری منفی جایگزین شود، باز هم برچسب x همان $y_{(i)}$ می‌شود.

قضیه ٦: فرض کنید مجموعه‌های عدم قطعیت متعلق به طبقات متفاوت، فاقد عضو مشترک باشند و $u_{(i)}$ یک ابردایره به فرم $\|x-m_{(i)}\|_2\le r_{(i)}$ به مرکز $m_{(i)}$ و شعاع $r_{(i)}$ باشد و $d(x,z)=\|x-z\|_2$، آن گاه تابع تصمیم مدل نزدیک‌ترین همسایه مرتبه صفر به صورت همان (۲۵) خواهد بود به شرطی که (۲۸) برقرار باشد

$$\|x-m_{(i)}\|_2-r_{(i)}\le \min\{\|x-m_{(k)}\|_2-r_{(k)}\,|\,k\neq i\} \qquad (۲۸)$$

قضیه ۷ نشان می‌دهد که مدل نزدیک‌ترین همسایه مرتبه صفر همچون مدل RSVM حاشیه سخت، یک مدل طبقه‌بندی نامتغیر است.

قضیه ۷: اگر مجموعه‌های عدم قطعیت، فاقد عضو مشترک باشند آن گاه تابع تصمیم مدل نزدیک‌ترین همسایه مرتبه صفر به ازای داده‌های توأم با عدم قطعیت، نامتغیر است.

لذا بنا به تعریف مدل نزدیک‌ترین همسایه مرتبه صفر به صورت (۲۹) خواهد بود

$$\forall x\in u_{(i)}: f(x)=y_{(i)}=f(x_{(i)}) \qquad (۲۹)$$

مدل نزدیک‌ترین همسایه مرتبه صفر در طبقه‌بندی مجموعه‌های عدم قطعیت از کلاس‌های متفاوت که دارای عضو مشترک هستند، دچار مشکل شده و برای طبقه‌بندی نقاطی از فضا که به بیش از یک کلاس نسبت داده شده‌اند از استراتژی طبقه‌بندی تصادفی استفاده می‌کند در حالی که این نقاط منتسب به چند طبقه، همانند نقاطی از فضا هستند که به هیچ طبقه‌ای نسبت داده نشده‌اند و لذا باید از یک استراتژی هدفمند مثل استراتژی مبتنی بر یادگیری، برای طبقه‌بندی آنها استفاده شود. برای این منظور مدل نزدیک‌ترین همسایه مرتبه یک پیشنهاد می‌شود که به جای استفاده از استراتژی طبقه‌بندی تصادفی برای طبقه‌بندی نقاط منتسب به چند طبقه، از استراتژی طبقه‌بندی CRSVM استفاده می‌کند. در پایان نشان داده می‌شود که جواب مدل نزدیک‌ترین همسایه مرتبه یک به ازای انواعی از مجموعه‌های عدم قطعیت، به روش تحلیلی تعیین می‌شود.

ایجاد قوانین فازی مرتبه یک به ازای هر بخش

اگر هر بخش، فقط یک مجموعه عدم قطعیت را دربرگیرد فقط قسمت ب-۲ مدل CRSVM تکه‌ای اجرا خواهد شد. حال اگر به ازای هر بخش، قانون فازی مرتبه یک، مطابق با توضیح بند ب-۲-۲ ایجاد شود، همه قوانین فازی ایجادشده از نوع قوانین فازی مرتبه یک خواهند بود. در ادامه نشان داده می‌شود که در صورت انتخاب صحیح توابع عضویت، مدل حاصل، یک مدل نزدیک‌ترین همسایه مرتبه یک است.

تابع تصمیم مدل نزدیک‌ترین همسایه مرتبه یک به ازای داده‌های آموزشی قطعی و داده آزمون x را به صورت (۳۰) تعریف می‌کنیم

$$f(x)=sign(w_{(i,j)}^T x+b_{(i,j)}) \qquad (۳۰)$$

به شرطی که (۳۱) برقرار باشد

$$d(x,x_{(i)})\le \min\{d(x,x_{(k)})\,|\,k\neq i\} \qquad (۳۱)$$

که در آن $j=\arg\min\limits_k\{d(x,x_{(k)})\,|\,y_{(k)}\neq y_{(i)}\}$، $w(i,j)$ و $b(i,j)$ بردار وزن و بایاس ابرصفحه بهینه حاصل از اجرای مدل SVM به ازای مجموعه آموزشی $\{(x_{(i)},y_{(i)}),(x_{(i)},y_{(i)})\}$ است و $d(x,x_{(i)})$ یک معیار اندازه‌گیری فاصله است. این مدل از این جهت اولین نزدیک‌ترین همسایه مرتبه یک خوانده می‌شود که تابع تصمیم آن، حاوی یک چندجمله‌ای درجه یک است.

قضیه ۸: با فرض آن که مجموعه‌های عدم قطعیت، فاقد عضو مشترک باشند و در صورت استفاده از مدل SVM حاشیه سخت با تابع هموارسازی $\|w\|_2^2$ در مدل نزدیک‌ترین همسایه مرتبه یک، خروجی مدل نزدیک‌ترین همسایه مرتبه یک برابر است با مدل نزدیک‌ترین همسایه مرتبه صفر.

تابع تصمیم مدل اولین نزدیک‌ترین همسایه مرتبه یک به ازای داده‌های آموزشی توأم با عدم قطعیت و داده آزمون x برای طبقه‌بندی دوکلاسی به صورت (۳۲) تعریف می‌شود

$$f(x)=sign(w_{(i,j)}^T x+b_{(i,j)}) \qquad (۳۲)$$

به شرطی که (۳۳) برقرار باشد

$$\min_{z\in u_{(i)}} d(x,z)\le \min\{\min_{z\in u_{(k)}} d(x,z)\,|\,k\neq i\} \qquad (۳۳)$$

که $j=\arg\min\limits_k\{\min\limits_{z\in u_{(k)}} d(x,z)\,|\,y_{(k)}\neq y_{(i)}\}$، $w_{(i,j)}$ و $b_{(i,j)}$ بردار وزن و بایاس ابرصفحه بهینه حاصل از اجرای مدل CRSVM به ازای مجموعه آموزشی $\{(u_{(i)},y_{(i)}),(u_{(i)},y_{(i)})\}$، یعنی مدل (۳۴) است

$$\min_{w,b,\xi} \frac{1}{2}\|w\|_2^2 + C(\xi_{(i)}^2 + \xi_{(j)}^2)$$
$$s.t \begin{cases} \min_{x\in u_{(i)}} y_{(i)}(w^T x + b) \geq 1 - \xi_{(i)} \\ \min_{x\in u_{(j)}} y_{(j)}(w^T z + b) \geq 1 - \xi_{(j)} \\ w \neq 0 \end{cases} \quad (۳۴)$$

برای به دست آوردن جواب بهینه این مدل، نیازی به حل عددی آن نیست و می‌توان جواب بهینه این مدل را به روش تحلیلی به دست آورد. با حل تحلیلی مدل CRSVM به ازای یک زوج مجموعه عدم قطعیت، زمان آموزش مدل نزدیک‌ترین همسایه مرتبه یک تقریباً به کلی حذف می‌شود.

۳-۵ خلاصه مدل‌های پیشنهادی

به طور خلاصه مدل‌های طبقه‌بندی پیشنهادی در این مقاله عبارتند از:

- مدل طبقه‌بندی مقاوم بااحتیاط غیر استاندارد CRSVM.
- مدل طبقه‌بندی مقاوم بااحتیاط استانداردشده در فضای ویژگی به ازای مجموعه عدم قطعیت ابردایروی- CRSVM-HC.
- مدل طبقه‌بندی مقاوم بااحتیاط استانداردشده در فضای ویژگی به ازای مجموعه عدم قطعیت ابربیضوی- CRSVM-HE.
- مدل طبقه‌بندی مقاوم بااحتیاط استاندارد تکه‌ای فازی مرتبه صفر- PiCRSVM0.
- مدل طبقه‌بندی مقاوم بااحتیاط استاندارد تکه‌ای فازی مرتبه یک- PiCRSVM1.

٤- آزمایش‌ها و نتایج

سخت‌افزار مورد استفاده برای این آزمایش‌ها شامل CPU Core i۵ با حافظه اصلی ۲ GB و سرعت ۲٫۲ GHz بوده است. مدل‌های SOCP که در این مقاله به آنها اشاره شده است با استفاده از کتابخانه MOSEK [۳۴] در نرم‌افزار MATLAB نسخه ۲۰۱۴a پیاده‌سازی شده‌اند و در این فصل با یکدیگر مقایسه می‌شوند. هر مدل طبقه‌بندی دارای تعدادی پارامتر است که بهترین مقدار آنها به کمک روش اعتبارسنجی ده‌دسته‌ای (۱۰-fold cross) و جستجوی شبکه‌ای[1] تعیین می‌گردد. برای این منظور، بهترین مقدار پارامتر جریمه C از مجموعه $(10^{5},10^{-1},\ldots,10^{-2})$، بهترین مقدار پارامتر عرض کرنل گوسی σ از مجموعه $(2^{5},2^{-1},\ldots,2^{-2})$، بهترین مقدار k در روش k نزدیک‌ترین همسایه از مجموعه $(2,4,\ldots,16)$ و بهترین مقدار $y = y_{(i)}$ در مدل‌های مقاوم از مجموعه $(0,0125,\ldots,1)$ جستجو می‌شود.

برای طبقه‌بندی چندکلاسی با استفاده از مدل‌های مبتنی بر مدل SVM از روش «یک طبقه در برابر بقیه طبقات»[2]، برای سنجش نتایج آزمایش‌ها از روش آماری آزمون فرضیه و برای مقایسه دو الگوریتم بر روی چند مجموعه داده و همچنین مقایسه چند الگوریتم بر روی چند مجموعه داده از روش‌های معرفی‌شده در [۳۵] تا [۳۷] استفاده شده است. در مرحله آزمون فقط از ۳۰٪ هر مجموعه داده آموزشی استفاده گردیده و منظور از زمان آزمون هر مدل طبقه‌بندی، زمان لازم برای طبقه‌بندی این ۳۰٪ داده در مرحله آزمون است. در ادامه از مدل‌های پیشنهادی، اولاً برای کاهش نرخ خطای طبقه‌بندی به ازای داده‌های آموزشی ناقص و ثانیاً برای کاهش زمان آموزش طبقه‌بند به ازای داده‌های آموزشی کامل استفاده شده است.

٤-۱ طبقه‌بندی داده‌های ناقص

برای طبقه‌بندی داده‌های ناقص از روش انتساب ساده و چندگانه که در آن ویژگی‌های نامعلوم داده‌های ناقص تخمین زده می‌شود و همچنین روش جایگذاری‌شده استفاده گردیده است. انتساب میانگین، پایه‌ای‌ترین روش انتساب ساده برای ویژگی‌های عددی یا پیوسته است. در این روش، ویژگی نامعلوم داده ناقص با میانگین مقدار ویژگی مربوط در داده‌های کامل هم طبقه داده ناقص مقداردهی می‌شود. انتساب مُد[3] نیز پایه‌ای‌ترین پایه‌ای‌ترین روش انتساب ساده برای ویژگی‌های اسمی[4] یا طبقه‌ای[5] است. است. در این روش، ویژگی نامعلوم داده ناقص با مقداری که بیشترین تکرار را در داده‌های کامل هم‌طبقه داده ناقص دارد مقداردهی می‌شود [۳] و [۳۸]. به ویژگی نامعلوم، مقدار صفر نسبت داده شده است. استفاده از میانگین یا مد k تا از نزدیک‌ترین همسایگان، روش دیگری برای انتساب ویژگی نامعلوم می‌باشد [۳].

در روش انتساب ساده EM[6] [۳۹]، داده ناقص، یک متغیر تصادفی فرض می‌شود که میانگین و کواریانس آن به روش تکراری EM حاصل می‌گردد و نهایتاً مقدار میانگین حاصل‌شده به عنوان مقدار تخمینی ویژگی‌های نامعلوم داده ناقص استفاده می‌شود. ایراد روش EM آن است که فقط در فضای ورودی قابل اعمال است. در ضمن، روش EM به ازای هر داده ناقص، یک بار اجرا می‌شود. زمان اجرای هر تکرار از روش تکراری EM به بُعد داده بستگی دارد، لذا در صورتی که بعد داده و تعداد داده‌های ناقص زیاد باشد روش زمان‌بری خواهد شد.

در انتساب چندگانه، برای ویژگی نامعلوم، یک مجموعه گسسته یا پیوسته تخمین زده می‌شود و ویژگی نامعلوم به نوعی مجموعه تبدیل می‌شود [۴۰]. در [۹] هر ویژگی نامعلوم با بازه همان ویژگی در ویژگی‌های معلوم داده‌های هم‌طبقه داده ناقص جایگذاری می‌شود. به تعبیری می‌توان گفت که در این روش، هر داده با یک ابرمستطیل جایگزین می‌شود که هر ضلع آن، بازه یک ویژگی نامعلوم را مشخص می‌کند.

در روش جاسازی‌شده[7] [۳]، مدل (۳۵) برای طبقه‌بندی داده‌های ناقص ناقص پیشنهاد شده است

$$\max_{w,b} \min_{i} \frac{y_{(i)}(w^{(i)^T} x_{(i)}^a + b)}{\|w^{(i)}\|_2} \quad (۳۵)$$

که در آن $x_{(i)}^a$ بردار متشکل از مؤلفه‌های معلوم $x_{(i)}$ و $w^{(i)}$ مؤلفه‌هایی از بردار وزن متناظر با مؤلفه‌های معلوم $x_{(i)}$ می‌باشد. مدل فوق یک مدل غیر استاندارد است که یک روش تکراری برای حل دوگان آن پیشنهاد شده که در هر تکرار آن، یک مدل درجه دوم حل می‌شود. البته همان طور که در [۳] نیز اشاره شده است تضمینی برای همگرایی این الگوریتم به نقطه بهینه مدل (۳۴) وجود ندارد.

1. Grid Search
2. Versus-Rest
3. Mod
4. Nominal
5. Categorical
6. Expectation Maximization
7. Embedded

جدول ۱: مشخصات تعدادی از مجموعه داده‌های ناقص مخزن دانشگاه آیروین کالیفرنیا.

ردیف	نام مجموعه داده	تعداد داده‌ها	تعداد ویژگی‌ها	تعداد داده‌های ناقص	درصد داده‌های ناقص	تعداد طبقات داده
۱	Breast Cancer Wisconsin (B.C.W.)	۶۹۹	۱۰	۱۶	۲٫۲۹	۲
۲	Census-Income (KDD)	۲۹۹۲۸۵	۴۰	۱۵۶۷۶۴	۵۲٫۳۸	۲
۳	Credit Approval (C.A.)	۶۹۰	۱۵	۳۷	۵٫۳۶	۲
۴	Heart disease (Cleveland)	۳۰۳	۱۳	۶	۱٫۹۸	۵
۵	Heart disease (Hungarian)	۲۹۴	۱۳	۲۹۳	۹۹٫۶۶	۲
۶	Heart disease (Switzerland)	۱۲۳	۱۳	۱۲۳	۱۰۰	۵
۷	Heart disease (Long Beach VA)	۲۰۰	۱۳	۱۹۹	۹۹٫۵۰	۵
۸	Hepatitis	۱۵۵	۱۹	۷۵	۴۸٫۳۹	۲
۹	Pittsburgh bridges (Material)	۱۰۸	۱۳	۳۸	۳۵٫۱۹	۳
۱۰	Pittsburgh bridges (REL-L)	۱۰۸	۱۳	۳۸	۳۵٫۱۹	۳
۱۱	Pittsburgh bridges (SPAN)	۱۰۸	۱۳	۳۸	۳۵٫۱۹	۳
۱۲	Pittsburgh bridges (T-OR-D)	۱۰۸	۱۳	۳۸	۳۵٫۱۹	۲
۱۳	Pittsburgh bridges (TYPE)	۱۰۸	۱۳	۳۸	۳۵٫۱۹	۶
۱۴	Postoperative Patient (P.P.)	۹۰	۸	۳	۳٫۳۳	۳
۱۵	SECOM	۱۵۶۷	۵۹۱	۱۵۶۷	۱۰۰	۲
۱۶	Thyroid Disease (Allhypo)	۳۷۷۲	۲۹	۳۷۷۲	۱۰۰	۵
۱۷	Thyroid Disease (Sick)	۳۵۳۵	۲۹	۳۵۳۵	۱۰۰	۲

٤-۱-۱ مدل‌های استفاده‌شده در آزمایش‌های داده‌های ناقص

در آزمایش‌ها از مدل‌های زیر برای طبقه‌بندی داده‌های ناقص استفاده شده و با یکدیگر مقایسه می‌شوند:

۱) روش جاسازی‌شده معرفی‌شده در [۳]، (رابطه (۳۵)).

۲) انتساب ساده مُد یا میانگین $k-$نزدیک‌ترین همسایه و سپس استفاده از مدل DSVM برای یادگیری طبقه‌بند که به طور خلاصه MK[1]+DSVM نامیده می‌شود.

۳) استفاده از روش انتساب ساده EM برای تخمین مقدار ویژگی‌های نامعلوم داده‌های ناقص و سپس استفاده از مدل DSVM برای یادگیری طبقه‌بند. این روش EM+DSVM نامیده می‌شود.

۴) انتساب چندگانه با تبدیل هر داده ناقص به یک مجموعه ابربیضوی در فضای ویژگی (تخمین مرکز و کواریانس ابربیضی بر اساس $k-$ نزدیک‌ترین همسایه) و سپس استفاده از مدل ARSVM-HE [۴۱] برای یادگیری طبقه‌بند که MIK[2]+ARSVM-HE نامیده می‌شود.

۵) انتساب چندگانه با تبدیل هر داده ناقص به یک مجموعه ابربیضوی در فضای ویژگی (تخمین مرکز و کواریانس ابربیضی بر اساس $k-$ نزدیک‌ترین همسایه) و و سپس استفاده از مدل پیشنهادی CRSVM-HE برای یادگیری طبقه‌بند. این روش MIK+CRSVM-HE نامیده می‌شود.

۶) انتساب چندگانه با تبدیل هر داده ناقص به یک مجموعه ابردایروی در فضای ویژگی (تخمین مرکز و کواریانس ابربیضی بر اساس $k-$ نزدیک‌ترین همسایه) و سپس استفاده از مدل پیشنهادی CRSVM-HC برای یادگیری طبقه‌بند. این روش MIK+CRSVM-HC نامیده می‌شود.

۷) انتساب چندگانه با تبدیل هر داده ناقص به یک مجموعه ابردایروی در فضای ویژگی (تخمین مرکز و شعاع ابردایره بر اساس $k-$ نزدیک‌ترین همسایه) و سپس استفاده از مدل تکه‌ای پیشنهادی PiCRSVM0 در فضای ویژگی برای یادگیری طبقه‌بند که MIK+PiCRSVM0-HC نامیده می‌شود.

۸) انتساب چندگانه با تبدیل هر داده ناقص به یک مجموعه ابربیضوی در فضای ورودی (تخمین مرکز و کواریانس ابربیضی به روش EM) و سپس استفاده از مدل تکه‌ای پیشنهادی PiCRSVM0 در فضای ورودی برای یادگیری طبقه‌بند که MEM[3]+PiCRSVM0-HE نامیده می‌شود.

٤-۱-۲ مجموعه داده‌های ناقص استفاده‌شده در آزمایش‌ها

جدول ۱ مشخصات تعدادی از مجموعه‌های داده‌های آموزشی ناقص مخزن دانشگاه آیروین کالیفرنیا[4] را نشان می‌دهد. در هر مجموعه ناقص، تعدادی از ویژگی‌های برخی از داده‌ها، از دست رفته یا نامعلوم است. یکی از مسایل مطرح در زمینه طبقه‌بندی، چگونگی یادگیری طبقه‌بند بر اساس چنین مجموعه آموزشی‌ای است. تعداد داده‌های برخی از این مجموعه‌ها بسیار زیاد است، لذا با توجه به محدودیت زمان و سخت‌افزار در هر آزمایش حداکثر از ۱۰۰۰ داده از هر مجموعه آموزشی استفاده گردید. این ۱۰۰۰ داده به صورت تصادفی و با حفظ نسبت داده‌های هر طبقه به کل داده‌ها و نسبت داده‌های ناقص به داده‌های کامل انتخاب شدند.

٤-۱-۳ نتایج آزمایش‌های طبقه‌بندی مجموعه داده‌های ناقص

٤-۱-۳-۱ نرخ خطا

جدول ۲ نرخ خطای مدل‌های مختلف طبقه‌بندی به ازای هر یک از مجموعه‌های داده یادشده را نشان می‌دهد. رتبه هر مدل به ازای هر مجموعه داده در پرانتز مشخص شده است. استفاده از میانگین نرخ خطا برای مقایسه کارایی چند طبقه‌بند بر روی بیش از دو مجموعه داده، نادرست است [۳۶] لذا از روش رتبه‌بندی طبقه‌بندها استفاده می‌گردد [۴۲]. همچنین با توجه به عدم استقلال آزمایش‌ها در روش اعتبارسنجی ۱۰-، باید از روش‌های آماری غیر پارامتری جهت انجام آزمون فرضیه استفاده کرد [۳۶]. یکی از روش‌های مطرح غیر پارامتری جهت انجام

1. Mean or Mod of K-Nearest Neighbors
2. Multiple Imputation Based on K-Nearest Neighbors
3. Multiple Imputation Based on EM
4. UCI Repository of Machine Learning Databases, Available: http://www.ics.uci.edu/mlearn/MLRepository.html, 2014

جدول ۲: میانگین نرخ خطای مدل‌های طبقه‌بندی به ازای داده‌های آموزشی ناقص.

نام مجموعه داده	Embedded [۳]	MK+ DSVM	EM+ DSVM	MIK+ ARSVM-HE	MIK+ CRSVM-HE	MIK+ CRSVM-HC	MIK+ PiCRSVM0-HC	MEM+ PiCRSVM0-HE
B.C.W.	۶٫۹۱(۱۳)	۷٫۲۱(۱۴)	۶٫۱۷(۱۲)	۵٫۱۸(۸)	۵٫۰۱(۹)	۵٫۲۱(۱۰)	۵٫۴۱(۱۱)	۵٫۶۴(۱۱)
KDD	۱۰٫۷۳(۱۳)	۱۲٫۱۵(۱۴)	۸٫۴۹(۱۰)	۱۰٫۴۸(۱۲)	۵٫۱۲(۵)	۵٫۳(۱۱)	۱۰٫۲۱(۹)	۷٫۱۲(۹)
C.A.	۵٫۱۹(۱۰)	۵٫۱۲(۶)	۵٫۲۱(۱۲)	۶٫۰۹(۱۴)	۵٫۰۱(۶)	۵٫۱۲(۳)	۵٫۰۵(۱۰)	۵٫۱۹(۱۰)
Cleveland	۲۱٫۱۲(۲)	۲۳٫۲۵(۱۴)	۲۱٫۳۱(۳)	۲۲٫۲۳(۱۱)	۲۱٫۰۴(۹)	۲۲٫۰۲(۱۱)	۲۲٫۲۳(۱۰)	۲۲٫۱۷(۱۰)
Hungarian	۲۳٫۴۲(۶)	۲۱٫۰۴(۱)	۲۳٫۷۱(۹)	۲۲٫۲۷(۳)	۲۲٫۱۳(۴)	۲۲٫۹۵(۱۳)	۲۴٫۷(۱۴)	۲۴٫۸۵(۱۴)
Switzerland	۱۸٫۵۱(۳)	۱۹٫۰۹(۱۳)	۱۸٫۹۷(۱۲)	۱۹٫۱۷(۱۴)	۱۸٫۲۸(۴)	۱۸٫۵۳(۱)	۱۸٫۱۱(۵)	۱۸٫۷۲(۵)
L.B.VA	۲۲٫۳۶(۱۲)	۲۲٫۳۹(۱۳)	۲۲٫۰۱(۱۱)	۲۰٫۱۷(۴)	۲۰٫۳۴(۶)	۲۰٫۵۱(۱۴)	۲۲٫۸۲(۱۰)	۲۱٫۹۵(۱۰)
Hepatitis	۱۷٫۴۲(۱۱)	۱۵٫۷۲(۱)	۱۵٫۷۶(۲)	۱۶٫۵۲(۵)	۱۶٫۴۳(۱۰)	۱۶٫۹۹(۴)	۱۶٫۴۹(۶)	۱۶٫۵۵(۶)
Material	۱۵٫۹(۱۴)	۱۳٫۹۷(۲)	۱۳٫۲۶(۱)	۱۴٫۲۷(۵)	۱۴٫۲۶(۳)	۱۴٫۲۲(۶)	۱۵٫۲۷(۷)	۱۵٫۲۹(۷)
REL-L	۳۱٫۶۲(۲)	۳۷٫۴۸(۱۱)	۳۹٫۹۶(۱۳)	۳۶٫۸۱(۰)	۳۱٫۶۴(۱)	۳۱٫۵۹(۱۲)	۳۹٫۱۶(۱۴)	۴۱٫۱۳(۱۴)
SPAN	۳۹٫۴۳(۱۱)	۳۸٫۸۶(۸)	۳۹٫۷۳(۱۲)	۳۹٫۰۲(۱۰)	۳۷٫۶۷(۵)	۳۸٫۸۱(۱۳)	۴۰٫۲۹(۱۴)	۴۱٫۰۶(۱۴)
T-OR-D	۱۴٫۷۱(۱۰)	۱۴٫۹۳(۱۲)	۱۵٫۸(۱۴)	۱۳٫۵۳(۱)	۱۳٫۸۲(۲)	۱۳٫۸(۱۳)	۱۵٫۱۱(۴)	۱۴٫۳(۴)
TYPE	۴۸٫۵۳(۱)	۵۲٫۱۶(۱۳)	۴۹٫۸۲(۵)	۵۱٫۹۱(۱۲)	۵۱٫۰۹(۱۴)	۵۲٫۳۴(۱۰)	۵۱٫۱۵(۱۱)	۵۱٫۱۷(۱۱)
P.P.	۴۷٫۳(۵)	۴۷٫۱۲(۸)	۴۷٫۹۳(۱۱)	۴۹٫۲۸(۱۴)	۴۶٫۳۱(۱۳)	۴۸٫۶(۵)	۴۷٫۰۳(۱۲)	۴۸٫۱۵(۱۲)
SECOM	۳۷٫۵۳(۱۴)	۳۳٫۰۷(۱)	۳۶٫۱۸(۷)	۳۵٫۱۸(۴)	۳۴٫۹۶(۲)	۳۴٫۹۲(۵)	۳۵٫۹۸(۶)	۳۶٫۱۲(۶)
Allhypeo	۳٫۵۱(۶)	۳٫۰۹(۴)	۳٫۶۲(۷)	۳٫۷۹(۸)	۲٫۵۷(۳)	۲٫۸۷(۱)	۲٫۳۹(۵)	۳٫۴۵(۵)
Sick	۵٫۹۷(۱۲)	۵٫۷۶(۹)	۵٫۷۹(۱۰)	۶٫۰۸(۱۴)	۴٫۹۶(۷)	۵٫۱۵(۱۰)	۵٫۷۹(۱۳)	۵٫۹۹(۱۳)
میانگین خطای طبقه‌بند	۲۱٫۷۶	۲۱٫۹۱	۲۱٫۹۸	۲۱٫۸۸	۲۰٫۶۳	۲۱٫۱۱	۲۲٫۱۹	۲۲٫۲۹
میانگین رتبه طبقه‌بند	۸٫۵۳	۸٫۴۷	۸٫۸۸	۸٫۷۶	۲٫۷۱	۶٫۰۶	۸٫۳۵	۹٫۴۷

آزمون فرضیه در مقایسه کارایی چند طبقه‌بند بر روی بیش از دو مجموعه داده، آزمون Friedman است [۴۳] که بدین شرح است: فرض کنید $\mathbb{K}$ تعداد مدل‌های مورد مقایسه و $\mathbb{N}$ تعداد مجموعه داده‌ها باشد که در آزمایش اخیر مقادیر $\mathbb{K}$ و $\mathbb{N}$ به ترتیب ۱۵ و ۱۷ است. همچنین فرض کنید که $r_{(i)}^{(j)}$ بیانگر رتبه $j-$امین مدل بر روی مجموعه داده شماره i باشد. آزمون Friedman میانگین رتبه‌ها را بر اساس $R_j = (1/\mathbb{N})\sum_j r_{(i)}^{(j)}$ محاسبه کرده و ملاک مقایسه قرار می‌دهد. برای انجام آزمون آماری فرضیه، دو فرضیه زیر در نظر گرفته می‌شود:

$H_{(\cdot)}$: تمام مدل‌های استفاده‌شده در آزمایش‌ها از نظر کارایی با یکدیگر یکسان هستند.

$H_{(1)}$: تمام مدل‌های استفاده‌شده در آزمایش‌ها از نظر کارایی با یکدیگر یکسان نیستند.

آماره Friedman به صورت (۳۶) می‌باشد

$$x_F^2 = \frac{12\times\mathbb{N}}{\mathbb{K}(\mathbb{K}+1)}\left(\sum_j R_{(j)}^2 - \frac{\mathbb{K}(\mathbb{K}+1)^2}{4}\right) \qquad (۳۶)$$

مقدار آماره فوق برای آزمایش مزبور برابر است با

$$x_F^2 = \frac{12\times17}{14\times15}(804/55 - \frac{14\times225}{4}) = 16/5$$

بنابراین چون مقدار محاسبه‌شده آماره x_F^2 بزرگ‌تر از ۱۵٫۹۸ (مقدار بحرانی توزیع chi-square با درجه آزادی $\mathbb{K}-1=13$) است و فرضیه $H_{(\cdot)}$ با سطح معنادار $\alpha = 0/25$ یا با درجه اطمینان ۷۵٪ رد می‌شود.

حال برای تحلیل بیشتر و تعیین بهترین مدل از روش‌های Post-hoc می‌توان استفاده نمود. برای این منظور در روش Friedman می‌توان از آماره Z یعنی (۳۷) استفاده کرد

$$z = (R_{(i)} - R_{(j)})\sqrt{\frac{6}{\mathbb{K}(\mathbb{K}+1)}} \qquad (۳۷)$$

در رابطه فوق $R_{(i)}$ و $R_{(j)}$ به ترتیب مقدار میانگین رتبه دو مدل i و j است. روال آماری مورد استفاده، Finner نام دارد [۴۴]. در این روال، طی یک رویه گام به گام، مقدار سطح معنادار α تنظیم می‌گردد.

در جدول ۳ روال Finner برای مقایسه MIK+CRSVM-HE با سایر مدل‌ها با درجه اطمینان ۹۹٪ نشان داده شده است. با توجه به این که تمام مقادیر $p-$value از مقادیر سطح معنادار α تنظیم‌شده در ستون پنجم جدول، کمتر است بنابراین به لحاظ آماری اثبات می‌شود که مدل MIK+CRSVM-HE از نظر نرخ خطای طبقه‌بندی کمترین رتبه را داراست.

نتایج زیر از جدول ۳ به دست می‌آید:

– از جنبه نرخ خطا، مدل بااحتیاط MIK+CRSVM-HE دارای کمترین میانگین رتبه نرخ خطا بوده و پس از آن مدل‌های بااحتیاط MIK+CRSVM-HC و MIK+PiCRSVM0-HC به ترتیب در جایگاه دوم و سوم کمترین رتبه نرخ خطا را نسبت به بقیه مدل‌ها دارند.

– برای توجیه نرخ خطای خوب مدل بااحتیاط MIK+CRSVM-HC یادآوری می‌شود که یک ابرصفحه یا ابردایره در فضای ویژگی به خوبی می‌تواند داده‌ها را توصیف کند و آزمایش‌ها نشان می‌دهد که تقریباً هر نوع توصیف داده با ابربیضی را می‌توان با استفاده از ابردایره یا ابرصفحه در فضای ویژگی و تنظیم دقیق پارامترهای تابع کرنلی چون کرنل گوسی انجام داد [۴۵] و [۴۶].

– میانگین رتبه نرخ خطای مدل تکه‌ای بااحتیاط پیشنهادی MEM+PiCRSVM0-HE بدتر از میانگین رتبه نرخ خطای مدل بااحتیاط تکه‌ای پیشنهادی MIK+PiCRSVM0-HC است، چرا که مدل MEM+PiCRSVM0-HE در فضای ورودی و مدل MIK+PiCRSVM0-HC در فضای ویژگی کار می‌کند.

جدول ۳: روال FINNER برای تأیید مدل MIK+CRSVM-HC به عنوان بهترین مدل طبقه‌بندی از نظر نرخ خطا.

i	مدل طبقه‌بندی	مقدار آماره Z	$p-\text{value}$	$۱-(۱-\alpha)^{\frac{K-۱}{i}}$
۱	MEM+PiCRSVM0-HE	۶٫۵۱	۰٫۰۰۰۰	۰٫۱۲۲۵
۲	EM+DSVM	۵٫۹۵	۰٫۰۰۰۰	۰٫۰۶۳۲
۳	MIK+ARSVM-HE	۵٫۸۳	۰٫۰۰۰۰	۰٫۰۴۲۶
۴	Embedded	۵٫۶۱	۰٫۰۰۰۰	۰٫۰۳۲۱
۵	MK+DSVM	۵٫۵۵	۰٫۰۰۰۰	۰٫۰۲۵۸
۶	MIK+PiCRSVM0-HC	۵٫۴۴	۰٫۰۰۰۰	۰٫۰۲۱۵
۷	MIK+CRSVM-HC	۳٫۲۳	۰٫۰۰۱۲	۰٫۰۱۰۰

جدول ۴: میانگین زمان آموزش مدل‌های طبقه‌بندی مختلف به ازای هر یک از مجموعه‌های آموزشی (بر حسب ثانیه).

نام مجموعه داده	Embedded	MK+ DSVM	EM+ DSVM	MIK+ ARSVM-HE	MIK+ CRSVM-HE	MIK+ CRSVM-HC	MIK+ PiCRSVM0-HC	MEM+ PiCRSVM0-HE
B.C.W.	۱۱۶٫۸۳(۱۰)	۱۶٫۳۴(۳)	۱۶٫۵(۴)	۶۳٫۰۷(۷)	۹۹۲٫۶۹(۱۳)	۶۳٫۲۴(۸)	۸٫۱۷(۲)	۲٫۲۶(۱)
KDD	۴۳۴٫۰۸(۱۰)	۵۳٫۵۹(۲)	۱۳۷٫۴(۵)	۲۲۷٫۲۲(۸)	۴۶۴۳٫۵۸(۱۳)	۲۳۸٫۰۸(۹)	۲۷٫۹۷(۱)	۹۰٫۰۶(۴)
C.A.	۱۴۹٫۴۷(۱۰)	۱۷٫۲۸(۳)	۱۸٫۱۱(۴)	۷۷٫۰۷(۸)	۱۲۵۵۲٫۰۴(۱۳)	۸۲٫۷۶(۹)	۹٫۲۴(۲)	۲٫۷۵(۱)
Cleveland	۴۷٫۲۳(۱۰)	۶٫۳۴(۳)	۶٫۴۴(۴)	۲۵٫۲۳(۷)	۱۸۵۷٫۱۸(۱۳)	۲۵٫۶۶(۹)	۳٫۲۱(۲)	۰٫۸۹(۱)
Hungarian	۴۱٫۸۸(۱۰)	۶٫۰۷(۳)	۱۱٫۰۲(۵)	۲۲٫۸۲(۸)	۳۷۶۶٫۳۹(۱۳)	۲۲٫۶۱(۶)	۳(۱)	۵٫۷۵(۲)
Switzerland	۸٫۸۶(۱۰)	۱٫۰۳(۲)	۳٫۱۱(۵)	۴٫۵۷(۸)	۲۷۸٫۹۲(۱۳)	۴٫۹(۹)	۰٫۵۵(۱)	۲٫۱۹(۴)
L.B.VA	۱۸٫۵۱(۱۰)	۲٫۱۴(۲)	۵٫۵(۵)	۹٫۵۴(۸)	۸۴٫۵۸(۱۳)	۱۰٫۲۵(۹)	۱٫۱۴(۱)	۳٫۶(۴)
Hepatitis	۱۳٫۸۳(۱۰)	۱٫۵۲(۲)	۴٫۲۳(۵)	۷٫۰۵(۸)	۷۸۳٫۸۵(۱۳)	۷٫۷۲(۹)	۰٫۸۳(۱)	۲٫۸۷(۴)
Material	۸٫۴۵(۱۰)	۱٫۰۵(۳)	۱٫۶۹(۵)	۴٫۴۳(۸)	۳۳٫۸۸(۱۳)	۴٫۶۳(۹)	۰٫۵۵(۱)	۰٫۷۷(۲)
REL-L	۹٫۲۶(۱۰)	۱٫۲۱(۳)	۱٫۸۵(۵)	۴٫۹۱(۸)	۵۴٫۸۸(۱۳)	۵٫۰۴(۹)	۰٫۶۲(۱)	۰٫۷۹(۲)
SPAN	۱۱٫۲۹(۱۰)	۱٫۴۲(۳)	۲٫۰۶(۵)	۵٫۹۴(۸)	۷۱٫۸۴(۱۳)	۶٫۱۸(۹)	۰٫۷۴(۱)	۰٫۸۱(۲)
T-OR-D	۹٫۱۵(۱۰)	۱٫۰۷(۳)	۱٫۷۱(۵)	۴٫۷۳(۸)	۴۶۳٫۰۳(۱۳)	۵٫۰۶(۹)	۰٫۵۷(۱)	۰٫۷۹(۲)
TYPE	۱۱٫۱۹(۱۰)	۱٫۶۱(۳)	۲٫۲۵(۵)	۶٫۰۹(۸)	۲۳۷٫۹۲(۱۳)	۶٫۰۴(۶)	۰٫۸(۱)	۰٫۸۵(۲)
P.P.	۴٫۴۳(۱۰)	۰٫۶۲(۳)	۰٫۶۴(۴)	۲٫۳۹(۷)	۹۹٫۴(۱۳)	۲٫۴(۸)	۰٫۳۱(۲)	۰٫۱(۱)
SECOM	۳۵۶٫۵۵(۸)	۴۵٫۴۲(۲)	۳۴۹۷۳٫۵۲(۱۴)	۱۸۸٫۰۴(۶)	۳۶۰۱٫۲۱(۱۱)	۱۹۴٫۸(۷)	۲۳٫۴۵(۱)	۳۴۹۳۳٫۵۳(۱۳)
Allhypeo	۵۸۷٫۵۷(۱۰)	۷۲٫۵۴(۲)	۱۵۶٫۶۴(۵)	۳۰۷٫۵۷(۸)	۷۱۱۱۳٫۴۴(۱۳)	۳۲۲٫۲۷(۹)	۳۷٫۸۷(۱)	۹۲٫۵۶(۴)
Sick	۵۹۸٫۰۷(۱۰)	۸۷٫۳۱(۲)	۱۷۱٫۴۱(۵)	۳۲۶٫۵۴(۸)	۲۸۸۶٫۴۶(۱۳)	۳۲۲٫۷۷(۶)	۴۳٫۰۴(۱)	۹۵٫۶۹(۳)
میانگین زمان آموزش	۱۴۲٫۷۴	۱۸٫۶۲	۲۰۸۹٫۰۶	۷۵٫۷۲	۶۰۸۹٫۴۹	۷۷٫۹۱	۹٫۵۳	۲۰۷۲٫۷۲
میانگین رتبه	۹٫۸۸	۲٫۵۹	۵٫۲۹	۷٫۷۱	۱۲٫۸۸	۸٫۲۴	۱٫۲۴	۳٫۰۶

- میانگین رتبه نرخ خطای مدل‌های بااحتیاط تکه‌ای پیشنهادی MIK+PiCRSVM0-HC و MEM+PiCRSVM0-HE بدتر از میانگین رتبه نرخ خطای مدل‌های بااحتیاط غیر تکه‌ای پیشنهادی MIK+CRSVM-HC و MIK+CRSVM-HE است چرا که مدل‌های تکه‌ای، نگاهی محلی به داده‌ها دارند (هر تکه از طبقه‌بند با استفاده از بخشی از داده‌های آموزشی ایجاد می‌شود نه با استفاده از همه داده‌های آموزشی) و تبعاً به جواب بهینه محلی خواهند رسید.

٤-۱-۳-۲ زمان آموزش

جدول ۴ میانگین زمان آموزش مدل‌های طبقه‌بندی مختلف را به ازای هر یک از مجموعه‌های آموزشی بر حسب ثانیه نشان می‌دهد. نتایج زیر از جدول ۴ به دست می‌آید:

- میانگین رتبه زمان آموزش مدل بااحتیاط تکه‌ای MEM+PiCRSVM0-HE بهتر از میانگین رتبه زمان آموزش مدل بااحتیاط غیر تکه‌ای MIK+CRSVM-HE است.
- میانگین رتبه زمان آموزش مدل بااحتیاط تکه‌ای MIK+PiCRSVM0-HC بهتر از میانگین رتبه زمان آموزش مدل بااحتیاط غیر تکه‌ای MIK+CRSVM-HC است. قبلاً پیچیدگی محاسباتی یک مدل تکه‌ای با پیچیدگی محاسباتی مدل غیر تکه‌ای مقایسه شد. از این رو تفاوت زمان آموزش مدل غیر تکه‌ای و مدل تکه‌ای قابل توجیه است.
- میانگین رتبه زمان آموزش مدل CRSVM-HE بدتر از میانگین رتبه زمان آموزش مدل CRSVM-HC و RSVM-HE است و میانگین رتبه زمان آموزش مدل CRSVM-HC بدتر از میانگین رتبه زمان مدل RSVM-HC است. هر یک از این نتایج، ناشی از تفاوت این مدل‌ها از نظر تعداد متغیرها و قیود است.

٤-۱-۳-۳ تعداد بردارهای پشتیبان، تعداد تقسیمات فضا

زمان آزمون مدل‌های مبتنی بر SVM به تعداد بردارهای پشتیبان (تعداد ضرایب غیر صفر بردار وزن) و زمان آزمون مدل‌های تکه‌ای به تعداد تقسیمات فضا بستگی دارد. این نسبت‌ها در جدول ۵ نشان داده شده است. نتایج زیر از جدول ۵ به دست می‌آید:

- درصد بردارهای پشتیبان مدل‌های Embedded، MK+DSVM و EM+DSVM، معمولاً بسیار کمتر از درصد بردارهای پشتیبان مدل‌های

جدول ۵: نسبت تعداد بردارهای پشتیبان به تعداد داده‌های آموزشی استفاده‌شده در بسط بردار وزن در مدل‌های مبتنی بر SVM یا نسبت تعداد تقسیمات فضا به تعداد داده‌های آموزشی در مدل‌های تکه‌ای.

نام مجموعه داده	Embedded	MK+ DSVM	EM+ DSVM	MIK+ ARSVM-HE	MIK+ CRSVM-HE	MIK+ CRSVM-HC	MIK+ PiCRSVM0-HC	MEM+ PiCRSVM0-HE
B.C.W.	۱۵٫۵۲(۷)	۱۴٫۰۷(۴)	۱۰٫۵۲(۱)	۹۹٫۱۲(۱۰)	۱۰۰(۱۲)	۱۰۰(۱۲)	۱۱٫۶۵(۳)	۱۴٫۹۵(۶)
KDD	۱۱٫۴۳(۶)	۷٫۱۳(۲)	۶٫۴۳(۱)	۱۰۰(۱۱)	۹۹٫۸۳(۱۰)	۱۰۰(۱۱)	۷٫۹۹(۳)	۱۳٫۱۹(۷)
C.A.	۱۲٫۳۷(۶)	۶٫۴۲(۱)	۷٫۳۷(۳)	۱۰۰(۱۲)	۱۰۰(۱۲)	۹۹٫۹۹(۱۱)	۶٫۵۲(۲)	۱۲٫۸۲(۷)
Cleveland	۴۶٫۲۸(۷)	۴۳٫۹۳(۵)	۴۱٫۲۸(۳)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۶٫۹۱(۱)	۲۰٫۸۱(۲)
Hungarian	۵۳٫۴۴(۷)	۵۲٫۷۴(۶)	۴۸٫۴۴(۴)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۹٫۹۹(۱)	۲۲٫۷۹(۲)
Switzerland	۳۶٫۸۱(۷)	۳۱٫۰۱(۳)	۳۱٫۸۱(۴)	۱۰۰(۱۲)	۹۹٫۹۹(۱۱)	۹۹٫۸۶(۱۰)	۱۱٫۵۶(۱)	۱۷٫۷۶(۲)
L.B.VA	۴۷٫۸۴(۷)	۴۱٫۸۹(۳)	۴۲٫۸۴(۴)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۳٫۶۲(۱)	۱۹٫۹۲(۲)
Hepatitis	۵۵٫۷۳(۶)	۴۸٫۴۳(۳)	۵۰٫۷۳(۴)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۳٫۸۵(۱)	۲۱٫۰۵(۲)
Material	۲۷٫۴(۷)	۲۳٫۲۵(۴)	۲۲٫۴(۳)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۱٫۳۳(۱)	۱۶٫۴۳(۲)
REL-L	۴۴٫۴(۷)	۴۱٫۴۵(۵)	۳۹٫۴(۳)	۹۹٫۳۸(۱۱)	۹۹٫۸۲(۱۲)	۹۹٫۲۶(۱۰)	۱۵٫۹۳(۱)	۲۰٫۲۳(۲)
SPAN	۴۱٫۱۶(۷)	۳۷٫۳۱(۴)	۳۶٫۱۶(۳)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۴٫۳۸(۱)	۱۹٫۲۸(۲)
T-OR-D	۲۲٫۹(۷)	۱۷٫۲۵(۳)	۱۷٫۹(۴)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۰۰(۱۰)	۸٫۹۳(۱)	۱۵٫۰۳(۲)
TYPE	۴۴٫۹(۷)	۴۴٫۰۵(۶)	۳۹٫۹(۴)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۰۰(۱۱)	۱۸٫۱۳(۱)	۲۱٫۰۳(۲)
P.P.	۵۹٫۲۱(۷)	۵۷٫۷۶(۵)	۵۴٫۲۱(۳)	۱۰۰(۱۳)	۹۹٫۷۲(۱۰)	۹۹٫۹۷(۱۲)	۲۰٫۳۹(۱)	۲۳٫۶۹(۲)
SECOM	۱۲٫۰۷(۶)	۸٫۵۲(۲)	۷٫۰۷(۱)	۹۹٫۹۹(۱۰)	۱۰۰(۱۲)	۹۹٫۹۹(۱۰)	۸٫۸۶(۳)	۱۳٫۵۶(۷)
Allhypeo	۱۱٫۱۴(۶)	۶٫۸۴(۲)	۶٫۱۴(۱)	۱۰۰(۱۰)	۱۰۰(۱۰)	۱۰۰(۱۰)	۷٫۹۳(۳)	۱۳٫۱۳(۷)
Sick	۱۷٫۵۶(۷)	۱۷٫۰۱(۶)	۱۲٫۵۶(۲)	۱۰۰(۱۲)	۱۰۰(۱۲)	۱۰۰(۱۲)	۱۲٫۹۶(۳)	۱۵٫۶۶(۴)
میانگین نسبت	۳۲٫۹۵	۲۹٫۳۶	۲۷٫۹۵	۹۹٫۹۱	۹۹٫۹۶	۹۹٫۹۵	۱۳٫۰۰	۱۷٫۷۳
میانگین رتبه	۶٫۷۱	۳٫۷۶	۲٫۸۲	۱۰٫۸۸	۱۰٫۸۸	۱۰٫۷۱	۱٫۶۵	۳٫۵۳

MIK+ARSVM-HE، MIK+CRSVM-HE و MIK+CRSVM-HC است. این موضوع قابل پیش‌بینی بود چرا که برخلاف مدل‌های ARSVM-HE، CRSVM-HE و CRSVM-HC، تنک‌شدن بردار وزن مدل‌های دوگانی چون DSVM تضمین شده است.

– کمترین درصد بردارهای پشتیبان متعلق به مدل بااحتیاط تکه‌ای پیشنهادی MIK+PiCRSVM0-HC است.

– درصد بردارهای پشتیبان مدل‌های بااحتیاط تکه‌ای پیشنهادی کمتر از مدل‌های بااحتیاط غیر تکه‌ای است.

٤-۳-۱-٤ زمان آزمون

جدول ۶ میانگین زمان آزمون مدل‌های مختلف را نشان می‌دهد. نتایج زیر از جدول ۶ به دست می‌آید:

– تعداد تقسیمات فضا در مدل‌های تکه‌ای معمولاً بسیار کمتر از تعداد بردارهای پشتیبان است. این موضوع، پیشتر در مورد مدل‌های طبقه‌بندی داده‌های قطعی در [۳۳] و [۴۷] گزارش شده است. به همین دلیل، میانگین رتبه زمان آزمون مدل بااحتیاط تکه‌ای پیشنهادی MEM+PiCRSVM0-HE بهتر از میانگین رتبه زمان آزمون سایر مدل‌ها است.

– میانگین رتبه زمان آزمون مدل تکه‌ای MEM+PiCRSVM0-HE بهتر از میانگین رتبه زمان آزمون مدل تکه‌ای MIK+PiCRSVM0-HE است چرا که مدل تکه‌ای MEM+PiCRSVM0-HE در فضای ورودی کار می‌کند و لذا پیچیدگی تابع تصمیم آن، فقط وابسته به تعداد ویژگی‌های داده (و مستقل از تعداد همسایگان داده ناقص و تعداد داده‌هاست) است در حالی که پیچیدگی تابع تصمیم مدل تکه‌ای MIK+PiCRSVM0-HC (همچون تابع تصمیم مدل RSVM-HC)، ضمن وابستگی به تعداد ویژگی‌های داده، با افزایش تعداد همسایگان داده ناقص افزایش می‌یابد.

– به دلیل مشابه، میانگین رتبه زمان آزمون مدل‌های Embedded، MK+DSVM، EM+DSVM و MIK+ARSVM-HE بهتر از میانگین رتبه زمان آزمون مدل‌های مقاوم غیر تکه‌ای MIK+CRSVM-HE، MIK+CRSVM-HC و MIK+PiCRSVM0-HC است.

شکل ۲ رتبه مدل‌های مختلف را از نظر سه معیار نرخ خطا، زمان آموزش و زمان آزمون مقایسه می‌کند. می‌توان گفت که:

– هزینه کسب کمترین نرخ خطا در مدل CRSVM-HE، صرف تقریباً بیشترین زمان آموزش است.

– رتبه زمان آموزش و زمان آزمون مدل‌های تکه‌ای پیشنهادی، بهتر از رتبه زمان آموزش و زمان آزمون مدل‌های غیر تکه‌ای پیشنهادی متناظر است و رتبه نرخ خطای مدل‌های تکه‌ای پیشنهادی، کمی بدتر از رتبه نرخ خطای مدل‌های غیر تکه‌ای پیشنهادی متناظر است.

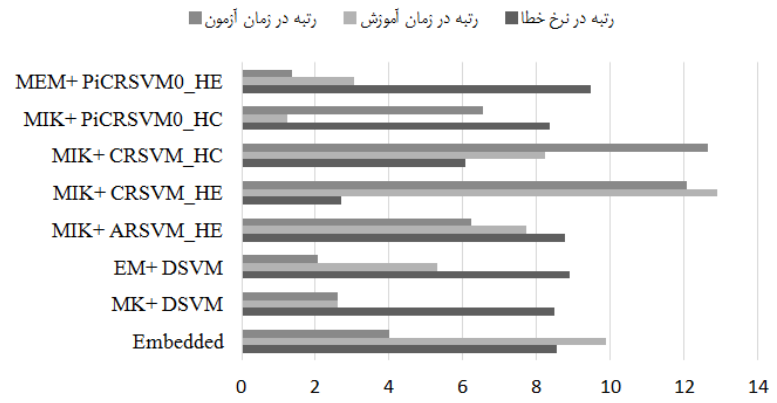
٤-۲ طبقه‌بندی داده‌های کامل قطعی

برای کاهش زمان آموزش مدل DSVM در طبقه‌بندی به ازای یک مجموعه داده آموزشی قطعی کامل می‌توان نخست به کمک الگوریتم‌های HCP یا HRP، مجموعه داده‌های آموزشی قطعی کامل را به یک مجموعه داده توأم با عدم قطعیت با تعداد عضوهای بسیار کمتر تبدیل نمود و سپس برای آموزش طبقه‌بند بر اساس مجموعه داده توأم با عدم قطعیت حاصل می‌توان از مدل‌های RSVM پیشنهادی استفاده کرد.

٤-۲-۱ مدل‌های استفاده‌شده در آزمایش‌های طبقه‌بندی داده‌های قطعی کامل

در این آزمایش‌ها ترکیب‌های مختلف الگوریتم‌های HCP یا HRP و مدل‌های مختلف مقاوم RSVM پیشنهادی، با مدل‌های غیر مقاوم DSVM و شبکه‌های عصبی حداقل–حداکثر فازی DCFMN[1] [۴۸]

1. Data-Core-Based Fuzzy Min-Max Neural Network



شکل ۲: رتبه مدل‌های طبقه‌بندی از نظر نرخ خطا، زمان آموزش و زمان آزمون.

جدول ۶: میانگین زمان آزمون مدل‌های مختلف (بر حسب ثانیه).

نام مجموعه داده	Embedded	MK+ DSVM	EM+ DSVM	MIK+ ARSVM-HE	MIK+ CRSVM-HE	MIK+ CRSVM-HC	MIK+ PiCRSVM0-HC	MEM+ PiCRSVM0-HE
B.C.W.	۰٫۰۰۳۱(۴)	۰٫۰۰۲۸(۳)	۰٫۰۰۲۱(۱)	۰٫۰۱۹۸(۸)	۰٫۲(۱۳)	۰٫۲(۱۳)	۰٫۰۱۵(۶)	۰٫۰۰۲۲(۲)
KDD	۰٫۰۰۹۱(۴)	۰٫۰۰۵۷(۲)	۰٫۰۰۵۱(۱)	۰٫۰۸(۷)	۰٫۷۹۸۶(۱۴)	۰٫۶۴(۱۳)	۰٫۰۸۴۴(۸)	۰٫۰۰۵۹(۳)
C.A.	۰٫۰۰۳۷(۴)	۰٫۰۰۱۹(۲)	۰٫۰۰۲۲(۳)	۰٫۰۳(۷)	۰٫۰۹(۱۱)	۰٫۱۵(۱۲)	۰٫۰۳۴۶(۹)	۰٫۰۰۱۷(۱)
Cleveland	۰٫۰۱۲(۴)	۰٫۰۱۱۴(۳)	۰٫۰۱۰۷(۲)	۰٫۰۲۶(۵)	۰٫۲۶(۱۴)	۰٫۱۰۴(۱۰)	۰٫۰۲۷(۶)	۰٫۰۰۳۷(۱)
Hungarian	۰٫۰۱۳۹(۴)	۰٫۰۱۳۷(۳)	۰٫۰۱۲۶(۲)	۰٫۰۲۶(۵)	۰٫۲۶(۱۴)	۰٫۱۸۲(۱۱)	۰٫۰۳۵۵(۶)	۰٫۰۰۴۷(۱)
Switzerland	۰٫۰۰۹۶(۴)	۰٫۰۰۸۱(۲)	۰٫۰۰۸۳(۳)	۰٫۰۲۶(۶)	۰٫۱۵۶(۱۱)	۰٫۲۳۳۷(۱۳)	۰٫۰۱۳۹(۵)	۰٫۰۰۲۱(۱)
L.B.VA	۰٫۰۱۲۴(۴)	۰٫۰۱۰۹(۲)	۰٫۰۱۱۱(۳)	۰٫۰۲۶(۶)	۰٫۰۷۸(۱۰)	۰٫۲۶(۱۴)	۰٫۰۲۵۹(۵)	۰٫۰۰۲۱(۱)
Hepatitis	۰٫۰۲۱۲(۴)	۰٫۰۱۸۴(۲)	۰٫۰۱۹۳(۳)	۰٫۰۳۸(۵)	۰٫۱۱۴(۸)	۰٫۲۲۸(۱۳)	۰٫۰۴۸(۶)	۰٫۰۰۲۹(۱)
Material	۰٫۰۰۷۱(۴)	۰٫۰۰۶(۳)	۰٫۰۰۵۸(۲)	۰٫۰۲۶(۵)	۰٫۱۰۴(۱۳)	۰٫۱۸۲(۱۴)	۰٫۰۲۹۹(۶)	۰٫۰۰۲۴(۱)
REL-L	۰٫۰۱۱۵(۴)	۰٫۰۱۰۸(۳)	۰٫۰۱۰۲(۲)	۰٫۰۲۵۸(۵)	۰٫۲۵۹۵(۱۴)	۰٫۲۳۲۳(۱۳)	۰٫۰۵۲۶(۶)	۰٫۰۰۳۴(۱)
SPAN	۰٫۰۱۰۷(۴)	۰٫۰۰۹۷(۳)	۰٫۰۰۹۴(۲)	۰٫۰۲۶(۵)	۰٫۲۰۸(۱۳)	۰٫۲۶(۱۴)	۰٫۰۳۵۱(۷)	۰٫۰۰۲۹(۱)
T-OR-D	۰٫۰۰۶(۴)	۰٫۰۰۴۵(۲)	۰٫۰۰۴۷(۳)	۰٫۰۲۶(۶)	۰٫۲۰۸(۱۳)	۰٫۱۵۶(۱۱)	۰٫۰۲۳۴(۵)	۰٫۰۰۱۸(۱)
TYPE	۰٫۰۱۱۷(۴)	۰٫۰۱۱۵(۳)	۰٫۰۱۰۴(۲)	۰٫۰۲۶(۵)	۰٫۲۳۴(۱۴)	۰٫۱۸۲(۱۳)	۰٫۰۴۹۲(۷)	۰٫۰۰۴۳(۱)
P.P.	۰٫۰۰۹۵(۴)	۰٫۰۰۹۲(۳)	۰٫۰۰۸۷(۲)	۰٫۰۱۶(۵)	۰٫۰۴۹۲(۸)	۰٫۱۴۴(۱۴)	۰٫۰۱۹(۶)	۰٫۰۰۲۸(۱)
SECOM	۰٫۱۴۲۷(۴)	۰٫۱۰۰۷(۳)	۰٫۰۸۳۶(۱)	۱٫۱۸۱۹(۹)	۳٫۵۴۶(۱۱)	۵٫۹۰۹۴(۱۳)	۰٫۹۲۶(۸)	۰٫۰۹۷۱(۲)
Allhypeo	۰٫۰۰۶۵(۴)	۰٫۰۰۴(۲)	۰٫۰۰۳۶(۱)	۰٫۰۵۸(۱۰)	۰٫۳۴۸(۱۲)	۰٫۲۳۲(۱۱)	۰٫۰۳۰۵(۶)	۰٫۰۰۴۲(۳)
Sick	۰٫۰۱۰۲(۴)	۰٫۰۰۹۹(۳)	۰٫۰۰۷۳(۲)	۰٫۰۵۸(۷)	۰٫۱۷۴(۱۲)	۰٫۲۹(۱۳)	۰٫۰۷۲۷(۹)	۰٫۰۰۷۲(۱)
میانگین زمان آزمون	۰٫۰۱۷۷	۰٫۰۱۴۱	۰٫۰۱۲۷	۰٫۱۰۰۹	۰٫۴۱۶۸	۰٫۵۶۳۸	۰٫۰۹۱۷	۰٫۰۰۸۹
میانگین رتبه	۴٫۰۰	۰٫۵۹	۲٫۰۶	۶٫۲۴	۱۲٫۰۶	۱۲٫۶۵	۶٫۵۳	۱٫۳۵

مقایسه می‌گردند. ترکیب‌های یادشده عبارتند از:

۱) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابردایروی در فضای ویژگی با استفاده از الگوریتم HCP و سپس استفاده از مدل CRSVM-HC برای آموزش طبقه‌بند. این روش، HCP+CRSVM-HC نامیده می‌شود.

۲) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابردایروی در فضای ورودی با استفاده از الگوریتم HCP و سپس استفاده از مدل ARSVM-HC برای آموزش طبقه‌بند. این روش، HCP+ARSVM-HC نامیده می‌شود.

۳) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابردایروی در فضای ورودی با استفاده از الگوریتم HCP و سپس استفاده از DARSVM-HC برای آموزش طبقه‌بند. این روش، HCP+DARSVM-HC نامیده می‌شود.

۴) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابردایروی در فضای ورودی با استفاده از الگوریتم HCP و سپس استفاده از CARSVM-HC برای آموزش طبقه‌بند. این روش، HCP+CARSVM-HC نامیده می‌شود.

۵) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابردایروی در فضای ورودی با استفاده از الگوریتم HCP و سپس استفاده از مدل تکه‌ای PiCRSVM0 برای آموزش طبقه‌بند. این روش، HCP+PiCRSVM0 نامیده می‌شود.

۶) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابرمستطیلی در فضای ورودی با استفاده از الگوریتم HRP و سپس استفاده از مدل تکه‌ای PiCRSVM0 برای آموزش طبقه‌بند. این روش، HRP+PiCRSVM0 نامیده می‌شود.

۷) تبدیل مجموعه داده آموزشی قطعی به یک مجموعه داده آموزشی توأم با عدم قطعیت ابرمستطیلی در فضای ورودی با استفاده از الگوریتم HRP و سپس استفاده از مدل تکه‌ای PiCRSVM1 برای آموزش طبقه‌بند. این روش، HRP+PiCRSVM1 نامیده می‌شود.

جدول ۷: مشخصات تعدادی از مجموعه‌های داده کامل قطعی از مخزن دانشگاه آیروین کالیفرنیا.

ردیف	نام مجموعه داده	تعداد داده‌ها	تعداد ویژگی‌ها	تعداد طبقات داده
۱	Iris	۱۵۰	۴	۳
۲	Sonar	۲۰۸	۶۰	۲
۳	Glass	۲۱۴	۹	۷
۴	Heart (SPECTF Heart)	۲۶۷	۴۴	۲
۵	Haberman	۳۰۶	۳	۲
۶	Ecoli	۳۳۶	۷	۸
۷	Liver (Liver Disorders)	۳۴۵	۶	۲
۸	Musk (Version 1)	۴۷۶	۱۶۷	۲
۹	Transfusion (Blood Transfusion Center Service)	۷۴۸	۴	۲
۱۰	Statlog1 (Vehicle Silhouettes)	۹۴۶	۱۸	۴
۱۱	Yeast	۱۴۸۴	۸	۱۰
۱۲	Statlog2 (Image Segmentation)	۲۳۱۰	۱۹	۷

جدول ۸: میانگین نرخ خطای مدل‌های طبقه‌بندی به ازای داده‌های آموزشی کامل قطعی.

نام مجموعه داده	DSVM	DCFMN	HCP+ ARSVM-HC	HCP+ DARSVM-HC	HCP+ CRSVM-HC	HCP+ PiCRSVM0	HRP+ PiCRSVM0	HRP+ PiCRSVM1
Iris	۴٫۵۴(۷)	۶٫۴۵(۱۲)	۵٫۳۶(۷)	۵٫۳۳(۱۲)	۴٫۵۴(۱)	۴٫۶۲(۱)	۵٫۳۱(۲)	۵٫۳۱(۶)
Sonar	۱۵٫۰۳(۵)	۱۷٫۳۴(۲)	۱۶٫۳۲(۵)	۱۶٫۳۴(۱۲)	۱۵٫۰۳(۱)	۱۶٫۱۳(۱)	۱۵٫۱(۷)	۱۵٫۱(۴)
Glass	۲۸٫۴۲(۲)	۲۷٫۹۳(۱۱)	۲۶٫۹۱(۲)	۲۶٫۹۲(۱۱)	۲۸٫۴۲(۱۲)	۲۷٫۲۳(۱۲)	۲۷٫۱۹(۹)	۲۷٫۱۹(۸)
Heart	۱۸٫۰۱(۵)	۱۹٫۰۳(۹)	۱۹٫۶۲(۵)	۱۹٫۶۲(۹)	۱۸٫۰۱(۳)	۱۷٫۶۱(۳)	۱۹٫۶۲(۲)	۱۹٫۶۲(۱۰)
Haberman	۲۳٫۶۷(۴)	۲۴٫۳۴(۱۰)	۲۵٫۸۱(۴)	۲۵٫۸۵(۱۰)	۲۳٫۶۷(۴)	۲۳٫۶۲(۴)	۲۳٫۶۱(۳)	۲۳٫۶۱(۲)
Ecoli	۱۲٫۹۸(۱۰)	۱۳٫۵۶(۹)	۱۳٫۸۲(۱۰)	۱۳٫۸۲(۹)	۱۲٫۹۸(۲)	۱۳٫۵۳(۲)	۱۲٫۹۱(۸)	۱۲٫۹۱(۱)
Liver	۲۵٫۷۱(۱)	۲۵٫۰۱(۶)	۲۴٫۷۳(۱)	۲۴٫۷۴(۶)	۲۵٫۷۱(۱۱)	۲۴٫۳۵(۱۱)	۲۷٫۶۴(۲)	۲۷٫۶۴(۱۲)
Musk	۳۵٫۹۱(۱)	۳۵٫۹۳(۳)	۳۸٫۹۱(۱)	۳۸٫۹۱(۳)	۳۵٫۹۱(۲)	۳۸٫۳۲(۲)	۳۷٫۹۳(۱۰)	۳۷٫۹۳(۹)
Transfusion	۳۱٫۱۹(۲)	۳۲٫۳۷(۹)	۳۲٫۶۹(۲)	۳۲٫۶۸(۹)	۳۱٫۱۹(۳)	۳۲٫۶۲(۳)	۳۱٫۰۶(۱۰)	۳۱٫۰۶(۱)
Statlog۱	۲۷٫۰۹(۳)	۳۰٫۵۲(۱۲)	۲۷٫۰۲(۳)	۲۷٫۰۲(۱۲)	۲۷٫۰۹(۷)	۲۹٫۶۴(۷)	۲۹٫۴۵(۱۱)	۲۹٫۴۵(۹)
Yeast	۲۹٫۸۲(۱)	۳۰٫۹۱(۱۱)	۳۰٫۰۳(۱)	۳۰٫۰۴(۱۱)	۲۹٫۵۲(۲)	۳۰٫۶۸(۲)	۳۰٫۰۲(۱۰)	۳۰٫۰۲(۶)
Statlog۲	۳٫۱۸(۷)	۴٫۳۵(۱۰)	۴٫۷۹(۷)	۴٫۷۹(۱۰)	۳٫۱۸(۱)	۳٫۵۶(۱)	۳٫۴۶(۳)	۳٫۴۶(۲)
Statlog۳	۹٫۰۹(۱)	۱۰٫۷۲(۱۲)	۹٫۰۵(۱)	۹٫۰۷(۱۲)	۹٫۰۹(۷)	۹٫۹۵(۷)	۹٫۵۲(۱۱)	۹٫۵۲(۹)
میانگین نرخ خطا	۲۰٫۱۸	۲۱٫۴۲	۲۱٫۱۶	۲۱٫۱۶	۲۰٫۳۳	۲۱٫۰۳	۲۰٫۹۱	۲۰٫۹۹
میانگین رتبه طبقه‌بند	۳٫۷۷	۹٫۶۹	۸٫۳۸	۸٫۶۹	۴٫۳۱	۸٫۰۸	۶٫۷۷	۶٫۰۸

٤-۲-۲ مجموعه داده‌های کامل قطعی استفاده‌شده در آزمایش‌ها

داده‌های آموزشی قطعی کامل که در جدول ۷ لیست شده‌اند از مخزن دانشگاه آیروین کالیفرنیا هستند. تعداد داده‌های برخی از این مجموعه‌های داده بسیار زیاد است. لذا با توجه به محدودیت زمان و سخت‌افزار، در هر آزمایش حداکثر از ۱۰۰۰ داده از هر مجموعه آموزشی به صورت تصادفی و با حفظ نسبت داده‌های هر طبقه به کل داده‌ها انتخاب گردیدند.

٤-۲-۳ نتایج آزمایشات طبقه‌بندی داده‌های کامل قطعی

٤-۲-۳-۱ نرخ خطا

جدول ۸ نرخ خطای مدل‌های مختلف طبقه‌بندی به ازای هر یک از مجموعه‌های داده یادشده را نشان می‌دهد. رتبه هر مدل به ازای هر مجموعه داده در پرانتز مشخص شده است. نتایج زیر از جدول ۸ به دست می‌آید:

– بر اساس آزمون Friedman، نرخ خطای این مدل‌ها یکسان نیست و بر اساس روال Finner، برتری مدل DSVM از نظر نرخ خطا نسبت به سایر مدل‌ها به سختی و با بازه اطمینان ۶۲٪ اثبات می‌شود.

– مدل DSVM از مجموعه آموزشی تقریبی استفاده نمی‌کند و به همین دلیل از مدل‌های مقاومی چون HCP+CRSVM-HC که از یک مجموعه آموزشی تقریبی استفاده می‌کنند برتر است (به یاد بیاورید که در الگوریتم‌های HCP یا HRP، هر مجموعه داده مجاور هم‌طبقه با یک مجموعه ابردایروی یا ابرمستطیلی تقریب زده می‌شود).

– در مجموع، میانگین رتبه خطای مدل‌های بااحتیاط از مدل‌های بی‌احتیاط (به استثنای مدل DSVM)، پایین‌تر (بهتر) می‌باشد و این نشان‌دهنده اثر افزودن احتیاط به این مدل‌ها است.

– میانگین رتبه نرخ خطای مدل بااحتیاط غیر تکه‌ای پیشنهادی HCP+CRSVM-HC بهتر است از میانگین رتبه نرخ خطای مدل‌های طبقه‌بندی بااحتیاط تکه‌ای پیشنهادی HCP+PiCRSVM0، HRP+PiCRSVM0 و HRP+PiCRSVM1 چرا که مدل‌های تکه‌ای، نگاهی محلی به داده‌ها دارند (هر تکه از طبقه‌بند با استفاده از بخشی از داده‌های آموزشی ایجاد می‌شود نه با استفاده از همه داده‌های آموزشی) و تبعاً به جواب بهینه محلی خواهند رسید.

٤-۲-۳-۲ زمان آموزش

جدول ۹ زمان آموزش هر یک از مدل‌های مختلف طبقه‌بندی به ازای

جدول ۹: زمان آموزش مدل‌های طبقه‌بندی مختلف به ازای هر یک از مجموعه‌های آموزشی (بر حسب ثانیه).

نام مجموعه داده	DSVM	DCFMN	HCP+ ARSVM-HC	HCP+ DARSVM-HC	HCP+ CRSVM-HC	HCP+ PiCRSVM0	HRP+ PiCRSVM0	HRP+ PiCRSVM1
Iris	۱٫۵۳(۱۲)	۰٫۰۱(۱)	۰٫۴۲(۷)	۰٫۲۲(۹)	۱٫۳۲(۱۱)	۰٫۰۲(۴)	۰٫۰۱(۲)	۰٫۰۲(۳)
Sonar	۴٫۶۸(۱۲)	۰٫۰۳(۱)	۱٫۵۶(۹)	۰٫۵۶(۷)	۲٫۷۱(۱۱)	۰٫۰۸(۲)	۰٫۶۲(۵)	۰٫۷۳(۶)
Glass	۳٫۹۹(۱۲)	۰٫۰۳(۱)	۱٫۰۶(۷)	۰٫۴۲(۸)	۲٫۶۱(۱۰)	۰٫۰۸(۲)	۰٫۰۸(۳)	۰٫۰۹(۴)
Heart	۴٫۰۳(۱۲)	۰٫۰۳(۱)	۱٫۲۹(۹)	۰٫۶۱(۸)	۲٫۶۸(۱۱)	۰٫۱۱(۲)	۰٫۴۸(۳)	۰٫۵۶(۵)
Haberman	۶٫۲۳(۱۲)	۰٫۰۵(۱)	۲٫۲۶(۸)	۱٫۱۲(۷)	۳٫۹۲(۱۰)	۰٫۲۲(۴)	۰٫۰۶(۲)	۰٫۰۷(۳)
Ecoli	۷٫۱۲(۱۲)	۰٫۰۵(۵)	۲٫۲۹(۹)	۱٫۲۷(۷)	۵٫۱۲(۱۱)	۰٫۲۳(۴)	۰٫۱۳(۲)	۰٫۱۶(۳)
Liver	۶٫۳۹(۱۲)	۰٫۰۵(۱)	۱٫۹۶(۸)	۰٫۸۹(۷)	۳٫۷۲(۱۱)	۰٫۳۵(۴)	۰٫۱۵(۲)	۰٫۱۷(۳)
Musk	۱۲٫۶۵(۱۲)	۰٫۱(۱)	۶٫۴۹(۷)	۲٫۷۵(۶)	۸٫۹۳(۸)	۱٫۱۱(۲)	۱۰٫۳۸(۱۰)	۱۲٫۱۲(۱۱)
Transfusion	۱۷٫۶۵(۱۲)	۰٫۱۵(۱)	۶٫۶۵(۷)	۲٫۱۱(۸)	۱۳٫۴۱(۱۱)	۱٫۴۹(۴)	۰٫۱۸(۲)	۰٫۲۱(۳)
Statlog۱	۳۲٫۶۵(۱۲)	۰٫۲۴(۱)	۱۲٫۳۲(۹)	۵٫۱۷(۸)	۲۰٫۸۱(۱۱)	۳٫۴۶(۴)	۱٫۸۹(۲)	۲٫۲(۳)
Yeast	۳۰٫۹۱(۱۲)	۰٫۲۵(۱)	۱۶٫۹۲(۹)	۶٫۲۲(۸)	۲۲٫۹۱(۱۱)	۲٫۶۷(۴)	۰٫۵۵(۲)	۰٫۶۴(۳)
Statlog۲	۳۵٫۲۶(۱۲)	۰٫۲۸(۱)	۱۲٫۶۵(۷)	۹٫۱۲(۸)	۱۹٫۹۲(۱۱)	۱۰٫۲۶(۵)	۵٫۲۴(۲)	۶٫۱۲(۳)
Statlog۳	۴۱٫۳۲(۱۲)	۰٫۳۳(۱)	۱۷٫۶۲(۷)	۷٫۳۷(۹)	۲۹٫۰۱(۱۱)	۵٫۵۸(۳)	۵٫۲۹(۲)	۶٫۱۸(۴)
میانگین زمان آموزش	۱۵٫۷۱	۰٫۱۳	۶٫۴۲	۲٫۹۱	۱۰٫۵۴	۲٫۰۰	۱٫۹۳	۲٫۲۵
میانگین رتبه طبقه‌بند	۱۲٫۰۰	۱٫۰۰	۷٫۹۲	۵٫۳۱	۱۰٫۶۲	۳٫۳۸	۳٫۰۰	۴٫۱۵

هر یک از مجموعه‌های داده یادشده را نشان می‌دهد. نتایج زیر از جدول ۹ به دست می‌آید:

– میانگین رتبه زمان آموزش مدل‌های تکه‌ای بهتر از میانگین رتبه زمان آموزش مدل‌های غیر تکه‌ای مبتنی بر مدل SVM است.

٤-۲-۳-۳ تعداد بردارهای پشتیبان، تعداد تقسیمات فضا

زمان آزمون مدل‌های مبتنی بر SVM به تعداد بردارهای پشتیبان (تعداد ضرایب غیر صفر بردار وزن) و زمان آزمون مدل‌های تکه‌ای به تعداد تقسیمات فضا بستگی دارد. جدول ۱۰ نسبت تعداد بردارهای پشتیبان به تعداد داده‌های آموزشی در مدل‌های مبتنی بر SVM و نسبت تعداد تقسیمات فضا به تعداد داده‌های آموزشی در مدل‌های تکه‌ای را نشان می‌دهد. نتایج زیر از جدول ۱۰ به دست آمده‌اند:

– درصد بردارهای پشتیبان مدل‌های DSVM و HCP+DARSVM-HC بسیار کمتر از تعداد بردارهای پشتیبان مدل‌های HCP+ARSVM-HC و HCP+CRSVM-HC است. این موضوع قابل پیش‌بینی بود چرا که برخلاف مدل‌های ARSVM-HC و CRSVM-HC تنک‌شدن بردار وزن مدل‌های دوگانی چون DSVM و DARSVM-HC تضمین شده است.

– تعداد تقسیمات فضا در مدل‌های تکه‌ای بسیار کمتر از تعداد بردارهای پشتیبان در مدل‌های مبتنی بر SVM است. این موضوع پیشتر در مورد مدل‌های طبقه‌بندی داده‌های قطعی در [۳۳] و [۴۷] گزارش شده است.

٤-۳-۲-٤ زمان آزمون

جدول ۱۱ زمان آزمون مدل‌های مختلف طبقه‌بندی به ازای هر یک از مجموعه‌های داده یادشده را نشان می‌دهد. نتایج زیر از جدول ۱۱ به دست آمده‌اند:

– میانگین رتبه زمان آزمون مدل‌های DSVM و HCP+CRSVM-HC بدتر از میانگین رتبه زمان آزمون سایر مدل‌ها است.

– میانگین رتبه زمان آزمون مدل‌های تکه‌ای و مدل DCFMN بهتر از میانگین رتبه زمان آزمون سایر مدل‌هاست.

– میانگین رتبه زمان آزمون مدل HCP+ARSVM-HC بهتر از میانگین رتبه مدل DSVM است چرا که:

۱) زمان آزمون مدل SVM به تعداد داده‌های آموزشی قطعی بستگی دارد در حالی که زمان آزمون مدل HCP+ARSVM-HC به تعداد داده‌های آموزشی توأم با عدم قطعیت وابسته است که با استفاده از الگوریتم ادغام HCP از روی داده‌های آموزشی قطعی به دست می‌آید. تعداد داده‌های آموزشی توأم با عدم قطعیت به دست آمده از الگوریتم ادغام HCP معمولاً بسیار کمتر از تعداد داده‌های آموزشی قطعی است.

۲) اگرچه زمان آزمون DSVM کمتر از زمان آزمون مدل SVM است (زمان آزمون مدل DSVM به تعداد بردارهای پشتیبان بستگی دارد که کمتر یا مساوی تعداد داده‌های آموزشی قطعی است) ولی همان طور که در جدول ۱۰ ملاحظه شد بردارهای پشتیبان، بخش بزرگی از داده‌های آموزشی هستند.

۳) نسبت تعداد داده‌های آموزشی توأم با عدم قطعیت به دست آمده با الگوریتم ادغام به تعداد داده‌های آموزشی معمولاً بسیار کمتر از نسبت تعداد بردارهای پشتیبان به تعداد داده‌های آموزشی است.

– میانگین رتبه زمان آزمون مدل HCP+ARSVM-HC و HCP+DARSVM-HC به ترتیب بهتر از میانگین رتبه زمان آزمون مدل HCP+CRSVM-HC است چرا که توابع تصمیم مدل‌های ARSVM-HC و DARSVM-HC مستقل از داده‌های آموزشی قبل از ادغام هستند، در حالی که تابع تصمیم مدل CRSVM-HC به داده‌های آموزشی قبل از ادغام وابسته است.

شکل ۳ رتبه مدل‌های مختلف را از نظر سه معیار نرخ خطا، زمان آموزش و زمان آزمون نشان می‌دهد. در مجموع می‌توان گفت:

– مدل HCP+CRSVM-HC از نظر کمترین نرخ خطا، دومین رتبه را دارد که پس از مدل DSVM قرار می‌گیرد.

– رتبه زمان آزمون و رتبه زمان آموزش مدل‌های HCP+ARSVM-HC، HCP+DARSVM-HC و مدل‌های مقاوم تکه‌ای، بهتر از رتبه زمان آزمون و رتبه زمان آموزش مدل DSVM است.

– رتبه زمان آزمون مدل‌های HCP+ARSVM-HC و HCP+DARSVM-HC بهتر از رتبه زمان آزمون HCP+CRSVM-HC است.

جدول ۱۰: نسبت تعداد بردارهای پشتیبان به تعداد داده‌های آموزشی در مدل‌های مبتنی بر SVM و نسبت تقسیمات فضا به تعداد داده‌های آموزشی در مدل‌های تکه‌ای.

نام مجموعه داده	DSVM	DCFMN	HCP+ ARSVM-HC	HCP+ DARSVM-HC	HCP+ CRSVM-HC	HCP+ PiCRSVM0	HRP+ PiCRSVM0	HRP+ PiCRSVM1
Iris	۳۴٫۵۷(۵)	۱۶٫۲۸(۳)	۱۰۰(۱۲)	۴۶٫۹۵(۱۱)	۹۹٫۹۳(۸)	۱۶٫۷۹(۹)	۱۳٫۶۳(۴)	۱۳٫۶۳(۱)
Sonar	۵۹٫۳۶(۵)	۱۲٫۲۴(۱)	۹۹٫۶۲(۹)	۹۵٫۱۹(۱۱)	۱۰۰(۶)	۱۳٫۵۵(۱۱)	۱۷٫۲۹(۲)	۱۷٫۲۹(۳)
Glass	۷۲٫۹(۵)	۲۱٫۰۲(۱)	۹۹٫۹۹(۸)	۱۰۰(۷)	۱۰۰(۶)	۲۲٫۴۶(۹)	۲۳٫۵۶(۲)	۲۳٫۵۶(۳)
Heart	۴۶٫۷۸(۵)	۱۶٫۶(۳)	۱۰۰(۱۰)	۶۸٫۹۹(۱۰)	۹۹٫۶(۷)	۱۷٫۴۴(۹)	۱۶٫۳۳(۴)	۱۶٫۳۳(۱)
Haberman	۴۸٫۱۹(۵)	۱۴٫۰۴(۱)	۱۰۰(۱۲)	۷۳٫۴۷(۱۱)	۹۹٫۳۱(۶)	۱۴٫۹۹(۹)	۱۵٫۶۴(۲)	۱۵٫۶۴(۱)
Ecoli	۳۹٫۳۱(۵)	۱۲٫۴۸(۱)	۱۰۰(۹)	۵۸٫۴۳(۹)	۱۰۰(۷)	۱۳٫۲۳(۹)	۱۳٫۱۵(۴)	۱۳٫۱۵(۲)
Liver	۳۴٫۴۶(۵)	۱۷٫۴(۳)	۱۰۰(۱۰)	۴۵٫۹۲(۹)	۱۰۰(۸)	۱۷٫۸۷(۱۰)	۱۴٫۰۴(۴)	۱۴٫۰۴(۱)
Musk	۵۹٫۸۷(۵)	۱۶٫۱۱(۱)	۹۹٫۹۸(۱۰)	۹۳٫۲۴(۱۱)	۹۹٫۸۲(۶)	۱۷٫۳۳(۹)	۱۸٫۹۱(۲)	۱۸٫۹۱(۳)
Transfusion	۵۵٫۶۷(۵)	۲۱٫۷۷(۳)	۹۹٫۴۶(۹)	۸۱٫۳۶(۱۰)	۱۰۰(۶)	۲۲٫۷۱(۱۲)	۲۰٫۲۲(۴)	۲۰٫۲۲(۱)
Statlog۱	۳۹٫۶۵(۵)	۱۰٫۲۷(۱)	۱۰۰(۱۰)	۶۰٫۷(۱۰)	۹۹٫۹۳(۶)	۱۱٫۰۸(۹)	۱۲٫۳۶(۲)	۱۲٫۳۶(۳)
Yeast	۵۸٫۱۲(۵)	۱۷٫۹۶(۱)	۱۰۰(۱۱)	۸۸٫۶۷(۹)	۱۰۰(۶)	۱۹٫۰۸(۱۱)	۱۹٫۲۶(۲)	۱۹٫۲۶(۳)
Statlog۲	۲۸٫۴(۵)	۱۴٫۳۲(۳)	۹۹٫۶۷(۹)	۳۷٫۱۶(۱۱)	۹۹٫۹۹(۸)	۱۴٫۷۱(۱۰)	۱۱٫۵۶(۴)	۱۱٫۵۶(۱)
Statlog۳	۳۱٫۶۲(۵)	۱۳٫۹۶(۳)	۱۰۰(۱۰)	۴۳٫۳(۱۰)	۱۰۰(۸)	۱۴٫۴۵(۱۰)	۱۲٫۱۱(۴)	۱۲٫۱۱(۱)
میانگین نسبت	۴۶٫۸۴	۱۵٫۷۳	۹۹٫۹۰	۶۸٫۷۲	۹۹٫۸۹	۱۶٫۵۹	۱۶٫۰۰	۱۶٫۰۰
میانگین رتبه طبقه‌بند	۵٫۰۰	۱٫۹۲	۹٫۹۲	۷٫۷۷	۶٫۷۷	۹٫۷۷	۳٫۰۸	۲٫۰۰

جدول ۱۱: زمان آزمون مدل‌های مختلف (بر حسب ثانیه).

نام مجموعه داده	DSVM	DCFMN	HCP+ ARSVM-HC	HCP+ DARSVM-HC	HCP+ CRSVM-HC	HCP+ PiCRSVM0	HRP+ PiCRSVM0	HRP+ PiCRSVM1
Iris	۰٫۰۱۳۸(۸)	۰٫۰۰۱۳(۳)	۰٫۰۰۸(۷)	۰٫۰۰۳۸(۵)	۰٫۰۵۶(۱۱)	۰٫۰۰۱۳(۴)	۰٫۰۰۱۱(۱)	۰٫۰۰۱۲(۲)
Sonar	۰٫۴۲۷۴(۸)	۰٫۱۴۷(۱)	۰٫۱۱۹۵(۶)	۰٫۱۱۴۲(۵)	۰٫۶(۱۱)	۰٫۰۱۶۳(۲)	۰٫۰۲۰۷(۳)	۰٫۰۲۳۱(۴)
Glass	۰٫۰۹۱۹(۸)	۰٫۰۰۳۸(۱)	۰٫۰۱۸(۵)	۰٫۰۱۸(۶)	۰٫۱۴۴(۱۱)	۰٫۰۰۴(۲)	۰٫۰۰۴۲(۳)	۰٫۰۰۴۷(۴)
Heart	۰٫۱۶۴۷(۸)	۰٫۰۱۴۶(۲)	۰٫۰۸۸(۶)	۰٫۰۶۰۷(۵)	۰٫۲۶۲۹(۱۱)	۰٫۰۱۵۳(۳)	۰٫۰۱۴۴(۱)	۰٫۰۱۶(۴)
Haberman	۰٫۰۲۳۱(۱۰)	۰٫۰۰۰۸(۱)	۰٫۰۰۶(۷)	۰٫۰۰۴۴(۵)	۰٫۰۲۳۸(۱۱)	۰٫۰۰۰۹(۲)	۰٫۰۰۰۹(۳)	۰٫۰۰۱(۴)
Ecoli	۰٫۰۳۳(۸)	۰٫۰۰۱۷(۱)	۰٫۰۱۴(۶)	۰٫۰۰۸۲(۵)	۰٫۰۸۴(۱۱)	۰٫۰۰۱۹(۳)	۰٫۰۰۱۸(۲)	۰٫۰۰۲(۴)
Liver	۰٫۰۴۱۴(۸)	۰٫۰۰۲۱(۳)	۰٫۰۱۲(۶)	۰٫۰۰۵۵(۵)	۰٫۱۰۸(۱۲)	۰٫۰۰۲۱(۴)	۰٫۰۰۱۷(۱)	۰٫۰۰۱۹(۲)
Musk	۱٫۷۹۹۷(۱۲)	۰٫۰۵۳۸(۱)	۰٫۳۳۳۹(۶)	۰٫۳۱۱۴(۵)	۰٫۶۶۶۸(۱۰)	۰٫۰۵۷۹(۲)	۰٫۰۶۳۱(۳)	۰٫۰۷۰۲(۴)
Transfusion	۰٫۰۲۲۳(۸)	۰٫۰۰۱۷(۲)	۰٫۰۰۸(۶)	۰٫۰۰۶۵(۵)	۰٫۰۸(۱۲)	۰٫۰۰۱۸(۴)	۰٫۰۰۱۶(۱)	۰٫۰۰۱۸(۳)
Statlog۱	۰٫۰۸۵۶(۸)	۰٫۰۰۳۷(۱)	۰٫۰۳۶(۶)	۰٫۰۲۱۹(۵)	۰٫۳۳۳۸(۱۱)	۰٫۰۰۴(۲)	۰٫۰۰۴۵(۳)	۰٫۰۰۴۹(۴)
Yeast	۰٫۰۳۷۲(۸)	۰٫۰۰۲۹(۱)	۰٫۰۱۶(۷)	۰٫۰۱۴۲(۵)	۰٫۰۸(۱۲)	۰٫۰۰۳۱(۲)	۰٫۰۰۳۱(۳)	۰٫۰۰۳۴(۴)
Statlog۲	۰٫۰۸۶۳(۱۰)	۰٫۰۰۵۴(۳)	۰٫۰۳۷۹(۶)	۰٫۰۱۴۱(۵)	۰٫۱۵۲(۱۱)	۰٫۰۰۵۶(۴)	۰٫۰۰۴۴(۱)	۰٫۰۰۴۹(۲)
Statlog۳	۰٫۲۰۴۹(۸)	۰٫۰۱۰۱(۳)	۰٫۰۷۲(۷)	۰٫۰۳۱۲(۵)	۰٫۵۷۶(۱۱)	۰٫۰۱۰۴(۴)	۰٫۰۰۸۷(۱)	۰٫۰۰۹۷(۲)
میانگین زمان آزمون	۰٫۲۳۳۲	۰٫۰۰۹۰	۰٫۰۵۹۲	۰٫۰۴۷۲	۰٫۲۴۲۹	۰٫۰۰۹۶	۰٫۰۱۰۰	۰٫۰۱۱۱
میانگین رتبه طبقه‌بند	۸٫۶۲	۱٫۷۷	۶٫۲۳	۵٫۰۸	۱۱٫۱۵	۲٫۹۲	۲٫۰۰	۳٫۳۱

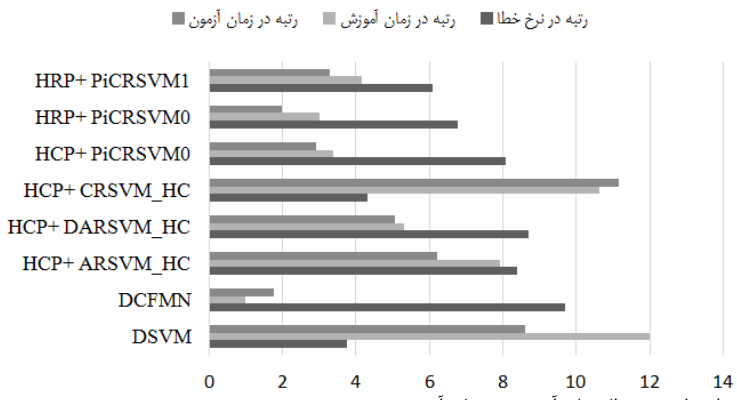
۵- نتیجه‌گیری

۱) برای رفع چالش وجود جواب بی‌معنی در مجموعه جواب‌های ممکن مدل RSVM، مدل RSVM بااحتیاط (CRSVM) معرفی شد. جواب بهینه مدل CRSVM ضرورتاً یک جواب معنی‌دار است.

۲) مدل CRSVM پیشنهادی در فضای ویژگی به ازای مجموعه داده‌های توأم با عدم قطعیت ابردایروی و ابربیضوی به دو مدل استاندارد SOCP تبدیل شد.

۳) زمان آموزش مدل CRSVM پس از استانداردسازی، همچنان بالا است. برای غلبه بر مشکل زمان آموزش، یک مدل طبقه‌بندی تکه‌ای پیشنهاد شد که هر تکه از طبقه‌بند آموزش‌یافته توسط این مدل، حاصل اجرای یک مدل CRSVM به ازای داده‌های آموزشی واقع در بخشی از فضا است. زمان آموزش مدل تکه‌ای، تقریباً رابطه معکوسی با تعداد تقسیمات فضا دارد و زمان آزمون مدل تکه‌ای، تقریباً رابطه مستقیمی با تعداد تقسیمات فضا دارد.

۴) ثابت شد که در حالت خاص، این مدل تکه‌ای به یک مدل نزدیک‌ترین همسایه تبدیل می‌شود که فاقد زمان آموزش است و نشان داده شد که جواب مدل نزدیک‌ترین همسایه به ازای انواع خاصی از داده‌های توأم با عدم قطعیت را می‌توان به روش تحلیلی به دست آورد. مدل نزدیک‌ترین همسایه در طبقه‌بندی مجموعه‌های عدم قطعیت از کلاس‌های متفاوت که دارای عضو مشترک هستند، دچار مشکل شده و برای طبقه‌بندی نقاطی از فضا که به بیش از یک کلاس نسبت داده شده‌اند از استراتژی طبقه‌بندی تصادفی استفاده می‌کند در حالی که این نقاط منتسب به چندطبقه، همانند نقاطی از فضا هستند که به هیچ طبقه‌ای نسبت داده نشده‌اند و لذا باید مثلاً از یک استراتژی طبقه‌بندی مبتنی بر یادگیری برای طبقه‌بندی این نقاط منتسب به چند طبقه استفاده



شکل ۳: رتبه مدل‌های طبقه‌بندی مختلف از نظر نرخ خطا، زمان آموزش و زمان آزمون.

شود. برای همین منظور مدل نزدیک‌ترین همسایه مرتبه یک پیشنهاد شد که به جای استفاده از استراتژی طبقه‌بندی تصادفی برای طبقه‌بندی نقاط منتسب به چند طبقه، از استراتژی طبقه‌بندی CRSVM استفاده می‌کند. سپس نشان داده شد که جواب مدل نزدیک‌ترین همسایه مرتبه یک به ازای انواع خاصی از داده‌های توأم با عدم قطعیت را می‌توان به روش تحلیلی به دست آورد.

۵) پس از اثبات و تحلیل ریاضی مدل‌های پیشنهادی و مزیت‌های آن در کاهش زمان آموزش و زمان آزمون، برای آزمون مدل پیشنهادی و تعیین میزان تأثیر آن در موارد مذکور در مقایسه با مدل‌های بی‌احتیاط، دو دسته آزمایش‌ها انجام گردید.

در دسته اول از آزمایش‌ها، از مدل پیشنهادی برای طبقه‌بندی داده‌های آموزشی ناقص استفاده گردید و نتایج زیر حاصل شد:

- زمان آموزش با نرخ خطا رابطه معکوس دارد یعنی هزینه کسب کمترین نرخ خطا در مدل پیشنهادی CRSVM-HE، صرف تقریباً بیشترین زمان آموزش است.
- رتبه زمان آموزش و زمان آزمون مدل‌های تکه‌ای پیشنهادی، بهتر از رتبه زمان آموزش و زمان آزمون مدل‌های غیر تکه‌ای پیشنهادی است.
- رتبه نرخ خطای مدل‌های تکه‌ای پیشنهادی، کمی بدتر از رتبه نرخ خطای مدل‌های غیر تکه‌ای پیشنهادی است.

در دسته دوم از آزمایش‌ها، از مدل‌های مقاوم و الگوریتم‌های ادغام پیشنهادی برای طبقه‌بند به ازای داده‌های قطعی کامل استفاده شد و نتایج زیر حاصل گردید:

- مدل HCP+CRSVM-HC دومین رتبه کمترین نرخ خطا را دارد که پس از مدل DSVM قرار می‌گیرد.
- رتبه زمان آزمون و رتبه زمان آموزش مدل HCP+ARSVM-HC و مدل‌های مقاوم تکه‌ای، بهتر از رتبه زمان آزمون و رتبه زمان آموزش مدل DSVM است.
- رتبه زمان آزمون مدل HCP+ARSVM-HC بهتر از رتبه زمان آزمون مدل HCP+CRSVM-HC است.

۷) مشکل مدل CRSVM تکه‌ای، نگاه محلی آن به مسئله طبقه‌بندی است و از این رو معمولاً نرخ خطای بالاتری نسبت به مدل CRSVM غیر تکه‌ای دارد.

۸) از طرفی، مشکل مدل CRSVM غیر تکه‌ای، زمان آزمون و زمان آموزش است. شاید بتوان برای بهبود این دو مشکل، از دوگان مدل CRSVM استفاده کرد ولی استفاده از دوگان CRSVM مستلزم به دست آوردن روابط بین متغیرهای مدل اولیه و مدل دوگان است که کار دشواری است و تا کنون نتوانسته‌ایم این روابط را به دست آوریم و لذا انجام این کار به آینده موکول می‌شود.

مراجع

[1] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebert, "Applications of second-order cone programming," *Linear Algebra and Its Applications*, vol. 284, no. 1-3, pp. 193-228, Nov. 1998.

[2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, 2013.

[3] G. Chechik, G. Heitz, G. Elidan, P. Abbeel, and D. Koller, "Max-margin classification of data with absent features," *J. of Machine Learning Research*, vol. 9, no. pp. 1-21, Jan. 2008.

[4] S. Kambar, Generating Synthetic Data by Morphing Transformation for Handwritten Numeral Recognition (with V-SVM), Concordia University, 2005.

[5] P. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. on Document Analysis and Recognition,* pp. 958 963, Edinburgh, UK, 6-6 Aug. 2003.

[6] G. Loosli, A. S. Canu, and L. Bottou, *Large-Scale Kernel Machines*, Massachusetts: MIT Press, 2007.

[7] G. M. Fung, O. L. Mangasarian, and J. W. Shavlik, "Knowledge-based support vector machine classifiers," in *Proc. of the 15th Int. Conf. on Neural Information Processing* Systems, NIPS'02, pp. 537-544, Vancouver, Canada, 9-14 Dec. 2002.

[8] V. Jeyakumar, J. Ormerod, and R. S. Womersley, "Knowledge-based semidefinite linear programming classifiers," *Optimisation Methods Software*, vol. 21, no. 5, pp. 693-706, 2006.

[9] E. Carrizosa, J. Gordillo, and F. Plastria, *Classification Problems with Imprecise Data through Separating Hyperplanes*, MOSI Department, Vrije Universiteit Brussel MOSI/33, 2007.

[10] T. B. Trafalis and R. C. Gilbert, "Robust support vector machines for classification and computational issues," *Optimisation Methods Software*, vol. 22, no. 1, pp. 187-198, 2007.

[11] T. B. Trafalis and R. C. Gilbert, "Robust classification and regression using support vector machines," *European J. of Operational Research*, vol. 173, no. 3, pp. 893-909, 16 Sept. 2006.

[12] H. Xu, Robust Decision Making and Its Applications in Machine Learning, McGill University, 2009.

[13] H. Xu, C. Caramanis, and S. Mannor, "Robustness and regularization of support vector machines," *J. of Machine Learning Research*, vol. 10, pp. 1485-1510, Jul. 2009.

[14] Y. Forghani and H. Sadoghi, "Comment on robustness and regularization of support vector machines by H. Xu et al. (J. of machine learning research, vol. 10, pp. 1485-1510, 2009)," *The J. of Machine Learning Research*, vol. 14, no. 1, pp. 3493-3494, Nov. 2013.

[15] Y. Shi, Y. Tian, G. Kou, Y. Peng, and J. Li, *Optimization Based Data Mining: Theory and Applications*, Springer Science & Business Media, 2011.

[16] H. Xu, C. Caramanis, S. Manor, and S. Yun, "Risk sensitive robust support vector machines," in *Proc. of the 48th IEEE Conf. on Decision and Control, held jointly with the 2009 28th Chinese Control Conf., CDC/CCC'09,* pp. 4655-4661, Shanghai, China, 15-18 Dec. 2009.

[17] H. Xu, S. Mannor, and C. Caramanis, "Robustness, risk, and regularization in support vector machines," *J. of Machine Learning Research*, vol. 10, pp. 1485-1510, Dec. 2009.

[18] H. Xu and S. Mannor, "Robustness and generalization," *Machine Learning*, vol. 86, no. 3, pp. 391-423, Mar. 2012.

[19] W. An and Y. Sun, "PKSVR: a novel prior knowledge-based support vector regression," *Asian J. of Information Technology*, vol. 4, no. 11, pp. 978-980, 2005.
[20] B. Liu, Y. Xiao, L. Cao, and P. S. Yu, "One-class-based uncertain data stream learning," in *Proc. of the 11th SIAM Int. Conf. on Data Mining,* pp. 992-1003, Mesa, AZ, USA, 28-30 Apr. 2011.
[21] J. B. Pothin and C. Richard, "Incorporating prior information into support vector machines in the form of ellipsoidal knowledge sets," in *Proc. 14 th IEEE European Signal Processing Conf., EUSIPCO'06,* 4 pp., Florence, Italy, 4-8 Sept. 2006.
[22] Q. V. Le, A. J. Smola, and T. Gartner, "Simpler knowledge-based support vector machines," in *Proc. of the 23rd Int. Conf. on Machine Learning,* pp. 521-528, Pittsburgh, PA, USA, 25-29 Jun. 2006.
[23] C. H. Teo, *et al.*, Convex learning with invariances, in Advances in Neural Information Processing Systems, 2008.
[24] A. B. Ji, J. H. Pang, and H. J. Qiu, "Support vector machine for classification based on fuzzy training data," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3495-3498, Apr. 2010.
[25] L. Yang and H. Dong, "Support vector machine with truncated pinball loss and its application in pattern recognition," *Chemometrics Intelligent Laboratory Systems*, vol. 177, pp. 89-99, 15 Jun. 2018.
[26] B. B. Gao and J. J. Wang, *A Fast and Robust TSVM for Pattern Classification*, arXiv preprint arXiv:.05406, 2017.
[27] J. Chen, T. Takiguchi, and Y. Ariki, "A robust SVM classification framework using PSM for multi-class recognition," *EURASIP J. on Image Video Processing*, vol. 2015, no. 7, 12 pp., Dec. 2015.
[28] S. Katsumata and A. Takeda, "Robust cost sensitive support vector machine," in *Proc. of the 18th Int. Conf. on Artificial Intelligence and Statistics (AISTATS), JMLR: W&CP,* vol. 38, no. 18, pp.434-443, 2015.
[29] C. Tzelepis, V. Mezaris, and I. Patras, "Linear maximum margin classifier for learning from uncertain data," *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. 40, no. 12, pp. 2948 -2962, Dec. 2017.
[30] Y. Forghani and H. S. Yazdi, "Fuzzy min-max neural network for learning a classifier with symmetric margin," *Neural Processing Letters*, vol. 42, no. 2, pp. 317-353, Oct. 2015.
[31] S. Jing and L. Yang, "A robust extreme learning machine framework for uncertain data classification," *The J. of Supercomputing*, pp. 1-27, 2018. https://doi.org/10.1007/s11227-018-2430-6
[32] E. Adeli, *et al.*, "Semi-supervised discriminative classification robust to sample-outliers and feature-noises," *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. 41, no. 2, pp. 515-522, Feb. 2018.
[33] W. Y. Cheng and C. F. Juang, "An incremental support vector machine-trained TS-type fuzzy system for online classification problems," *Fuzzy Sets Systems*, vol. 163, no. 1, pp. 24-44, 16 Jan. 2011.
[34] MOSEK, Available: https://www.mosek.com/.
[35] D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, 2017.
[36] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *J. of Machine Learning Research*, vol. 7, no. 1, pp. 1-30, Jan. 2006.
[37] O. T. Yildiz and E. Alpaydin, "Ordering and finding the best of K>2 supervised learning algorithms," *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. 28, no. 3, pp. 392-402, Mar. 2006.
[38] B. M. Marlin, *Missing Data Problems in Machine Learning*, Department of Computer Science, University of Toronto, 2008.
[39] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1-38, 1997.
[40] D. B. Rubin, "Multiple imputation after 18+ years," *J. of the American Statistical Association*, vol. 91, no. 434, pp. 473-489, Jun. 1996.
[41] P. K. Shivaswamy, C. Bhattacharyya, and A. J. Smola, "Second order cone programming approaches for handling missing and uncertain data," *J. of Machine Learning Research*, vol. 7, pp. 1283-1314, Jul. 2006.
[42] S. Garcia, A. Fernandez, J. Luengo, and F Herrera., "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044-2064, 15 May 2010.
[43] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Florida, US: Chapman & Hall, 2007.
[44] H. Finner, "On a monotonicity problem in step-down multiple test procedures," *J. of the American Statistical Association*, vol. 88, no. 423, pp. 920-923, Sept. 1993.
[45] Y. Forghani, S. Effati, H. Sadoughi Yazdi, and R. Sigari Tabrizi, "Support vector data description by using hyper-ellipse instead of hyper-sphere," in *Proc. 1st IEEE Int. eConf. on Computer and Knowledge Engineering, ICCKE'11*, pp. 22-27, Mashad, Iran, 13-14 Oct. 2011.
[46] L. M. Manevitz, and M. Yousef, "One-class SVMs for document classification," *J. of Machine Learning Research*, vol. 2, pp. 139-154, Dec. 2001.
[47] C. F. Juang, S. H. Chiu, and S. W. Chang, "A self-organizing TS-type fuzzy network with support vector learning and its application to classification problems," *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 5, pp. 998-1008, Oct. 2007.
[48] H. Zhang, J. Liu, D. Ma, D. Ma, Z. Wang., "Data-core-based fuzzy min-max neural network for pattern classification," *IEEE Trans. on Neural Networks*, vol. 22, no. 12, pp. 2339-2352, Dec. 2011.

یحیی فرقانی در سال ۱۳۷۹ مدرک کارشناسی مهندسی کامپیوتر خود را از دانشگاه فردوسی مشهد و در سال ۱۳۸۴ مدرک کارشناسی ارشد مهندسی کامپیوتر خود را از دانشگاه صنعتی اصفهان دریافت نمود. پس از آن به دوره دکترای مهندسی برق و کامپیوتر در دانشگاه فردوسی مشهد وارد گردید و در سال ۱۳۹۳ موفق به اخذ درجه دکترا در مهندسی کامپیوتر از دانشگاه مذکور گردید. وی از سال ۱۳۸۳ در دانشکده مهندسی کامپیوتر دانشگاه آزاد اسلامی مشهد مشغول به فعالیت گردید و اینک نیز عضو هیأت علمی این دانشکده می‌باشد. زمینه‌های علمی مورد علاقه نام‌برده بهینه‌سازی، یادگیری ماشین و پردازش تصویر است.

میثاق سادات حجازی تحصیلات خود را در مقطع کارشناسی مهندسی فناوری اطلاعات در سال ۱۳۹۶ از دانشگاه آزاد اسلامی واحد مشهد دریافت کرده و هم‌اکنون دانشجوی مقطع کارشناسی ارشد هوش مصنوعی در دانشگاه فردوسی مشهد می‌باشد.

هادی صدوقی یزدی در سال ۱۳۷۳ مدرک کارشناسی مهندسی برق را از دانشگاه فردوسی مشهد و کارشناسی ارشد و دکترای مهندسی برق را به ترتیب در سال‌های ۱۳۷۵ و ۱۳۸۴ از دانشگاه تربیت مدرس دریافت نمود. هم‌اکنون استاد گروه مهندسی کامپیوتر در دانشگاه فردوسی مشهد می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از شناسایی آماری الگو، یادگیری ماشین، بینایی ماشین و پردازش سیگنال.