

ارائه هسته چندجمله‌ای متعامد هرمیت با هدف کاهش بردارهای پشتیبان در طبقه‌بند ماشین بردار پشتیبان (SVM)

وحید هوشمندمقدم^۱، جواد حمیدزاده^۲

^۱ دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بین‌المللی امام رضا (علیه‌السلام)، مشهد
V.hooshmand@imamreza.ac.ir

^۲ عضو هیئت علمی، گروه مهندسی کامپیوتر، موسسه آموزش عالی سجاد، مشهد
J_hamidzadeh@sadjad.ac.ir

چکیده

ماشین بردار پشتیبان روش مطلوبی برای طبقه‌بندی انواع داده‌هاست، اما مشکل اساسی این روش کاهش چشم‌گیر سرعت طبقه‌بندی آن در ازای افزایش ابعاد مسئله و افزایش تعداد نمونه‌ها است. در این مقاله تابع هسته‌ای که از چندجمله‌ای متعامد هرمیت مشتق شده، برای ماشین بردار پشتیبان ارائه شده است. این تابع، علاوه بر کاهش تعداد بردارهای پشتیبان که موجب افزایش سرعت خواهد شد، صحت طبقه‌بندی را افزایش می‌دهد. نتایج آزمایش‌ها نشان می‌دهد که تابع پیشنهادی در مقایسه با سایر توابع هسته رایج برای طبقه‌بندی ماشین بردار پشتیبان، در همه‌ی مجموعه داده‌های مورد آزمایش برگرفته از مجموعه داده واقعی UCI، دارای کم‌ترین تعداد بردارهای پشتیبان بوده و به طور کلی بهترین عملکرد را نسبت به دیگر روش‌ها داراست.

کلمات کلیدی

ماشین بردار پشتیبان (SVM)، تابع هسته، هسته چندجمله‌ای متعامد هرمیت

۱- مقدمه

و یا اصلاح تابع هسته اشاره کرد. البته باید به این موضوع توجه داشت که کاهش تعداد بردارهای پشتیبان می‌تواند باعث کاهش صحت طبقه‌بندی شود. هدف این مقاله ارائه تابع هسته‌ای است که بتواند تعداد بردارهای پشتیبان را کاهش داده و صحت طبقه‌بندی را بهبود ببخشد. حال به بررسی مقالات ارائه‌شده در این زمینه می‌پردازیم. در مقاله [۱]، ماشین بردار پشتیبان را به عنوان یکی از پرکاربردترین روش‌های طبقه‌بندی در زمینه سیستم‌های انتقال هوشمند مطرح کرده است؛ اما نیاز این روش به محاسبات زیاد کارایی آن را تحت تأثیر قرار داده است. روش ماشین بردار پشتیبان به رغم صحت بهتر در طبقه‌بندی داده‌ها، نیازمند محاسبات بالا در مرحله آزمایش نسبت به سایر رقبای موجود مانند شبکه‌های عصبی است. محاسبات در این مرحله متناسب با تعداد بردارهای پشتیبان است. در این مقاله نشان داده‌شده که تعداد بردارهای پشتیبان یک ماشین بردار پشتیبان استاندارد، متناسب با اندازه مسئله داده شده است. افزایش نمونه‌های یادگیری، به منظور افزایش تعمیم طبقه‌بند در کاربرد واقعی مورد استفاده قرار می‌گیرند، این امر نیازمند زمان اجرای بالاتری در مرحله آزمایش است؛ بنابراین، بلافاصله پس از انتشار ماشین بردار پشتیبان، روش‌های بسیاری برای کاهش تعداد بردارهای پشتیبان در یک ماشین بردار پشتیبان استاندارد بدون کاهش

از ابزارهای پرکاربرد برای اهداف رگرسیون و طبقه‌بندی، ماشین بردار پشتیبان است که کارایی عمومی آن به استفاده از دو گام عمده بستگی دارد. گام اول، ساخت تابع هزینه است که باید بهینه بوده و محدودیت‌هایی برای آن معین شود و مورد دوم استفاده از توابع هسته که داده ورودی را به فضای ویژگی با ابعاد بالاتر نگاشت می‌کند. به طوری که بتوان داده‌ها را به طور خطی جدا کرد که در این مقاله تمرکز بر روی گام دوم کارایی ماشین بردار پشتیبان است.

روش ماشین بردار پشتیبان به رغم صحت بالاتر در طبقه‌بندی داده‌ها، نیازمند محاسبات بالا در مقایسه با روش‌های دیگر است، در نتیجه زمان اجرای بالاتری را نسبت به آن‌ها دارد. این محاسبات، متناسب با تعداد بردارهای پشتیبان است. از طرف دیگر اندازه مسئله ارتباط مستقیمی با تعداد بردارهای پشتیبان دارد [۱]؛ بنابراین هرچه مسئله بزرگ‌تر باشد، تعداد بردارهای پشتیبان افزایش یافته و موجب افزایش میزان محاسبات می‌شود و در نتیجه زمان اجرا بیش‌تر خواهد شد؛ بنابراین اهمیت توسعه روش‌هایی برای کاهش تعداد بردارهای پشتیبان به خوبی قابل درک است. از جمله این روش‌ها، می‌توان به تغییر