运筹学与控制论

DOI:10.11721/cqnuj20210116

基于二阶锥规划的鲁棒投影支持向量机*

吴至友，尹　林
(重庆师范大学 数学科学学院，重庆 401331)

摘要:【目的】投影支持向量机是通过将线性判别分析的思想应用到双子支持向量机,从而提出的一种新的非平行分类模型,旨在寻找两个不平行的投影方向而非超平面。然而该模型不够鲁棒,当训练数据集中存在大量的异常点或噪声时,投影支持向量机所学习的投影方向往往会受到影响而发生偏移,从而分类性能下降,需要进一步降低模型对异常点或噪声的敏感性,提升模型的鲁棒性。【方法】在模型中引入机会约束,在投影空间中允许部分投影样本到它的样本中心投影的距离大于它到另一类样本中心的投影距离,即给出了分错样本概率的一个上界。【结果】得到一个新的带有机会约束的鲁棒投影支持向量机,并等价地转化为二阶锥规划问题,从而只需求解一对线性二阶锥规划问题即可训练出两个非平行投影方向。【结论】在有关 UCI 数据集以及增加噪声的该数据集的数值实验中,上述基于二阶锥规划的鲁棒投影支持向量机与其他算法相比,准确率变化很小,相对稳定,具有更好的鲁棒性和泛化能力。

关键词:投影支持向量机;机会约束;二阶锥规划;鲁棒性
中图分类号:O221　　**文献标志码**:A　　**文章编号**:1672-6693(2021)01-0001-10

支持向量机(SVM)[1-4]作为一种基于核的经典分类方法,越来越受到人们的关注并被广泛应用于文本识别、人脸识别、时间序列预测等[5-8]多个领域。SVM 的主要思想是通过极小化结构风险寻求两个平行的分离超平面,从而尽可能地将正负两类样本点分开。但随着社会的发展,数据复杂度急剧增加,SVM 应用的劣势逐渐显现:它在训练过程中需要求解一个大规模的二次规划问题,这将大大增加它的计算成本。为了提升 SVM 的泛化表现,一些非平行支持向量机相继被提出,包括临近支持向量机(PSVM)[9]、稀疏临近支持向量机(SPSVM)[10]、双子支持向量机(TWSVM)[11-12]等。与 SVM 不同,TWSVM 是求解两个规模较小、结构与 SVM 相似的二次规划问题,所以 TWSVM 的计算速度比 SVM 快了大约 4 倍。由于具有优异的性能,TWSVM 已被应用于医学检测等多个领域。但是,TWSVM 仅考虑了经验风险,并且在求解过程中可能出现奇异矩阵。为了解决这个问题,双边界支持向量机(BTWSVM)[13]在模型中引入了一个正则项,不仅实现了结构风险最小化,而且有效避免了奇异矩阵的出现。而基于二阶锥规划的双子支持向量机(SOCP-TSVM)[14]则增强了 TWSVM 的鲁棒性,使它不受异常点或噪声的影响,从而找到最优的两个不平行的超平面。

与 TWSVM 不同的是,投影双子支持向量机(PTSVM)[15]是为每一个类寻找一个投影方向,而不是传统的寻找两个不平行的超平面;这使得该类别的样本投影在各自的投影子空间中,并聚集在均值投影的周围,且与另一类的投影样本尽量远离。由于 PTSVM 在训练时只考虑了经验风险极小化,因此会出现奇异矩阵和过拟合的情况。为此,带有正则项投影支持向量机(RPTSVM)[16]则在 PTSVM 的模型中引入一个正则项来克服上述缺陷。然而,关于提升 PTSVM 的鲁棒性的研究成果到目前为止还很少,其中之一为 V 投影双子支持向量机(V-PTSVM)[17]。而事实上,当训练集中存在异常点或噪声时,PTSVM 获得的投影方向往往不是最优的投影方向。为了降低 PTSVM 对噪声的敏感性,受 SOCP-TSVM 的启发,本文将引入机会约束,在投影空间中允许部分投影样本到它样本中心投影的距离大于它到另一类样本中心的投影距离,即通过机会约束给出了分错样本概率的一

* **收稿日期**:2020-11-05　**网络出版时间**:2021-03-16 16:25
资助项目:国家自然科学基金(No. 11871128;No. 11991024)
第一作者简介:吴至友,女,教授,博士生导师,研究方法为最优化理论与算法,E-mail:zywu@cqnu.edu.cn;**通信作者**:尹林,女,E-mail:yinlinxue0@163.com
网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20210316.1611.002.html

个上界,由此使得该模型在训练存在大量异常点的数据集时不受异常点或噪声影响,找到最优的投影方向,从而提升模型的鲁棒性。

1 SVM 的相关工作

本节首先介绍将使用的部分符号,然后再简要回顾 TWSVM,SOCP-TSVM 以及 PTSVM。在本文中,所有的向量均为列向量,除非通过$(\cdot)^{\mathrm{T}}$转置为行向量。$\boldsymbol{e}$表示分量全为1的向量,$\boldsymbol{I}$表示单位矩阵,此外,$\|\cdot\|$表示二范数。本文训练数据集表示为$T=\{(\boldsymbol{x}_j^{(i)},y_j)\mid i=1,2;j=1,\cdots,m_i\}$,其中正负类点的样本个数分别为$m_1,m_2$,且$m=m_1+m_2$,而$\boldsymbol{x}_j^{(i)}\in\mathbf{R}^n$表示第$i$类的第$j$个样本点,$y_j\in\{+1,-1\}$表示相应的样本标签。为了方便起见,用$k=1,2$分别表示正类或者负类,$\boldsymbol{A}\in\mathbf{R}^{m_1\times n}$是正类样本点观测值矩阵,$\boldsymbol{B}\in\mathbf{R}^{m_2\times n}$是负类样本点观测值矩阵。

1.1 TWSVM

TWSVM 旨在通过求解两个较小规模的二次规划问题获得两个不平行的超平面:

$$\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}+b_1=0,\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}+b_2=0。$$

BTWSVM[13]对 TWSVM 做了进一步改进,在目标函数中考虑了结构风险,它的模型为:

$$\begin{aligned}\min_{\boldsymbol{w}_1,b_1,\boldsymbol{\xi}_2}\quad & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}_1b_1\|^2+\frac{C_1}{2}(\|\boldsymbol{w}_1\|^2+b_1^2)+C_3\boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{\xi}_2,\\ \text{s.t.}\quad & -(\boldsymbol{B}\boldsymbol{w}_1+b_1)\geqslant\boldsymbol{e}_2-\boldsymbol{\xi}_2,\boldsymbol{\xi}_2\geqslant0。\end{aligned}\tag{1}$$

$$\begin{aligned}\min_{\boldsymbol{w}_2,b_2,\boldsymbol{\xi}_1}\quad & \frac{1}{2}\|\boldsymbol{B}\boldsymbol{w}_2+\boldsymbol{e}_2b_2\|^2+\frac{C_2}{2}(\|\boldsymbol{w}_2\|^2+b_2^2)+C_4\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{\xi}_1,\\ \text{s.t.}\quad & (\boldsymbol{A}\boldsymbol{w}_2+b_2)\geqslant\boldsymbol{e}_1-\boldsymbol{\xi}_1,\boldsymbol{\xi}_1\geqslant0。\end{aligned}\tag{2}$$

其中:$C_1,C_2,C_3,C_4>0$是参数。令$\boldsymbol{H}=[\boldsymbol{A},\boldsymbol{e}_1],\boldsymbol{G}=[\boldsymbol{B},\boldsymbol{e}_2]$,则问题(1)和问题(2)的对偶问题如下:

$$\begin{aligned}\max_{\boldsymbol{\alpha}}\quad & \boldsymbol{e}_2^{\mathrm{T}}\alpha-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+C_1\boldsymbol{I})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\alpha},\\ \text{s.t.}\quad & \boldsymbol{0}\leqslant\boldsymbol{\alpha}\leqslant\boldsymbol{C}_1。\end{aligned}$$

$$\begin{aligned}\max_{\boldsymbol{\gamma}}\quad & \boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{\gamma}-\frac{1}{2}\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+C_2\boldsymbol{I})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\gamma},\\ \text{s.t.}\quad & \boldsymbol{0}\leqslant\boldsymbol{\gamma}\leqslant\boldsymbol{C}_2。\end{aligned}$$

因此通过解上述对偶问题即可获得如下两个解:

$$\begin{aligned}\boldsymbol{v}_1&=-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+C_1\boldsymbol{I})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\alpha},\\ \boldsymbol{v}_2&=-(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+C_2\boldsymbol{I})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\gamma}。\end{aligned}$$

其中:$\boldsymbol{v}_k=[\boldsymbol{w}_k,b_k]^{\mathrm{T}}\in\mathbf{R}^{n+1},k=1,2$。

一个新的样本$\boldsymbol{x}\in\mathbf{R}^n$被分为正类或者负类,依赖于它到两个超平面的距离,靠哪个超平面更近就被分为哪一类,即决策函数如下:

$$y=\arg\min_{k=1,2}\frac{|\boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{x}+b_k|}{\|\boldsymbol{w}_k\|}。$$

1.2 SOCP-TSVM

设$\boldsymbol{X}_1,\boldsymbol{X}_2$分别表示产生正负类样本点的随机向量,它们分别服从$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$和$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$分布,其中$\boldsymbol{\mu}_k$为均值,$\boldsymbol{\Sigma}_k\in\mathbf{R}^{n\times n}$为协方差矩阵,$\boldsymbol{\Sigma}_k$为对称半正定矩阵$(k=1,2)$。

为提升 TWSVM 的鲁棒性,文献[14]给出带有机会约束的 TWSVM:

$$\begin{aligned}\min_{\boldsymbol{w}_1,b_1}\quad & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}_1b_1\|^2+\frac{C_1}{2}(\|\boldsymbol{w}_1\|^2+b_1^2),\\ \text{s.t.}\quad & \inf_{\boldsymbol{X}_2\sim(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)}\Pr\{\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{X}_2+b_1\leqslant-1\}\geqslant\eta_2。\end{aligned}\tag{3}$$

$$\begin{aligned}\min_{\boldsymbol{w}_2,b_2}\quad & \frac{1}{2}\|\boldsymbol{B}\boldsymbol{w}_2+\boldsymbol{e}_2b_2\|^2+\frac{C_2}{2}(\|\boldsymbol{w}_2\|^2+b_2^2),\\ \text{s.t.}\quad & \inf_{\boldsymbol{X}_1\sim(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)}\Pr\{\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{X}_1+b_2\geqslant1\}\geqslant\eta_1。\end{aligned}\tag{4}$$

问题(3)中的约束首先给出了真阴性概率的一个下界 η_2。类似地，问题(4)中的机会约束给出真阳性的概率下界为 η_1。再通过多元切比雪夫不等式[18]，将上述问题转化为二阶锥规划问题[19]，即：

$$\begin{aligned}&\min_{\boldsymbol{w}_1,b_1}\quad \frac{1}{2}\|\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}_1b_1\|^2+\frac{C_1}{2}(\|\boldsymbol{w}_1\|^2+b_1^2),\\&\text{s.t.}\quad -\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_2-b_1\geqslant 1+l_2\|\boldsymbol{S}_2^{\mathrm{T}}\boldsymbol{w}_1\|。\end{aligned}\tag{5}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_2,b_2}\quad \frac{1}{2}\|\boldsymbol{B}\boldsymbol{w}_2+\boldsymbol{e}_2b_2\|^2+\frac{C_2}{2}(\|\boldsymbol{w}_2\|^2+b_2^2),\\&\text{s.t.}\quad \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_1+b_2\geqslant 1+l_1\|\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{w}_2\|。\end{aligned}\tag{6}$$

这里的 $l_k=\sqrt{\dfrac{\eta_k}{1-\eta_k}}$，$\boldsymbol{\Sigma}_k=\boldsymbol{S}_k\boldsymbol{S}_k^{\mathrm{T}}$，$k=1,2$，$\boldsymbol{S}_k$ 为对称半正定矩阵 $\boldsymbol{\Sigma}_k$ 分解得到的方阵。

注意，设 $\boldsymbol{v}_k=[\boldsymbol{w}_k^{\mathrm{T}},b_k]^{\mathrm{T}}\in\mathbf{R}^{n+1}$，$\boldsymbol{H}=[\boldsymbol{A},\boldsymbol{e}_1]\in\mathbf{R}^{m_1\times(n+1)}$，$\boldsymbol{G}=[\boldsymbol{B},\boldsymbol{e}_2]\in\mathbf{R}^{m_2\times(n+1)}$，问题(5)和问题(6)的目标函数可以分别写为：

$$\frac{1}{2}\|\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}_1b_1\|^2+\frac{C_1}{2}(\|\boldsymbol{w}_1\|^2+b_1^2)=\frac{1}{2}\boldsymbol{v}_1^{\mathrm{T}}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+C_1\boldsymbol{I})\boldsymbol{v}_1,$$

$$\frac{1}{2}\|\boldsymbol{B}\boldsymbol{w}_2+\boldsymbol{e}_2b_2\|^2+\frac{C_2}{2}(\|\boldsymbol{w}_2\|^2+b_2^2)=\frac{1}{2}\boldsymbol{v}_2^{\mathrm{T}}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+C_2\boldsymbol{I})\boldsymbol{v}_2。$$

再通过引入新的变量 t_1 和 t_2，问题(5)和问题(6)可以转化为以下两个线性二阶锥规划问题：

$$\begin{aligned}&\min_{\boldsymbol{w}_1,b_1,t_1}\quad t_1,\\&\text{s.t.}\quad -\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_2-b_1\geqslant 1+l_2\|\boldsymbol{S}_2^{\mathrm{T}}\boldsymbol{w}_1\|,\\&\qquad\|(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+C_1\boldsymbol{I})^{\frac{1}{2}}\boldsymbol{v}_1\|\leqslant t_1。\end{aligned}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_2,b_2,t_2}\quad t_2,\\&\text{s.t.}\quad \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_1+b_2\geqslant 1+l_1\|\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{w}_2\|,\\&\qquad\|(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+C_2\boldsymbol{I})^{\frac{1}{2}}\boldsymbol{v}_2\|\leqslant t_2。\end{aligned}$$

进而用求解线性二阶锥规划的求解器进行求解上述两个问题，例如 SeDuMi 工具箱[20]等。SOCP-TSVM 的决策函数为：

$$y=\arg\min_{k=1,2}\frac{|\boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{x}+b_k|}{\|\boldsymbol{w}_k\|}。$$

1.3 PTSVM

与 TWSVM 寻找两个超平面不同的是，PTSVM 的思想旨在寻找两个不平行的投影方向，每个类别一个投影方向，各个类别的投影样本在各自的子空间中类内方差极小化，同时与另一个类别的投影样本很好地分离。PTSVM 模型如下：

$$\begin{aligned}&\min_{\boldsymbol{w}_1,\xi_l}\quad \frac{1}{2}\sum_{i=1}^{m_1}\left(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_i^{(1)}-\boldsymbol{w}_1^{\mathrm{T}}\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}\right)^2+C_1\sum_{l=1}^{m_2}\xi_l,\\&\text{s.t.}\quad \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_l^{(2)}-\boldsymbol{w}_1^{\mathrm{T}}\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}+\xi_l\geqslant 1,\xi_l\geqslant 0,l=1,2,\cdots,m_2。\end{aligned}\tag{7}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_2,\eta_l}\quad \frac{1}{2}\sum_{i=1}^{m_2}\left(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}_i^{(2)}-\boldsymbol{w}_2^{\mathrm{T}}\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)^2+C_2\sum_{l=1}^{m_1}\eta_l,\\&\text{s.t.}\quad -\left(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}_l^{(1)}-\boldsymbol{w}_2^{\mathrm{T}}\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)+\eta_l\geqslant 1,\eta_l\geqslant 0,l=1,2,\cdots,m_1。\end{aligned}\tag{8}$$

式中：ξ_l，η_l 分别表示正负两类投影样本点的非负松弛变量，$\boldsymbol{w}_k$ 表示第 $k\in\{1,2\}$ 类的投影方向。问题(7)的目标函数表示了第 1 类投影样本点到投影中心的距离以及误差尽可能小，约束条件表示第 2 类投影样本点到第 1 类投影中心的距离至少为 1，问题(8)与问题(7)类似。

然而，PTSVM 只关注了经验风险最小化，这会导致训练数据集时出现过拟合的情况以及矩阵病态问题。为

了解决上述问题,文献[16]对PTSVM做了进一步改进,在问题(7)和问题(8)的目标函数中加入了正则项来实现结构风险最小化,有:

$$\begin{aligned}&\min_{\boldsymbol{w}_1,\xi_l}\quad \frac{C_1}{2}\sum_{i=1}^{m_1}\left(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_i^{(1)}-\boldsymbol{w}_1^{\mathrm{T}}\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}\right)^2+\frac{1}{2}\|\boldsymbol{w}_1\|^2+C_3\sum_{l=1}^{m_2}\xi_l,\\&\text{s.t.}\quad \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_l^{(2)}-\boldsymbol{w}_1^{\mathrm{T}}\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}+\xi_l\geqslant 1,\xi_l\geqslant 0,l=1,2,\cdots,m_2。\end{aligned}\tag{9}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_2,\eta_l}\quad \frac{C_2}{2}\sum_{i=1}^{m_2}\left(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}_i^{(2)}-\boldsymbol{w}_2^{\mathrm{T}}\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)^2+\frac{1}{2}\|\boldsymbol{w}_2\|^2+C_4\sum_{l=1}^{m_1}\eta_l,\\&\text{s.t.}\quad -\left(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}_l^{(1)}-\boldsymbol{w}_2^{\mathrm{T}}\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)+\eta_l\geqslant 1,\eta_l\geqslant 0,l=1,2,\cdots,m_1。\end{aligned}\tag{10}$$

其中:$C_1,C_2,C_3,C_4>0$是权重参数。为了简化公式,定义类内协方差矩阵如下:

$$\boldsymbol{H}_1=\sum_{i=1}^{m_1}\left(\boldsymbol{x}_i^{(1)}-\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}\right)\left(\boldsymbol{x}_i^{(1)}-\frac{1}{m_1}\sum_{j=1}^{m_1}\boldsymbol{x}_j^{(1)}\right)^{\mathrm{T}},$$

$$\boldsymbol{H}_2=\sum_{i=1}^{m_2}\left(\boldsymbol{x}_i^{(2)}-\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)\left(\boldsymbol{x}_i^{(2)}-\frac{1}{m_2}\sum_{j=1}^{m_2}\boldsymbol{x}_j^{(2)}\right)^{\mathrm{T}}。$$

于是问题(9)和问题(10)可写为:

$$\begin{aligned}&\min_{\boldsymbol{w}_1,\boldsymbol{\xi}}\quad \frac{C_1}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{w}_1+\frac{1}{2}\|\boldsymbol{w}_1\|^2+C_3\boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{\xi},\\&\text{s.t.}\quad \boldsymbol{B}\boldsymbol{w}_1-\frac{1}{m_1}\boldsymbol{e}_2\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{\xi}\geqslant\boldsymbol{e}_2,\boldsymbol{\xi}\geqslant 0。\end{aligned}\tag{11}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_1,\boldsymbol{\eta}}\quad \frac{C_2}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{w}_2+\frac{1}{2}\|\boldsymbol{w}_2\|^2+C_4\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{\eta},\\&\text{s.t.}\quad -\left(\boldsymbol{A}\boldsymbol{w}_2-\frac{1}{m_2}\boldsymbol{e}_1\boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{B}\boldsymbol{w}_2\right)+\boldsymbol{\eta}\geqslant\boldsymbol{e}_1,\boldsymbol{\eta}\geqslant 0。\end{aligned}\tag{12}$$

在获得解$\boldsymbol{w}_1,\boldsymbol{w}_2$之后,一个未知的实例$\boldsymbol{x}$被分配第$k$类,这取决于它最接近哪个投影中心,即决策函数为:

$$y=\arg\min_{k=1,2}\left|\boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{w}_k^{\mathrm{T}}\frac{1}{m_k}\sum_{j=1}^{m_k}\boldsymbol{x}_j^{(k)}\right|。$$

2 基于二阶锥规划的鲁棒投影支持向量机(SOCP-PTSVM)

文献[16]对PTSVM加入正则实现了结构风险极小化以及解决了过拟合的情况,即问题(11)和问题(12),但该模型鲁棒性不强,在训练数据时易受异常点的影响,投影方向往往会发生偏移。机会约束可以用来提升分类器的鲁棒性,已被成功启用到SVM以及TWSVM。本节将该方法推广到PTSVM中,提出一个新的模型即SOCP-PTSVM。令$\boldsymbol{X}_1,\boldsymbol{X}_2$分别服从$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$和$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$分布,为正负类观测值的随机向量,对问题(11)和(12)引入机会约束得到新的模型如下:

$$\begin{aligned}&\min_{\boldsymbol{w}_1}\quad \frac{1}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{w}_1+\frac{C_1}{2}\|\boldsymbol{w}_1\|^2,\\&\text{s.t.}\quad \inf_{\boldsymbol{X}_2\sim(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)}\Pr((\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{X}_2-\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_1)\geqslant 1)\geqslant\eta_2。\end{aligned}\tag{13}$$

$$\begin{aligned}&\min_{\boldsymbol{w}_2}\quad \frac{1}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{w}_2+\frac{C_2}{2}\|\boldsymbol{w}_2\|^2,\\&\text{s.t.}\quad \inf_{\boldsymbol{X}_1\sim(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)}\Pr(-(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{X}_1-\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_2)\geqslant 1)\geqslant\eta_1。\end{aligned}\tag{14}$$

其中:$C_1,C_2>0$为调节结构风险的参数,而$\eta_1,\eta_2\in(0,1)$与SOCP-TSVM中的作用类似。在问题(13)中,目标函数的第1项表示在投影子空间中,正类样本投影点到投影中心的距离;第2项为正则项,防止分类器过拟合,实现结构风险极小化;约束条件则指出负类样本投影点到正类中心投影的距离至少为1的概率下界为η_1,降低

了 SOCP-PTSVM 对异常点或噪声的敏感性，从而具有更好的鲁棒性。问题(14)与问题(13)类似，故不再赘述。为求解该模型，利用多元切比雪夫不等式，上述问题可等价地写成如下 SOCP-PTSVM：

$$\begin{aligned}\min_{\boldsymbol{w}_1}\quad & \frac{1}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{w}_1+\frac{C_1}{2}\|\boldsymbol{w}_1\|^2,\\ \text{s.t.}\quad & \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_2-\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_1\geqslant 1+l_2\|\boldsymbol{S}_2^{\mathrm{T}}\boldsymbol{w}_1\|。\end{aligned}\tag{15}$$

$$\begin{aligned}\min_{\boldsymbol{w}_2}\quad & \frac{1}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{w}_2+\frac{C_2}{2}\|\boldsymbol{w}_2\|^2,\\ \text{s.t.}\quad & -(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_1-\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_2)\geqslant 1+l_1\|\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{w}_2\|。\end{aligned}\tag{16}$$

此时 $l_k=\sqrt{\dfrac{\eta_k}{1-\eta_k}}$，$\boldsymbol{\Sigma}_k=\boldsymbol{S}_k\boldsymbol{S}_k^{\mathrm{T}}$，其中 $k=1,2$。

值得注意的是，可以将问题(15)和问题(16)的目标函数写为如下形式：

$$\frac{1}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{H}_1\boldsymbol{w}_1+\frac{C_1}{2}\|\boldsymbol{w}_1\|^2=\frac{1}{2}\boldsymbol{w}_1^{\mathrm{T}}[\boldsymbol{H}_1+C_1\boldsymbol{I}_1]\boldsymbol{w}_1,$$

$$\frac{1}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{H}_2\boldsymbol{w}_2+\frac{C_2}{2}\|\boldsymbol{w}_2\|^2=\frac{1}{2}\boldsymbol{w}_2^{\mathrm{T}}[\boldsymbol{H}_2+C_2\boldsymbol{I}_2]\boldsymbol{w}_2。$$

利用 Cholessky 分解 $\boldsymbol{H}_1+C_1\boldsymbol{I}_1=\boldsymbol{L}_1^{\mathrm{T}}\boldsymbol{L}_1$，$\boldsymbol{H}_2+C_2\boldsymbol{I}_2=\boldsymbol{L}_2^{\mathrm{T}}\boldsymbol{L}_2$，并引入两个新的变量 g_1，g_2，问题(15)和(16)可以转化为如下两个二阶锥规划问题：

$$\begin{aligned}\min_{\boldsymbol{w}_1,g_1}\quad & g_1,\\ \text{s.t.}\quad & \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_2-\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_1\geqslant 1+l_2\|\boldsymbol{S}_2^{\mathrm{T}}\boldsymbol{w}_1\|,\\ & \left\|\begin{matrix}g_1-1\\ \sqrt{2}\boldsymbol{L}_1\boldsymbol{w}_1\end{matrix}\right\|\leqslant g_1+1。\end{aligned}\tag{17}$$

$$\begin{aligned}\min_{\boldsymbol{w}_2,g_2}\quad & g_2,\\ \text{s.t.}\quad & -(\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_1-\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_2)\geqslant 1+l_1\|\boldsymbol{S}_1^{\mathrm{T}}\boldsymbol{w}_2\|,\\ & \left\|\begin{matrix}g_2-1\\ \sqrt{2}\boldsymbol{L}_2\boldsymbol{w}_2\end{matrix}\right\|\leqslant g_2+1。\end{aligned}\tag{18}$$

问题(17)和(18)均是带有两个二阶锥约束，目标函数是线性的二阶锥规划问题，可以使用 SeDuMi 进行有效求解。综上所述，本文给出如下关于求解 SOCP-PTSVM 的算法。

算法 1　步骤 1：输入训练集 $T=\{(\boldsymbol{x}_j^{(i)},y_j)\mid i=1,2;j=1,\cdots,m_i\}$，这里 $\boldsymbol{x}_j^{(i)}\in\mathbf{R}^n$，以及 $y_j\in\{+1,-1\}$。

步骤 2：通过五折交叉验证的方法选定参数 $C_1,C_2>0$，$\eta_1,\eta_2\in(0,1)$。

步骤 3：计算 $\boldsymbol{L}_k$，$\boldsymbol{\mu}_k$，$\boldsymbol{S}_k$，使用 SeDuMi 求解器求解问题(17)和(18)得到投影方向 $\boldsymbol{w}_k$，$k=1,2$。

步骤 4：构建决策函数如下：

$$f_1(\boldsymbol{x})=\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{\mu}_1,$$

$$f_2(\boldsymbol{x})=\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{\mu}_2。$$

步骤 5：输出 y，即新样本 $\boldsymbol{x}$ 属于的类别：

$$y=\arg\min_{k=1,2}|f_k(\boldsymbol{x})|。$$

值得注意的是，在算法 1 中，步骤 2 是选择 SOCP-PTSVM 的最优参数，即限定每一类样本被分对的概率下界分别为 η_1，η_2，即允许将部分样本分错，特别是异常点，从而提高了 SOCP-PTSVM 的鲁棒性。算法 1 对于线性二阶锥规划是全局超线性收敛的，证明详情参考文献[19,21]。

3 数值实验

本节将在真实数据集上将 SOCP-PTSVM 与几个鲁棒分类器进行比较，包括：SOCP-TSVM[14]，鲁棒双子支持向量机[22](R-TWSVM)，鲁棒二阶锥非平行支持向量机[23](NPSVM)以及 V-PTSVM[17]。实验环境为：Intel

(R) Core(TM)i7-4790 (3.60 GHz)处理器,32 GB内存,使用Matlab 2018实现代码。利用五折交叉验证的方法训练超参数,所有模型中的参数C_1,C_2选择范围为:

$$C_1=C_2\in\{2^{-7},2^{-6},2^{-5},2^{-4},2^{-3},2^{-2},2^{-1},1,2^1,2^2,2^3,2^4,2^5,2^6,2^7\}。$$

SOCP-PTSVM,SOCP-TSVM,NPSVM的参数η_1,η_2以及V-PTSVM的参数v_1,v_2取值范围为{0.2,0.4,0.6,0.8},NPSVM的参数$\varepsilon_1=\varepsilon_2\in\{0.05,0.1,0.2,0.3,0.4,0.5\}$。准确率($R_{\mathrm{Acc}}$)作为评价算法性能的指标,可由真正率($R_{\mathrm{TP}}$)、真负率($R_{\mathrm{TN}}$)、假正率($R_{\mathrm{FP}}$)以及假负率($R_{\mathrm{FN}}$)计算得出,即:

$$R_{\mathrm{Acc}}=\frac{R_{\mathrm{TP}}+R_{\mathrm{TN}}}{R_{\mathrm{TP}}+R_{\mathrm{TN}}+R_{\mathrm{FP}}+R_{\mathrm{FN}}}。$$

所有数据集进行20次实验,并将正确率的平均值作为最终实验结果。

在UCI公布的数据集[24]中选取9个真实数据集Ionosphere,WPBC,Toes,Australian,Echocardiogra,Vehicle1,Spilce,Glass1和Chess进行了实验比较。样本个数及特征个数见表1。

表1 实验中使用的数据集

Tab.1 Benchmark datasets used in experiments

数据集	样本	特征	数据集	样本	特征	数据集	样本	特征
Ionosphere	351	34	Australian	690	14	Spilce	3 175	60
WPBC	198	33	Echocardiogra	74	10	Glass1	214	9
Toes	40	227	Vehicle1	846	18	Chess	3 196	36

表2给出了上述5个算法在9个数据集上的实验结果。从表2可以看出SOCP-PTSVM在7个数据集上的R_{Acc}最高,而其他4个算法仅在1个数据集上的R_{Acc}最高。在Spilce和Chess上SOCP-PTSVM的R_{Acc}比其他4个算法中最高R_{Acc}略低:在Spilce数据集上V-PTSVM的R_{Acc}最高(81.9%),而SOCP-PTSVM的R_{Acc}为81.5%,仅少0.4%,并明显高于SOCP-TSVM,R-TWSVM,NPSVM这几个鲁棒分类器的结果;同样地,NPSVM在Chess数据集上的R_{Acc}最高,但SOCP-PTSVM在该数据集上的R_{Acc}仅比它低0.16%,且高于其他几个算法。因此,表2显示的结果说明SOCP-PTSVM的分类性能具有一定优势。在训练时间方面,本文所提出的方法SOCP-PTSVM表现也很好,在所有训练集上的时间均不超过0.2 s,这得益于SOCP-PTSVM仅需求解两个线性SOCP。

表2 5个算法的R_{Acc}及训练时间

Tab.2 The R_{Acc} and training time of the five algorithms

数据集	SOCP-TSVM		R-TWSVM		NPSVM		V-PTSVM		SOCP-PTSVM	
	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s
Ionosphere	87.10	0.069	84.62	0.151	87.81	0.226	89.24	0.041	**93.35**	0.031
WPBC	76.43	0.066	76.59	0.103	76.26	0.146	76.92	0.031	**84.43**	0.057
Toes	55.40	0.112	61.53	0.125	62.89	0.640	65.50	0.013	**73.42**	0.211
Australian	85.69	0.046	74.56	0.112	87.97	0.638	65.40	1.374	**88.10**	0.040
Echocardiogra	84.92	0.034	87.26	0.089	87.94	0.093	88.09	0.011	**91.50**	0.049
Vehicle1	79.67	0.020	72.24	0.239	78.57	0.528	78.77	2.061	**80.07**	0.037
Spilce	81.11	0.071	64.09	0.480	80.68	0.787	81.90	3.400	81.50	0.040
Glass1	66.92	0.034	64.84	0.118	68.34	0.102	66.94	0.033	**68.72**	0.029
Chess	94.31	0.224	60.37	1.672	95.78	8.292	94.93	10.845	95.62	0.049

注:加粗数字为同一数据集中R_{Acc}最优结果,下同

为了比较 SOCP-PTSVM 与另外 4 种算法的差异性，采用配对 t 检验(Paired t-test)进行测试，用 R_{Acc} 作为度量算法性能的标准，当检验结果的 p 值小于 0.1 时，两种算法有统计学意义上的差异。从表 3 可知，SOCP-PTSVM 与另外 4 种算法在大部分数据集上的 R_{Acc} 配对 t 检验 p 值都小于 0.1，从而说明算法 SOCP-PTSVM 与其他 4 种算法明显不同。从 SOCP-PTSVM 分别与 SOCP-TSVM，R-TWSVM，NPSVM 在各数据集上的 R_{Acc} 配对 t 检验 p 值可以看出：SOCP-PTSVM 与 R-TWSVM 在 3 个数据集上的 R_{Acc} 配对 t 检验 p 值接近于 0.1，在其他 6 个数据集上的 R_{Acc} 配对 t 检验 p 值则远远小于 0.1，说明 SOCP-PTSVM 与 R-TWSVM 存在统计学意义上的差异。

表 3　SOCP-PTSVM 与其他 4 个算法的配对 t 检验 p 值比较

Tab. 3　Comparison of p value of paired t test on SOCP-PTSVM and another four algorithms

数据集	配对 t 检验 p 值			
	SOCP-PTSVM 与 SOCP-TSVM	SOCP-PTSVM 与 R-TWSVM	SOCP-PTSVM 与 NPSVM	SOCP-PTSVM 与 V-PTSVM
Ionosphere	0.016 265	9.73×10^{-5}	0.043 523	0.007 737
WPBC	0.972 047	0.972 047	0.972 047	0.404 815
Toes	0.034 565	0.151 025	0.126 098	0.025 887
Australian	0.227 115	1.05×10^{-5}	0.252 535	8.70×10^{-8}
Echocardiogra	0.106 482	0.059 885	0.059 885	0.054 084
Vehicle1	0.430 215	0.005 279	0.074 301	0.096 81
Spilce	0.905 324	1.85×10^{-5}	0.485 237	0.065 362
Glass1	0.235 039	0.619 384	0.672 079	0.957 8
Chess	2.14×10^{-14}	0.028 591	4.85×10^{-15}	1.82×10^{-14}

此外从 SOCP-PTSVM 与 V-PTSVM 在各数据集上的 R_{Acc} 配对 t 检验 p 值可以看出：虽然它们都是投影支持向量机，但在其中 7 个数据集上的 R_{Acc} 配对 t 检验 p 值远远小于 0.1，这说明二者存在统计学意义上的差异。再比较 SOCP-PTSVM 与 V-PTSVM 加噪前后 R_{Acc} 的变化(表 2～表 5)，得出 SOCP-PTSVM 在大部分的数据集上的 R_{Acc} 优于 V-PTSVM 的 R_{Acc}，且 SOCP-PTSVM 在噪声的影响下，R_{Acc} 变化较小，可以说明 SOCP-PTSVM 的泛化性能要优于 V-PTSVM。

表 4 和表 5 分别显示了对 9 个数据集加入噪声后，5 个算法训练得到的 R_{Acc} 以及 SOCP-PTSVM 与另外 4 种算法 R_{Acc} 配对 t 检验 p 值。这里引入了 20%的噪声，采用的是均值为 0，方差为 0.01 的正态分布。从表 4 中可以看到 SOCP-PTSVM 在加噪后的 8 个数据集上 R_{Acc} 达到最优。在 WPBC 上，SOCP-PTSVM 的 R_{Acc} 比其他算法高了接近 9%，在 Ionosphere 和 Toes 上 SOCP-PTSVM 的 R_{Acc} 也高于其他算法至少 4%。对于其他算法而言，仅有 V-PTSVM 在 Glass1 上的分类性能表现最好，在这一数据集上 SOCP-PTSVM 的 R_{Acc} 仅比 V-PTSVM 的 R_{Acc} 少 2%；并且与加噪前的情况相比，SOCP-PTSVM 的 R_{Acc} 仅有 0.49%的变化，而 V-PTSVM 的 R_{Acc} 则有 3.86%的变化。这表明了 SOCP-PTSVM 的鲁棒性比 V-PTSVM 强，同时 SOCP-PTSVM 的 R_{Acc} 明显高于 SOCP-TSVM，R-TWSVM，NPSVM 这几个鲁棒分类器的 R_{Acc}。而且，SOCP-PTSVM 在数据集加噪前后的 R_{Acc} 变化是平稳的(表 2，表 4)，变化幅度在 1%左右；其中该算法在 Toes，Australian，Vehicle1，Spilce，Glass1 加噪前后的 R_{Acc} 几乎没有明显改变。这是因为模型中加入了机会约束，使模型更具鲁棒性。其他算法在数据集加噪前后的 R_{Acc} 变化较大，这也说明了在有噪声的情况下，SOCP-PTSVM 的分类性能没有变化太大的同时，也优于其他算法。从表 5 可知，SOCP-PTSVM 与 V-PTSVM 比较在 5 个数据集上的 R_{Acc} 配对 t 检验 p 值都低于 0.1，与 SOCP-TSVM，R-TWSVM，NPSVM 相比至少有 4 个数据集上的 R_{Acc} 配对 t 检验 p 值低于 0.1。所以，加噪后的 SOCP-PTSVM 不仅具有稳定的泛化性能，而且与其他 4 个鲁棒分类器有着明显差异。

表4 加噪后5个算法的R_{Acc}及训练时间

Tab.4 The R_{Acc} and training time of the five algorithms after adding noise

数据集	SOCP-TSVM		R-TWSVM		NPSVM		V-PTSVM		SOCP-PTSVM	
	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s	R_{Acc}/%	时间/s
Ionosphere	86.62	0.097	84.15	0.082	86.9	0.147	88.13	0.028	**92.47**	0.029
WPBC	76.47	0.101	76.85	0.078	76.66	0.09	76.77	0.016	**85.53**	0.039
Toes	59.27	0.217	68.21	0.091	58.33	0.225	67.65	0.009	**73.62**	0.174
Australian	85.73	0.048	74.51	0.132	87.20	0.871	59.57	4.227	**88.38**	0.040
Echocardiogra	86.94	0.080	87.70	0.049	83.94	0.086	89.73	0.024	**90.12**	0.050
Vehicle1	79.69	0.150	74.16	0.258	77.88	0.492	78.61	6.189	**80.25**	0.055
Spilce	80.88	0.145	64.90	0.994	81.19	0.834	80.90	0.547	**81.30**	0.041
Glass1	66.09	0.058	64.90	0.161	67.86	0.104	70.80	0.028	68.23	0.038
Chess	91.30	0.135	59.33	3.032	94.43	9.008	66.80	74.357	**95.12**	0.127

表5 加噪后SOCP-PTSVM与其他4个算法的配对t检验p值比较

Tab.5 Comparison of p value of paired t test on SOCP-PTSVM and another four algorithms after adding noise

数据集	配对t检验p值			
	SOCP-PTSVM与SOCP-TSVM	SOCP-PTSVM与R-TWSVM	SOCP-PTSVM与NPSVM	SOCP-PTSVM与V-PTSVM
Ionosphere	0.782 495	0.017 621	0.154 227	0.022 912
WPBC	0.076 263	9.65×10^{-9}	0.003 722	0.418 817
Toes	0.698 863	0.857 045	0.190 902	0.873 596
Australian	0.003 067	0.001 102	0.028 989	0.006 128
Echocardiogra	0.239 843	0.239 843	0.239 843	0.267 51
Vehicle1	0.309 237	1	0.992 711	0.622 124
Spilce	0.209 651	0.000 175	0.263 806	2.98×10^{-7}
Glass1	0.049 21	0.016 71	0.029 057	0.022 763
Chess	6.49×10^{-5}	0.790 503	4.74×10^{-5}	5.10×10^{-5}

4 结论

为了降低分类器对异常点或噪声的敏感性,提升模型的鲁棒性,本文通过引入机会约束,以概率的形式允许一定概率的样本在投影空间中远离自身类别的投影中心,从而提出了一个新的鲁棒投影支持向量机模型。并将该模型等价地转化成两个带有二阶锥约束的线性二阶锥规划问题,这两个问题可以利用求解器SeDuMi进行快速求解。数值实验结果表明,本文所提出的方法具有很好的鲁棒性和泛化性。

参考文献:

[1] DENG N Y,TIAN Y J,ZHANG C H. Support vector machines:optimization based theory,algorithms,and extensions[M]. New York: Chapman and Hall/CRC,2012.

[2] TSANG I W,KWOK J T,CHEUNG P M. Core vector machines: fast SVM training on very large data sets[J]. Journal of

Machine LearningResearch，2005，6(2)：363-392.

[3] SCHOLKOPF B，SMOLA A J，WILLIAMSON R C，et al. New support vector algorithms[J]. Neural Computation，2000，12(5)：1207-1245.

[4] CORTES C，VAPNIK V. Support-vector networks[J]. Machine Learning，1995，20(3)：273-297.

[5] CHEN Z S，QI Z Q，WANG B，et al. Learning with label proportions based on nonparallel support vector machines[J]. Knowledge-Based Systems，2017，119：126-141.

[6] WANG X F，HUANG D S，XU H. An efficient local chan-vese model for image segmentation[J]. Pattern Recognition，2010，43(3)：603-618.

[7] LI B，ZHENG C H，HUANG D S. Locally linear discriminant embedding：an efficient method for facerecognition[J]. Pattern Recognition，2008，41(12)：3813-3821.

[8] JIANG W，HUANG D S，LI S. Random walk-based solution to triple level stochastic point location problem[J]. IEEE Transactions on Cybernetics，2017，46(6)：1438-1451.

[9] MANGASARIAN O L，WILD E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2006，28(1)：69-74.

[10] 杨琳希，李国权. 基于 DC 规划方法的稀疏临近支持向量机[J]. 运筹学学报，2020，24(3)：115-126.
YANG L X，LI G Q. Sparse proximal support vector machines via DC programming[J]. Operations Research Transactions，2020，24(3)：115-126.

[11] JAYADEVA，KHEMCHANDANI R，CHANDRA S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2007，29(5)：905-910.

[12] 廖均淋，白富生，马龙. 基于 L_1-范数的 v-双子限定支持向量机[J]. 重庆师范大学学报(自然科学版)，2020，37(2)：1-11.
LIAO J L，BAI F S，MA L. v-twin bounded support vector machines based on L_1-norm[J]. Journal of Chongqing Normal University (Natural Science)，2020，37(2)：1-11.

[13] SHAO Y H，ZHANG C H，WANG X B，et al. Improvements on twin support vector machines[J]. IEEE Transactions on Neural Networks，2011，22(6)：962-968.

[14] MALDONADO S，LÓPEZ J，CARRASCO M. A second-order cone programming formulation for twin support vector machines[J]. Applied Intelligence，2016，45(2)：265-276.

[15] CHEN X B，YANG J，YE Q L，et al. Recursive projection twin support vector machine via within-class variance minimization[J]. Pattern Recognition，2011，44(10/11)：2643-2655.

[16] SHAO Y H，WANG Z，CHEN W J，et al. A regularization for the projection twin support vector machine[J]. Knowledge-Based Systems，2013，37(2)：203-210.

[17] CHEN W J，SHAO Y H，LI C N，et al. v-projection twin support vector machine for pattern classification[J]. Neurocomputing，2020，376：10-24.

[18] LANCKRIET G R G，El GHAOUI L，BHATTACHARYYA C，et al. A robust minimax approach to classification[J]. Journal of Machine Learning Research，2003，3：555-582.

[19] ALIZADEH F，GOLDFARB D. Second-order cone programming[J]. Mathematical Programming，2001，95：3-51.

[20] SOKOLOVA M，JAPKOWICZ N，SZPAKOWICZ S. Beyond accuracy，F-score and ROC：a family of discriminant measures for performance evaluation[C]//SATTAR A，KANG B H. AI2006：advances in artificial intelligence：19th Australian joint conference on artificial intelligence，Hobart，Australia，December 4-8，2006. proceedings. Berlin：Springer-Verlag，2006：1015-1021.

[21] 方述诚，邢文训. 线性锥优化[M]. 北京：科学出版社，2013.
FANG S C，XING W X. Linear cone optimization[M]. Beijing：Science Press，2013.

[22] QI Z Q，TIAN Y J，SHI Y. Robust twin support vector machine for pattern classification[J]. Pattern Recognition，2013，46(1)：305-316.

[23] LÓPEZ J，MALDONADO S，CARRASCO M. Robust nonparallel support vector machines via second-order cone programming[J]. Neurocomputing，2019，364：227-238.

[24] DUA D，TANISKIDOU E K. UCI machine learning repository[DS/OL]. [2020-11-05]. http://archive.ics.uci.edu/ml/.

Operations Research and Cybernetics

Robust Projection Support Vector Machine Based on Second-Order Cone Programming

WU Zhiyou, YIN Lin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: [**Purposes**]Projection support vector machine is a new non-parallel classification model proposed by applying the idea of linear discriminate analysis to twin support vector machine. which aims to find two non-parallel projection directions instead of hyperplanes. However, the model is not robust enough. When there are a large number of abnormal points or noise in the training data set, the projection direction learned by the projection support vector machine will often be affected by it and shift, and the classification performance will decrease. In order to reduce the sensitivity of the model to abnormal points or noise, the robustness of the model is improved. [**Methods**]The chance constraint is introduced into the model. In the projection space, the projection distance from a part of the projection sample to its sample center is allowed to be greater than the projection distance from it to another kind of sample center. In other words, an upper bound of the probability of wrong sample separation is given. [**Findings**]Obtain a new robust projection support vector machine with chance constraints, and equivalently transform it into a second-order cone programming problem, so that only a pair of linear second-order cone programming problems can be solved to train two non-parallel projection directions. [**Conclusions**]In numerical experiments on UCI dataset and the dataset with increased noise, compared with other algorithms, the proposed robust projection support vector machine based on second-order cone programming has little change in accuracy, is relatively stable, and has better robustness and generalization ability.

Keywords: projection support vector machine; chance constraint; second-order cone programming; robustness

(责任编辑 方 兴)