优化样本分布的最接近支持向量机

杨 勃[1,2]
(1.湖南理工学院信息与通信工程学院,湖南岳阳 414006;
2.湖南理工学院复杂系统优化与控制湖南省普通高等学校重点实验室,湖南岳阳 414006)

摘 要: 当两类样本分布存在差异时,最接近支持向量机(Proximal Support Vector Machine,PSVM)等最小二乘类分类器分类结果将出现偏差,不能实现最小错误率分类.本文在分析 PSVM 等价广义特征值分解模型基础上,提出了一种改善原 PSVM 分类决策面的优化样本分布 PSVM,其基本思想是通过引入最大化正确分类样本距决策面距离,同时最小化错误分类样本距决策面距离的优化样本分布正则化项,构造优化样本分布 PSVM 的广义特征值分解模型.通过人工数据集和 UCI 数据集的 10 个数据子集上的对比实验,验证了该改进分类模型能够有效调整决策边界,从而获得更好的分类效果.

关键词: 最接近支持向量机;优化样本分布;正则化技术
中图分类号: TN911.23 **文献标识码:** A **文章编号:** 0372-2112 (2014)12-2429-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2014.12.014

Proximal Support Vector Machine Based on Optimizing Sample Distribution

YANG Bo[1,2]
(1. *School of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang, Hunan* 414006, *China*;
2. *Key Laboratory of Optimization and Control for Complex Systems, College of Hunan Province, Yueyang, Hunan* 414006, *China*)

Abstract: When the distributions of 2 class samples are different, the classification results will be biased by using least square classifiers, such as proximal support vector machine (PSVM). Inevitably, this decision bias will cause non-minimal classification error rates. In the present paper, based on equivalent generalized eigenvalue decomposition model of PSVM, a novel optimizing samples distribution PSVM model is proposed, which can improve original PSVM decision. The model is constructed as a generalized eigenvalue decomposition model and contains an optimal samples distribution regularization item. It can maximize distances between correctly classified samples and decision boundary and minimize distances between misclassified samples and decision boundary. Experimental results under artificial datasets and 10 data subsets from UCI datasets show that using this novel model can adjust decision effectively and achieve better classification effects.

Key words: proximal support vector machine(PSVM); optimizing sample distribution; regularization technique

1 引言

最接近支持向量机 PSVM[1]是支持向量机 SVM[2~4]的一类变体.与最小二乘支持向量机(Least Square Support Vector Machine, LSSVM)[5]类似,PSVM 采用等式约束,具有求解快速的优点.此外,与 LSSVM 相比,PSVM 将阈值平方纳入到优化目标中,使优化问题具有更为简洁的表达形式,有效增强了优化模型的凸性.由于 PSVM 具有模型简单,求解快速等优点,得到了相关学者的重视,在原 PSVM 模型基础上,又提出了多类 PSVM[6,7],模糊 PSVM[8,9]等扩展模型和其他改进模型[10,11],并在轴承、齿轮箱故障诊断、人脸识别等领域[11~13]取得了一定的应用成果.

模式识别理论指出,尽管在最小均方误差意义上 PSVM 等最小二乘分类器最优逼近 Bayes 分类器,但并不意味一定能使误分概率极小化[14].究其原因,主要是因为 PSVM 的最优分类性能建立在两类样本分布特性相同的假设基础上.在实际应用中,这种假设通常并不成立,甚至在某些应用场合,还会出现不同类别样本分布差异较大的极端情形.此时,受大方差类别样本影响,

收稿日期:2013-06-05;修回日期:2014-04-23;责任编辑:梅志强
基金项目:国家 973 预研项目(No.2014CB046300);湖南省科技计划(No.2012WK4015,No.2013GK3099,No.2014GK3026);湖南省教育厅科学研究优秀青年项目(No.11B055);湖南省创新团队支持计划(No.湘教通〔2012〕318 号)

采用 PSVM 学习会出现较大偏差,分类效果退化严重.因此,要提高实际应用中 PSVM 方法的分类性能,则必须要消除不同类别样本分布差异对 PSVM 学习的不利影响.

为此,本文在 PSVM 具有最优逼近 Bayes 分类意义的主优化项基础上,尝试将优化样本分布作为一个优化子项引入,提出了一种优化分布 PSVM.该优化分布 PSVM 通过构造等价广义特征值分解模型,优化正确、错误分类样本分布,有效缓解分布差异带来的决策偏差.最后,通过人工数据集、UCI 数据集上的对比实验,验证了本优化分布 PSVM 模型的有效性.

2 PSVM 及其等价模型

对二分类问题,设有 l 维训练样本集$\{(\boldsymbol{x}_i, y_i) \mid \boldsymbol{x}_i \in \mathbf{R}^{l\times 1}, y_i \in \{\pm 1\}\}_{1\leqslant i\leqslant N}$,其对应的样本矩阵为 $\boldsymbol{X}(\boldsymbol{X}\in \mathbf{R}^{l\times N})$,PSVM 模型优化准则可写为:

$$\min \frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + b^2) + \frac{c}{2}\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\zeta} \tag{1}$$

$$\text{s.t.} \quad \boldsymbol{D}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} + \boldsymbol{e}b) + \boldsymbol{\zeta} = \boldsymbol{e}$$

其中,$\boldsymbol{e}$ 为全 1 的 N 维列向量,矩阵 $\boldsymbol{D}$ 为 $N\times N$ 维对角阵,且对角元素 $\boldsymbol{D}_{ii}$ 为第 i 个样本的类别值 y_i.

首先对 PSVM 式(1)进行变形简化,可得定理 1.

定理 1 令 $\tilde{\boldsymbol{x}}_i = (\boldsymbol{x}_i^{\mathrm{T}} \quad 1)^{\mathrm{T}}$, $\tilde{\boldsymbol{w}} = (\boldsymbol{w}^{\mathrm{T}} \quad b)^{\mathrm{T}}$, $c_1 = 1/c$, $\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}\in \mathbf{R}^{(l+1)\times N})$为新样本$\{\tilde{\boldsymbol{x}}_i\}$构成的样本矩阵,此时 PSVM 模型等价于正则化最小二乘模型 $\min \frac{c_1}{2}\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \frac{1}{2}\|\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} - \boldsymbol{Y}\|^2$,其最优解为:$\tilde{\boldsymbol{w}}^* = (\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}} + c_1\boldsymbol{I})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{Y}$.

证明 将 $\tilde{\boldsymbol{w}} = (\boldsymbol{w}^{\mathrm{T}} \quad b)^{\mathrm{T}}$, $c_1 = 1/c$, $\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}\in \mathbf{R}^{(l+1)\times N})$代入式(1),整理可得如下等价模型:

$$\min \quad \frac{c_1}{2}\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \frac{1}{2}\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\zeta} \tag{2}$$

$$\text{s.t.} \quad \boldsymbol{D}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \boldsymbol{\zeta} = \boldsymbol{e}$$

$\boldsymbol{D}$ 满秩,因此对任意向量 $\boldsymbol{z}$,有 $\boldsymbol{D}\boldsymbol{z} = 0$ 等价于 $\boldsymbol{z} = 0$.因此,式(2)等式约束可改为 $\boldsymbol{D}\boldsymbol{D}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \boldsymbol{D}\boldsymbol{\zeta} = \boldsymbol{D}\boldsymbol{e}$.

又因 $\boldsymbol{D}\boldsymbol{D} = \boldsymbol{I}$, $\boldsymbol{D}\boldsymbol{e} = \boldsymbol{Y}$,令 $\boldsymbol{D}\boldsymbol{\zeta} = \boldsymbol{\eta}$,代入式(2)得式(3):

$$\min \quad \frac{c_1}{2}\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \frac{1}{2}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta} \tag{3}$$

$$\text{s.t.} \quad \tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \boldsymbol{\eta} = \boldsymbol{Y}$$

将式(3)中等式约束代入主优化式得式(4):

$$\min \quad \frac{c_1}{2}\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + \frac{1}{2}\|\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} - \boldsymbol{Y}\|^2 \tag{4}$$

显然,式(4)就是一个正则化最小二乘优化模型,正则化项为投影矢量 2 范数 $\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}}$,其最优解为 $\tilde{\boldsymbol{w}}^* = (\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}} + c_1\boldsymbol{I})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{Y}$.

证毕.

下面进一步分析式(4).首先分析无正则化项基本最小二乘分类模型 $\min\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{Y}\|^2$,其最优解为 $\boldsymbol{w}_{\mathrm{opt}} = (\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{Y}$.分析该模型可得定理 2.

定理 2 最小二乘分类模型 $\min\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{Y}\|^2$ 与优化式(5)等价.

$$\max \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}} \tag{5}$$

证明 式(5)对应于如下拉格朗日优化问题:

$$\max \boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} - \lambda\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}$$

其导数 0 点满足:

$$\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} = \lambda\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}$$

对任意列向量 $\boldsymbol{w}$,令 $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} = a$,则有:

$$a\boldsymbol{X}\boldsymbol{Y} = \lambda\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}$$

上式求得的最优解 $\boldsymbol{w}^* = \frac{a}{\lambda}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{Y}$,满足 $\boldsymbol{w}^* \propto \boldsymbol{w}_{\mathrm{opt}}(\boldsymbol{w}_{\mathrm{opt}} = (\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{Y})$,表明式(5)最优解与最小二乘最优解一致.

进一步限制 $a > 0$(若 $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}^* < 0$,则取 $\boldsymbol{w}^* = -\boldsymbol{w}^*$),因 $\lambda \geqslant 0$,此时 $\boldsymbol{w}^*$ 与最小二乘解同向.

对分类问题,决策的依据是 $\mathrm{sgn}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})$,与 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$ 值大小无关.因此,最小二乘分类模型与优化模型(5)等价.

证毕.

令 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{Z}$,还可将式(5)中的优化式改写为:

$$\frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}} \Leftrightarrow \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}} = \left(\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Y}}{\sqrt{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}\sqrt{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}}}\right) \tag{6}$$

式(6)描述了最小二乘分类的等价数学意义:对分类问题,最小二乘等价于向量 $\boldsymbol{Z}$、$\boldsymbol{Y}$ 的最优对齐.

依据定理 2,我们可得到 PSVM 等价正则化最小二乘分类器模型的广义特征值分解模型,见推论 1.

推论 1 PSVM 分类模型与优化式(7)等价.

$$\max \frac{\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}}}{\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{w}} + c_1\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{w}}} \tag{7}$$

3 模型优化样本分布 PSVM

在训练样本集上用 PSVM 学习,可得到分类器 $\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}} = 0$.若 $\mathrm{sgn}(\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}}_i) = y_i$,则样本 $\tilde{\boldsymbol{x}}_i$ 被正确分类,否则被误分.依分类结果,可将样本集划分为正分样本集$\{\tilde{\boldsymbol{x}}^i_{\text{正分}}\}$和误分样本集$\{\tilde{\boldsymbol{x}}^i_{\text{误分}}\}$.分类示意图如图 1.

图 1 中实心样本为误分样本,空心样本为正分样本.误分样本以值 $d_{\text{误分}} = \tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}}_{\text{误分}}$分布在分类器 $\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}} = 0$ 两边,正分样本以值 $d_{\text{正分}} = \tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}}_{\text{正分}}$分布在分类器 $\tilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\boldsymbol{x}} = 0$ 两边.

为改善分类效果,需优化误、正分样本分布,即压缩决策边界两边误分样本所占区域大小,以减少潜在

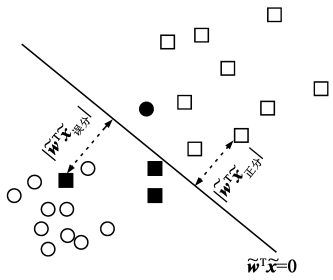


图1　分类器$\widetilde{w}^{\mathrm{T}}\widetilde{x}=0$下的分类示意图

错误率;同时使正分样本远离决策边界,获得大间隔,提高泛化性.

假设随机变量 $d_{误分}$,$d_{正分}$均满足 0 均值,从统计学角度,要求二阶统计量 $d_{误分}$的方差应尽可能小,同时要求 $d_{正分}$的方差应尽可能大,可描述为如下优化样本分布模型:

$$\max \frac{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}}{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{误分}\widetilde{\boldsymbol{X}}_{误分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}} \tag{8}$$

其中 $\widetilde{\boldsymbol{X}}_{正分}$,$\widetilde{\boldsymbol{X}}_{误分}$分别为正、误分样本矩阵.进一步分析式(8),可得如下等价模型.

定理 3　优化式(8)与式(9)等价.

$$\max \frac{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}}{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}} \tag{9}$$

其中,$\widetilde{\boldsymbol{X}}$ 为所有样本组成的样本矩阵,$\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}=\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}+\widetilde{\boldsymbol{X}}_{误分}\widetilde{\boldsymbol{X}}_{误分}^{\mathrm{T}}$.

证明　式(8)是一个广义特征值分解模型,其最优条件为:

$$\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}=\lambda\widetilde{\boldsymbol{X}}_{误分}\widetilde{\boldsymbol{X}}_{误分}^{\mathrm{T}}\widetilde{\boldsymbol{w}} \tag{10}$$

式(10)两边各加 $\lambda\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}$,经整理可写为:

$$\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}=\lambda'\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{w}} \tag{11}$$

其中 $\lambda'=\dfrac{\lambda}{(1+\lambda)}$.式(11)是式(9)的最优条件,这表明优化式(8)与式(9)等价.

证毕.

考虑到实际应用中有可能出现 $\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}$不满秩的情况,式(9)也可引入正则化项 $c_1\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}$:

$$\max \frac{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}}{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}+c_1\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}} \tag{12}$$

最终,综合考虑式(7)和式(12),提出如下优化样本分布 PSVM 模型:

$$\max \frac{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}+c_2\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}}{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}+c_1\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}} \tag{13}$$

其中,$c_2(c_2>0)$为优化样本分布项的正则化系数.

在已知当前分类器下正分样本矩阵 $\widetilde{\boldsymbol{X}}_{正分}$的前提下,优化样本分布 PSVM 式(13)是一个典型的广义特征值分解问题,可取最优投影矢量 $\widetilde{\boldsymbol{w}}^{*}$为最大特征值所对应的特征矢量.我们通过式(13)优化样本分布,改善分类器分类效果.具体算法设计如算法 1.

算法 1

(1)预置 c_1,c_2,求得 PSVM 最优解 $\widetilde{\boldsymbol{w}}_0=(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}+c_1\boldsymbol{I})^{-1}\widetilde{\boldsymbol{X}}\boldsymbol{Y}$;

(2)由最优解 $\widetilde{\boldsymbol{w}}_0$ 计算得到正分样本矩阵 $\widetilde{\boldsymbol{X}}_{正分}$;

(3)采用广义特征值分解计算优化样本分布 PSVM 模型式(13),$\widetilde{\boldsymbol{w}}_1$ 取最大特征值所对应特征矢量.若 $\widetilde{\boldsymbol{w}}_1^{\mathrm{T}}\widetilde{\boldsymbol{X}}\boldsymbol{Y}<0$,令 $\widetilde{\boldsymbol{w}}_1$ 取负,即 $\widetilde{\boldsymbol{w}}_1=-\widetilde{\boldsymbol{w}}_1$,否则 $\widetilde{\boldsymbol{w}}_1$ 不变;

(4)输出学习结果 $\widetilde{\boldsymbol{w}}_1$.

算法 1 表明,求解优化样本分布 PSVM 需在求得原始 PSVM 最优解基础上,再进一步微调优化.因此,优化样本分布 PSVM 相对于 PSVM,计算复杂度有所增加.下面分析算法 1 的时间复杂度.

算法 1 由三步实现.设样本矩阵 $\widetilde{\boldsymbol{X}}\in\mathbf{R}^{M\times N}$,维数 $M=l+1$,N 为样本数.第一步求 PSVM 最优解,其时间代价主要耗费在 $\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}$计算、$(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}+c_1\boldsymbol{I})^{-1}$求逆以及$(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{T}}+c_1\boldsymbol{I})^{-1}\widetilde{\boldsymbol{X}}$ 计算上,矩阵乘法计算时间复杂度均为 $o(M^2N)$,矩阵求逆为 $o(M^3)$;第二步求矩阵 $\widetilde{\boldsymbol{X}}_{正分}$,其时间代价主要耗费在计算一维投影 $\widetilde{\boldsymbol{w}}_0^{\mathrm{T}}\widetilde{\boldsymbol{X}}$ 上,时间复杂度为$o(M\times N)$;第三步求优化样本分布 PSVM 最优解,其时间代价主要耗费在 $\widetilde{\boldsymbol{X}}_{正分}\widetilde{\boldsymbol{X}}_{正分}^{\mathrm{T}}$计算以及模型式(13)的广义特征值分解计算上,设正分样本数为 N_+($N_+<N$),前者时间复杂度为 $o(M^2N_+)$,后者为 $o(M^3)$.

以上时间复杂度分析表明,为改善 PSVM 分类效果,优化样本分布 PSVM 的时间复杂度有所增加.但从指数大小层面看,优化样本分布 PSVM 仍属时间复杂度为 $o(M^3)$的算法,与原始 PSVM 相同.

4　优化样本分布 Kernel PSVM

优化样本分布 Kernel PSVM(KPSVM)是优化样本分布 PSVM 的非线性版本.通过核映射 ϕ,原始样本 $\boldsymbol{x}_i$ 被映射为$\phi(\boldsymbol{x}_i)$.令核样本 $\tilde{\phi}(\boldsymbol{x}_i)=(\phi(\boldsymbol{x}_i)^{\mathrm{T}}\quad 1)^{\mathrm{T}}$,投影矢量 $\widetilde{\boldsymbol{w}}=(\widetilde{\boldsymbol{w}}^{\mathrm{T}}\quad b)^{\mathrm{T}}$,此时优化样本分布 KPSVM 对应的优化准则为:

$$\max \frac{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{w}}+c_2\widetilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})_{正分}\tilde{\phi}(\boldsymbol{X})_{正分}^{\mathrm{T}}\widetilde{\boldsymbol{w}}}{\widetilde{\boldsymbol{w}}^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})\tilde{\phi}(\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{w}}+c_1\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}} \tag{14}$$

其中,$\tilde{\phi}(\boldsymbol{X})$为核空间中的样本矩阵,$\tilde{\phi}(\boldsymbol{X})_{正分}$为核空间中被分类器正确分类的样本所组成的样本矩阵.

进一步,我们采用式(15)将无穷维隐式核空间中的优化问题转换到有限维核内积空间中求解[15]:

$$\widetilde{\boldsymbol{w}}=\tilde{\boldsymbol{\phi}}(\boldsymbol{X})\alpha \tag{15}$$

将式(15)代入到优化式(14)中,同时类似 PSVM 将分母项中的 $c_1\widetilde{\boldsymbol{w}}^{\mathrm{T}}\widetilde{\boldsymbol{w}}$ 直接置换为$c_1\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}$,整理可得:

$$\max \frac{\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{K}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\widetilde{\boldsymbol{K}}\boldsymbol{\alpha} + c_2\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{K}}_{\text{正分}}\widetilde{\boldsymbol{K}}_{\text{正分}}^{\mathrm{T}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{K}}\boldsymbol{\alpha} + c_1\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}} \tag{16}$$

其中，$\widetilde{\boldsymbol{K}} = \tilde{\boldsymbol{\phi}}(\boldsymbol{X})^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})$，$\widetilde{\boldsymbol{K}}_{\text{正分}} = \tilde{\boldsymbol{\phi}}(\boldsymbol{X})^{\mathrm{T}}\tilde{\phi}(\boldsymbol{X})_{\text{正分}}$.

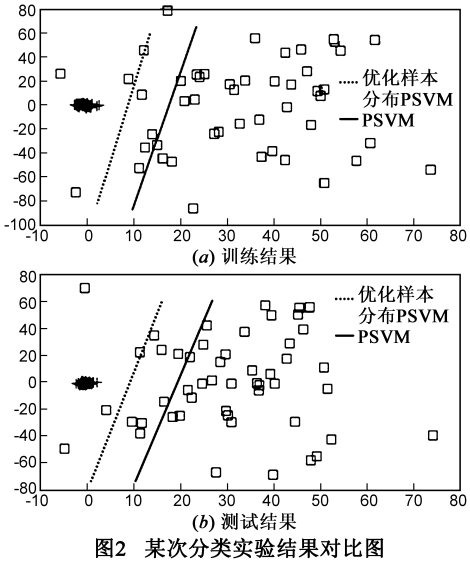
需要指出的是，原始 KPSVM 采用的转换式与式(15)不同，为 $\boldsymbol{w} = \phi(\boldsymbol{X})\boldsymbol{D}\boldsymbol{\alpha}$[1].

总之，通过核内积映射技巧，无穷维隐式核空间中的优化问题被转换到维数为训练样本数 N 的核内积空间进行计算. 此时采用算法 1 步骤，求解优化样本分布 KPSVM，其时间复杂度变为 $o(N^3)$，而基于二次规划的 SVM 时间复杂度要更高一些，为 $o(N^{3.5})$[16].

5 实验结果与分析

5.1 人工数据集实验

我们构造人工数据集以测试优化样本分布 PSVM 分类效果. 首先，我们随机生成两类样本各 100 个(前 50 个用于训练，后 50 个用于测试)，一类样本服从均值列向量为$(-0.5 \quad 0)^{\mathrm{T}}$，协方差矩阵为$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$的高斯分布；另一类样本服从均值列向量为$(30 \quad 0)^{\mathrm{T}}$，协方差矩阵为$\begin{pmatrix}20 & 0\\ 0 & 40\end{pmatrix}$的高斯分布. 分别采用线性 PSVM 和优化样本分布线性 PSVM 对上述样本进行分类训练和测试，其中线性 PSVM 的正则化系数 $c1$ 取 $c1 = 10^{-7}$，优化样本分布线性 PSVM 的正则化系数 c 取 $c = 1000$，$c1$ 取 $c1 = 10^{-7}$. 实验重复 50 次，某次分类训练和测试结果如图 2 所示.



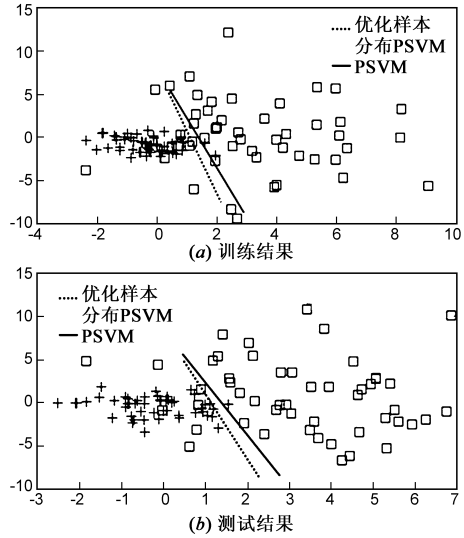
(a) 训练结果　(b) 测试结果

图2 某次分类实验结果对比图

从图 2 可以看出，由于两类样本分布存在显著差异，此时，PSVM 分类边界受大方差类别样本干扰，出现明显偏差；而优化样本分布 PSVM 分类边界相对更优. 优化样本分布 PSVM 和 PSVM50 次分类测试的平均正确率分别为 94.9%，89.3%，优化样本分布 PSVM 明显优于 PSVM.

进一步，我们减小大方差类别样本的分布尺度，改为服从均值列向量为$(3 \quad 0)^{\mathrm{T}}$，协方差矩阵为$\begin{pmatrix}2 & 0\\ 0 & 4\end{pmatrix}$的高斯分布；实验重复 50 次，某次分类结果如图 3 所示.

从图 3 可以看出，由于两类样本分布差异相对而言不如上一组实验显著，此时，PSVM 分类边界无明显偏差. 优化样本分布 PSVM 和 PSVM50 次分类测试的平均正确率分别为 90.5%，90.3%，优化样本分布 PSVM 分类效果无明显改善.



(a) 训练结果　(b) 测试结果

图3 某次分类实验结果对比图

5.2 UCI 数据集实验

进一步，我们采用 UCI 数据集[17]来测试算法的分类效果. 使用 6 个 2 分类 UCI 数据子集：Bupa(6 维/345 个样本)、Diabetes(8 维/768 个样本)、Monk1(6 维/432 个样本)、Monk2(6 维/432 个样本)、Monk3(6 维/432 个样本)、Sonar(60 维/208 个样本)和 4 个多分类数据子集：Glass(9 维/6 类别/214 个样本)、cmc(9 维/3 类别/1473 个样本)、tae(5 维/3 类别/151 个样本)、vehicle(18 维/4 类别/846 个样本)，分别测试线性、高斯核 SVM，线性、高斯核 LSSVM，线性、高斯核 PSVM，优化样本分布线性、高斯核 PSVM. 其中，SVM 采用 Steve Gunn 开源 SVM MATLAB 工具箱，其余三种分类器自编.

分类器可调参数取值范围设定为：高斯核参数 $\sigma = 2^{-7} \sim 2^{10}$，LSSVM、PSVM 参数 $c = 2^{-12} \sim 2^{12}$，优化样本分布 PSVM 参数 $c_1 = 2^{-10}$，$c_2 = 2^{-4} \sim 2^{15}$. 对后四个多类别数据子集使用 one-versus-rest 方式实现多分类. 实验采用 10 重交叉验证方式，结果见表 1.

在实验中，优化样本分布 PSVM 与其他三种分类器相比，尤其在高斯核非线性分类情形下，取得了相对更

好的分类结果.这说明优化样本分布 PSVM 通过压缩误分样本所占区域,同时使正分样本远离分类边界,能够改善决策偏差,从而进一步提高了分类正确率.

进一步,我们对各分类器的训练耗时进行了实验比较.分类器运行硬件平台为 CPU I7 640M 2.8GHz,内存 4GB;软件平台为 WINDOWS 7,MATLAB R2008.以下是 10 次训练平均耗时,见表 2.

从表 2 可以看出,在线性、高斯核非线性分类两种情形下,SVM 平均耗时最大.特别在线性分类情形下,由于 SVM 实际工作在与训练样本数相等的 N 维线性内积空间,远大于原始样本维数,此时 SVM 训练平均耗时显著大于其他三类分类器.而优化样本分布 PSVM 相比于 LSSVM 和 PSVM,其平均耗时有所增加,但远小于 SVM 平均耗时.

表 1　SVM、LSSVM、PSVM、优化样本分布 PSVM 平均测试正确率

	SVM		LSSVM		PSVM		优化样本分布 PSVM	
	线性	高斯核	线性	高斯核	线性	高斯核	线性	高斯核
Bupa	0.6759	0.7280	0.6714	0.7188	0.6698	0.7124	0.6773	0.7385
Diabetes	0.7565	0.7692	0.7582	0.7787	0.7615	0.7765	0.7534	0.7724
Monk1	0.6656	0.8674	0.6645	0.8610	0.6627	0.8657	0.6643	0.8850
Monk2	0.6387	0.7461	0.6199	0.7381	0.6212	0.7335	0.6450	0.7424
Monk3	0.7734	0.9705	0.7904	0.9742	0.7898	0.9722	0.7810	0.9785
Sonar	0.7023	0.8346	0.6823	0.8255	0.6887	0.8297	0.6952	0.8513
Glass	0.5833	0.6657	0.5914	0.6592	0.5897	0.6588	0.6128	0.6815
cmc	0.5165	0.6074	0.5161	0.5843	0.5172	0.5854	0.5170	0.6149
tae	0.5057	0.5314	0.5088	0.5298	0.5068	0.5285	0.5074	0.5501
vehicle	0.7258	0.7585	0.7269	0.7436	0.7256	0.7447	0.7269	0.7565

表 2　SVM、LSSVM、PSVM、优化样本分布 PSVM 平均耗时(s)

	SVM		LSSVM		PSVM		优化样本分布 PSVM	
	线性	高斯核	线性	高斯核	线性	高斯核	线性	高斯核
Bupa	13.758	13.234	0.0001	0.0184	0.0001	0.0162	0.0003	0.2536
Diabetes	112.65	107.72	0.0002	0.2369	0.0002	0.2042	0.0005	3.8605
Monk1	20.123	21.502	0.0001	0.0372	0.0001	0.0258	0.0003	0.6267
Monk2	23.682	21.998	0.0001	0.0368	0.0001	0.0293	0.0003	0.6153
Monk3	20.447	20.713	0.0001	0.0391	0.0001	0.0314	0.0003	0.4512
Sonar	2.3654	2.4306	0.0005	0.0043	0.0005	0.0043	0.0029	0.0740
Glass	13.461	14.367	0.0008	0.0271	0.0008	0.0284	0.0017	0.3585
cmc	3287.2	3724.7	0.0006	2.8862	0.0006	3.1217	0.0018	74.255
tae	3.7248	3.4291	0.0002	0.0106	0.0002	0.0097	0.0007	0.0893
vehicle	1024.4	1078.6	0.0013	0.7956	0.0012	0.8245	0.0031	24.415

6　总结

本文在最小二乘类分类器 PSVM 基础上,通过引入优化样本分布正则化项对原始 PSVM 决策面进行调整,提出了一种改进 PSVM 分类模型.人工数据集和部分 UCI 数据子集上的对比实验表明,该改进模型能够有效改善正分样本和误分样本分布,进一步提高了分类正确率,证明了将优化样本分布引入 PSVM 的有效性.进一步,考虑到实际应用中的小样本问题,此时样本分布统计量估计受野值干扰较大.为此,我们下一步将研究提高样本分布统计分析鲁棒性方法,以提高实际应用中优化样本分布估计的稳定性和有效性.

参考文献

[1] Fung G, Mangasarian O L. Proximal support vector machine classifier[A]. Proc of the 7th ACM SIGKDD International Coference on Knowledeg Discovery and Data Mining[C]. San Francisco: ACM, 2001. 77 – 82.

[2] Vapnik V N. The Nature of Statistical Learning Theory[M]. Second Edition. New York: Springer Press, 2000.

[3] 王国胜,钟义信.支持向量机的若干新进展[J].电子学报,2001,29(10):1397 – 1400.
Wang Guosheng, Zhong yixin. Somenew developments on support vector machine[J]. Acta Electronica Sinica, 2001, 29(10): 1397 – 1400. (in Chinese)

[4] 方景龙,陈铄,潘志庚,梁荣华.复杂分类问题支持向量机的简化[J].电子学报,2007,35(5):858 – 861.
FANG Jinglong, CHEN Shuo, PAN Zhigeng, LIANG Ronghua. Simplification to support vector machine for complicated recognition problem[J]. Acta Electronica Sinica, 2007, 35(5): 858 – 861. (in Chinese)

[5] Suykens J A K, Vandewalle J. Least square support vector machine classifier[J]. Neural Processing Letters, 1999, 9(3): 293 – 300.

[6] 杨绪兵,陈松灿.基于原型超平面的多类最接近支持向量机[J].计算机研究与发展,2006,43(10):1700 – 1705.
Yang Xubing, Chen Songcan. Proximal support vector machine based on prototypal multiclassification hyperplanes[J]. Journal of Computer Research and Development, 2006, 43(10): 1700 – 1705. (in Chinese)

[7] Niu Lingfeng. Parallel algorithm for training multiclass proximal Support Vector Machines[J]. Applied Mathematics and Computation, 2011, 217(12): 5328 – 5337.

[8] Jayadeva K R, Chandra S. Fuzzy linear proximal support vector machines for multi-category data classification[J]. Neurocomputing, 2005, 67(8): 426 – 435.

[9] Yang Xubing, Chen Songcan, Chen Bin, et al. Proximal support vector machine using local information[J]. Neurocomputing, 2009, 73(1 – 3): 357 – 365.

[10] Reshma K, Jayadeva, Suresh C. Knowledge based proximal support vector machines[J]. European Journal of Operational Research, 2009, 195(3): 914 – 923.

[11] Reshma K, Anuj K, Suresh C. Generalized eigenvalue proximal support vector regressor[J]. Expert Systems with Applications, 2011, 38(10): 13136 – 13142.

[12] Sugumaran V, Muralidharan V, Ramachandran K I. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing[J]. Mechanical Systems and Signal Processing, 2007, 21(2): 930 – 942.

[13] Saravanan N, Kumar V N S, Ramachandran K I. Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM)[J]. Applied Soft Computing, 2010, 10(1): 344 – 360.

[14] Duda R O, Hart P E, Stork D G. Pattern Classification[M]. Second Edition. New York: Wiley-Interscience, 2000. 239 – 245.

[15] Schölkopf B, Mika S, et al. Input space versus feature space in kernel based methods[J]. IEEE Transactions on Neural Networks, 1999, 10(5): 1000 – 1017.

[16] Mangasarian O L, Wild E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 69 – 74.

[17] Murphy P M, Aha D W. UCI machine learningrepository [OL]. http://www.ics.uci.edu/~mlearn/MLRepository.html, 1992.

作者简介

杨 勃 男,1974 年生于湖南岳阳.湖南理工学院信息与通信工程学院副教授、博士、硕士生导师.主要研究方向为模式识别、机器学习等.
E-mail: ybmengshen@163.com