

二次损失函数支持向量机性能的研究

朱永生 王成栋 张优云

(西安交通大学润滑理论及轴承研究所 西安 710049)

摘 要 通过比较二次损失函数支持向量机和标准支持向量机在模式识别问题上的表现,分析了二次损失函数支持向量机的性能.实验表明这两种支持向量机对平衡数据有相似的分类能力,但二次损失函数支持向量机的优化参数更小,支持向量更多;对不平衡数据,二次损失函数支持向量机的分类准确率随不平衡度的增加而急剧下降.研究同时表明基于RM界的梯度方法对某些数据无效.文中定性分析了导致上述各种现象的原因.最后提出了一种利用黄金分割原理缩减二次损失函数支持向量机支持向量的方法,该方法冗余的支持向量数不超过一个.

关键词 支持向量机;损失函数;支持向量缩减

中图法分类号 TP306

Experimental Study on the Performance of Support Vector Machine with Squared Cost Function

ZHU Yong-Sheng WANG Cheng-Dong ZHANG You-Yun

(Theory of Lubrication and Bearing Institute, Xi'an Jiaotong University, Xi'an 710049)

Abstract The parameter optimization is one of the main study directions of SVM. Recently, a gradient descent algorithm based on RM bound has been developed, which can tune multiple parameters of SVM with squared cost function automatically and efficiently. But till now, few issues related to practical use of this type SVM have been discussed. In this paper, the performance of SVM with squared cost function on pattern recognition is studied and compared with the standard SVM. The results indicate that for balanced data, both SVMs have almost the same classifying accuracy, but the SVM with square cost function possess more support vectors and smaller optimized parameters than standard SVM. For unbalanced data, when the unbalanced degree between two classes of training samples increases, the classifying accuracy of the SVM with squared cost function decreases rapidly. The experiments also show that the gradient descent algorithm based on RM bound is not suitable for some data. Some analysis on properties of the SVM with square cost function are also included in the paper. Finally, a pruning algorithm based on golden section rule is proposed and applied to increase the sparseness of SVM with squared cost function. Using this algorithm, the number of the redundant support vectors can be reduced to one or zero.

Keywords support vector machine; cost function; support vectors pruning

1 引 言

Vapnik V等^[1]在统计学习理论中指出,学习

机器的泛化能力不但与训练过程的经验风险有关,而且与学习机器实际复杂性有关,复杂性越高,则经验风险越小但泛化能力变差,据此提出了综合考虑经验风险和学习机器复杂度的结构风险最小化

收稿日期:2002-04-19;修改稿收到日期:2003-01-02. 本课题得到国家自然科学基金(59990472)和国家“八六三”高技术研究发展计划(2001AA411310)资助. 朱永生,男,1973年生,博士研究生,主要从事机械故障诊断方法、模式识别及相关领域的研究. E-mail: zhuyongsheng@tlbi.xjtu.edu.cn. 王成栋,男,1972年生,博士研究生,主要从事机械故障诊断及信号处理的研究. 张优云,女,1947年生,博士,教授,博士生导师,研究领域为机器的状态监测及故障诊断、滑动轴承-转子系统动力学、润滑理论、现代设计等.

(SRM)归纳原则.支持向量机是实现该原则的一种通用学习算法,其原理是利用非线性函数把输入矢量映射到高维特征空间,然后在此空间中构造分类间隔最大的线性分类超平面.根据统计学习理论,分类间隔最大就保证了以最简单的学习机器实现所允许的经验风险,因而支持向量机具有很好的泛化能力.另外,高维空间中的运算只涉及矢量点积,通过引入核函数^[1],避免了高维空间计算的“维数灾难”问题,不同类型的核函数构成了不同类型的支持向量机.在模式识别、函数自回归估计、特征提取等方面的应用表明,支持向量机比由经验风险最小化原则训练得到的神经网络等具有更好的性能^[1,2],成为近年来人工智能领域研究的一个热点.

作为一种新兴的技术,支持向量机仍有许多方面尚待进一步研究,核参数(包括核函数的参数和误差惩罚参数 C)的优化选择就是其中之一.核参数是影响支持向量机性能的关键因素,常用的优化方法是参数空间穷尽搜索法^[3],即用参数空间中每一组可能的参数组合去训练和测试支持向量机,找出效果最好的参数组合.由于搜索空间很大,穷尽搜索法不但需要极大的计算量,并且当参数超过两个时,将难于实现.最近 Chapelle O 等^[3]提出利用梯度算法,在参数空间最小化支持向量机风险上界的方法,具有寻优结果准确、计算量少等优点.风险上界函数梯度的计算是该方法的关键,Duan K 等^[4]研究了估计支持向量机风险上界的不同方法以及它们在核函数参数优化中的表现,指出 RM 界(Radius-Margin bound)是连续的且容易计算的一种风险上界,但却只适用于数据完全可分的“硬间隔”^[1]支持向量机.对数据非线性可分的情况,如果将标准支持向量机中的损失函数由一阶改为二阶,则通过改写核函数,可以将其转变为“硬间隔”支持向量机的形式^[5],从而可以利用基于 RM 界的梯度算法来优化其参数.

目前,支持向量机的研究主要集中于一次损失函数的情况,对二次损失函数支持向量机的性能了解还较少.由于传统的参数优化方法存在明显的缺点,并且大部分实际问题主要是数据非线性可分的情况,为了能在实际问题中应用基于 RM 界的梯度算法高效地优化支持向量机参数,有必要对二次损失函数支持向量机作进一步的探讨和研究.

本文第 2 节介绍了二次损失函数支持向量机的“硬间隔”形式和基于 RM 界的参数优化方法,给出了针对 RBF 核的 RM 界梯度计算公式;第 3 节通过对一次损失函数支持向量机和二次损失函数支持向

量机在平衡数据和不平衡数据上分类效果的实验比较,研究了二次损失函数支持向量机的特性并进行了定性分析;第 4 节提出了一种缩减二次损失函数支持向量机支持向量方法;第 5 节为结论.

2 二次损失函数支持向量机及其参数优化方法

2.1 二次损失函数支持向量机的“硬间隔”形式

所谓“ n 次”损失函数支持向量机,是指对 SVM 的损失函数^[1]

$$F_{\sigma}(\xi) = \sum_{i=1}^l \xi_i^{\sigma} \quad (1)$$

ξ_i 的阶次 $\sigma = n$.当 $\sigma = 1$ 时,损失函数为一次函数,这是计算上可行的最小的 σ ^[1],也是目前研究最多的一种支持向量机;当 $\sigma = 2$ 时,其对应的是二次损失函数支持向量机.对二次损失函数支持向量机,采用如下形式^[5]

$$\min \Phi(\mathbf{W}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \left(\sum_{i=1}^l \xi_i^2 \right) \quad (2)$$

$$\text{s. t. } y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad (3)$$

在由映射函数 $\phi(\mathbf{x})$ 产生的特征空间中构造“最优分类超平面”^[1]

$$\mathbf{w} \cdot \phi(\mathbf{x}) + b = 0 \quad (4)$$

式中, \mathbf{w} 为分类超平面的权矢量; $\mathbf{x}_i \in R^n, i = 1, 2, \dots, l$ 为训练样本; y_i 为 \mathbf{x}_i 对应的输出标量, $y_i \in \{-1, 1\}$; ξ_i 为 \mathbf{x}_i 的误差; \mathbf{l} 为最优分类超平面的权矢量; b 为最优分类超平面的阈值; C 为误差惩罚参数; $\phi(\mathbf{x})$ 为从输入空间到特征空间的非线性映射函数.

通过引入拉格朗日乘子和核函数^[1],式(2),(3)的优化问题被转变为 wolf 对偶问题(wolf dual problem)来求解.式(2),(3)对应的对偶问题为

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{2C} \sum_{i=1}^l \alpha_i^2 \quad (5)$$

$$\text{s. t. } \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, l \quad (6)$$

式中 α_i 为拉格朗日乘子; $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ 为核函数支持向量机中常用的核函数,有径向基函数、多项式函数、Sigmoid 函数等^[1].

若令

$$Q(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{C} \delta_{i,j} \quad (7)$$

其中

$$\delta_i = \begin{cases} 1, & x_i = x_j \\ 0, & x_i \neq x_j \end{cases}$$

并把式(7)代入式(5),得到

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j Q(x_i, x_j) \tag{8}$$

$$\text{s. t. } \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, l \tag{9}$$

此时,二次损失函数支持向量机的对偶问题(式(8),(9))和数据线性可分情况时的“硬间隔”支持向量机^[1]的对偶问题具有完全相同的形式.

和一次损失函数支持向量机一样,由对偶问题式(8),(9)解出的 α_i 中仅有一部分不为零,它们所对应的训练样本就是支持向量. 相应的最优分类超平面的权矢量为

$$w = \sum_{i \in \text{支持向量}} \alpha_i y_i \phi(x_i) \tag{10}$$

最后得到的决策函数为

$$f(x) = \text{sgn}(W \cdot \phi(x) + b) = \text{sgn}\left(\sum_{i \in \text{支持向量}} \alpha_i y_i Q(x_i, x) + b\right) \tag{11}$$

其中 $x \in R^n$ 为待识别样本, $\text{sgn}(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$.

2.2 支持向量机泛化能力的 RM 界

支持向量机的泛化能力一般不能准确计算,只能对其上界作出估计,常见的估算方法有 LOO(leave-one-out)估计、K 重交叉校验、GACV 界等^[4],其中 LOO 估计是支持向量机真实泛化能力的无偏估计,而 Vapnik V 等提出的 RM 界(Radius-Margin bound)^[6]则对“硬间隔”支持向量机的泛化能力有较精确的预测^[4,7]. 由于二次损失函数的支持向量机可以转变为“硬间隔”形式,因此 RM 界对二次损失函数支持向量机的泛化能力也能很好地逼近. RM 界的定义如下

$$LOOerr \leq \frac{1}{4l} (R^2 \|w\|^2) \tag{12}$$

式中, LOOerr 为 LOO 方法得到的泛化错误率; l 为训练样本的数目; R 为特征空间中包含所有训练样本的最小的球的半径; w 为支持向量机的权矢量. 其中, R 可由以下的优化问题求得^[5]:

$$\max_{\beta} R^2 = \sum_{i=1}^l \beta_i K(x_i, x_i) - \sum_{i,j=1}^l \beta_i \beta_j K(x_i, x_j) \tag{13}$$

$$\text{s. t. } \beta_i \geq 0, \sum_{i=1}^l \beta_i = 0, i = 1, 2, \dots, l \tag{14}$$

式中, β_i 为拉格朗日乘子.

2.3 基于 RM 界的径向基二次损失函数支持向量机的参数优化方法

核参数选择不合适,则支持向量机也可能出现“过学习”或者“欠学习”现象^[8]. 目前比较有效的参数优化方法是多重交叉检验方法. 近年来,Chapelle O^[3]通过研究指出支持向量机泛化能力上界的估计函数 $f(p)$ 是核参数矢量 p 的隐函数,并给出了 $f(p)$ 梯度的计算方法,基于此提出了一种新的利用梯度下降算法优化选择支持向量机参数的方法,克服了以往多重校验方法需要计算次数多,寻优范围不易确定等缺点. 其步骤如下:

1. 给定核函数参数的初始值 p_0 ;
2. 在此初始值处训练支持向量机并计算梯度 $\left. \frac{\partial f(p)}{\partial p} \right|_{p=p_0}$;
3. 利用梯度下降算法,计算使得 $f(p)$ 减小的新的参数矢量 p_i ;
4. 以 p_i 作为新的核函数参数值,返回步 2 或者当 $f(p)$ 达到极小值时终止迭代.

该方法中 $f(p)$ 梯度的计算是关键,但前述大部分泛化能力上界的估计函数要么难于计算,要么不连续,因而其梯度难以计算或无法直接计算. 相比而言, RM 界是核参数的连续函数,并且计算简单,成为实现 Chapelle O 方法的理想的选择^[3,5].

不同的核函数对应不同类型的支持向量机^[1],其中径向基函数(RBF)支持向量机已有很广泛的研究和应用,本文以使用二次损失函数的 RBF 核支持向量机为重点进行研究,以便于和已有的研究结果比较. RBF 核函数形式为

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{15}$$

对二次损失函数 RBF 核 SVM,使用式(7)的核函数将其改造为“硬间隔”形式,则它需要优化的核参数为 (C, σ) . 在优化过程中用 $(u = \ln C, v = \ln \sigma)$ 代替 (C, σ) ,可以解除约束条件 $C > 0, \sigma > 0$,更利于优化^[5]. RM 界可用函数形式表示为

$$f(u, v) = (R^2 \|w\|^2) / 4l \tag{16}$$

其梯度为

$$\frac{\partial f(u, v)}{\partial u} = \frac{1}{4l} \left(\|w\|^2 \frac{\partial R^2}{\partial u} + R^2 \frac{\partial \|w\|^2}{\partial u} \right) \tag{17}$$

$$\frac{\partial f(u, v)}{\partial v} = \frac{1}{4l} \left(\|w\|^2 \frac{\partial R^2}{\partial v} + R^2 \frac{\partial \|w\|^2}{\partial v} \right) \tag{18}$$

其中

$$\frac{\partial R^2}{\partial u} = \frac{\ln 2}{C} \sum (\bar{\beta}_i - \bar{\beta}),$$

$$\frac{\partial R^2}{\partial v} = \frac{\ln 2}{2\sigma^2} \sum_{i,j=1}^l \bar{\beta}_i \bar{\beta}_j Q_{i,j} \|x_i - x_j\|^2,$$

$$\frac{\partial \|w\|^2}{\partial u} = \frac{\ln 2}{C} \sum_{i=1}^l \bar{\alpha}_i^2,$$

$$\frac{\partial \|w\|^2}{\partial v} = \frac{\ln 2}{2\sigma^2} \sum_{i,j=1}^l \bar{\alpha}_i \bar{\alpha}_j y_i y_j Q_{i,j} \|x_i - x_j\|^2.$$

式中, $\bar{\beta}_i$ 为约束优化问题(13),(14)的最优解; $\bar{\alpha}_i$ 为约束优化问题(8),(9)的最优解; $Q_{i,j}$ 为核函数式(7)在 (x_i, x_j) 的值。

在文献[3,5]中,上述的基于 RM 界的梯度方法在优化二次损失函数支持向量机的参数时,以很少的寻优次数获得了很好的优化效果。

3 二次损失函数支持向量机性能研究

本节讨论二次损失函数支持向量机的性能,首先分析其支持向量的分布特点,然后通过实验,研究它在几个标准测试数据集上的表现以及对不平衡数据的识别能力。

3.1 不同损失函数支持向量机支持向量分布的特点

由一次损失函数和二次损失函数支持向量机对偶问题的 KKT^[7] 条件,可以推知表 1 结果。

表 1 α_i 与 ξ_i 的关系及支持向量分布

ξ	损失函数类型		支持向量分布情况
	一次函数	二次函数	
$\xi_i = 0$	$0 \leq \alpha_i < C$	$\alpha_i = 0$	$y_i(w \cdot \phi(x_i) + b) \geq 1$
$0 < \xi_i < 1$	$\alpha_i = C$	$0 < \alpha_i < C$	$0 < y_i(w \cdot \phi(x_i) + b) < 1$
$\xi_i = 1$	$\alpha_i = C$	$\alpha_i = C$	$y_i(w \cdot \phi(x_i) + b) = 0$
$\xi_i > 1$	$\alpha_i = C$	$\alpha_i > C$	$y_i(w \cdot \phi(x_i) + b) < 0$

由表中可以看出,对于一次损失函数支持向量机,其支持向量除了错分点 $(y_i(w \cdot \phi(x_i) + b) < 0)$ 和位于分类间隔中间的样本点 $(0 < y_i(w \cdot \phi(x_i) + b) < 1)$ 外,还包括分类间隔上的点 $(y_i(w \cdot \phi(x_i) + b) = 1)$, 这种点很少),而二次损失函数支持向量机只包含前两种类型支持向量。

式(10)表明,拉格朗日系数 α_i 可以看作支持向量对最优分类超平面的影响因子. 对于一次损失函数支持向量机,除了极少量位于分类间隔上的支持向量具有较小的影响 $(\alpha_i < C)$ 外,其它的支持向量对分类超平面具有同等的重要程度 $(\alpha_i = C)$, 而二次损失函数支持向量机,由优化问题(5),(6)的 KKT 条件可以推知以下的关系

$$\alpha_i = C \xi_i \tag{19}$$

可见随着距其对应的分类间隔面距离的增加和参数 C 的增大,二次损失函数支持向量机支持向量对线性分类超平面影响增大。

图 1 所示为线性核函数时两种损失函数支持向量机对人造数据的分类结果,两类数据在对角位置各有一个错误标识的点. 这两类数据理想的分类面是一条垂直线,但从图 1 的 4 幅图中可以看出由支持向量机得到的分类面都有向错误标记的样本点倾斜的迹象,这说明支持向量机具有容错性差的缺点。

另外对照图 1(a)(b)可以发现, C 的增加没有明显影响一次损失函数支持向量机分类面的角度,只是分类面移向更优化的位置. 这是因为一次损失函数支持向量机绝大部分支持向量对分类面的影响是相同的,每个支持向量的作用并不会随着 C 的变化而变化, C 的增加表示对经验误差的惩罚增大,导致经验误差减少而分类能力逐渐达到最优,表现在分类面上就是它移向更优化的位置(当然,过大的 C 将导致学习机器复杂度增加,可能出现“过学习”现象)。

对照图 1(a)(c),可以发现图(c)中二次损失函数支持向量机最优分类面向错分支持向量(错误标记的样本点)倾斜的程度明显增加,这说明二次损失函数支持向量机受错分支持向量的影响更严重,容错性能更差;而对照图(c)(d)可以看出,随着 C 的增加,二次损失函数支持向量机的分类面的倾斜程度进一步加强,证明参数 C 对其性能的影响远远超过了对一次损失函数支持向量机的影响. 文献[8]对

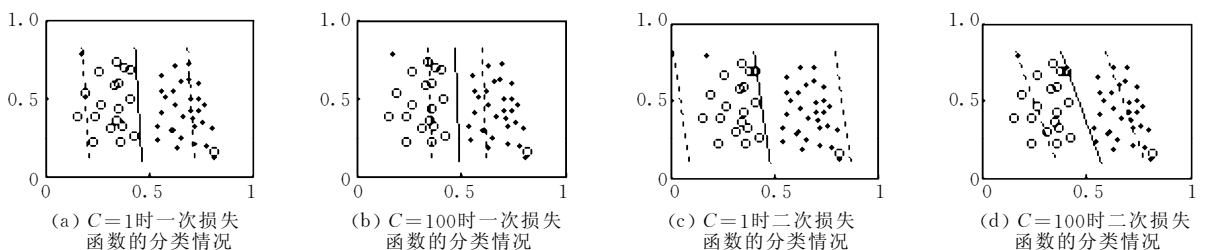


图 1 不同损失函数支持向量机最优分类面情况(线性核函数)

RBF 核函数二次损失函数支持向量机的实验也获得了同样的结果。

3.2 二次损失函数支持向量机分类实验

采用 4 种不同重叠度的数据集^①对比研究 RBF 核支持向量机在两种损失函数情况下的性能, 每种数据各有 100 组训练和测试数据, 数据集的具体情况见表 2。

表 2 数据集的情况

数据集	训练样本数	测试样本数	维数	平均错误率(%)
Thyroid	215	75	5	4.80±2.19
Banana	400	4900	2	11.53±0.66
Heart	170	100	13	15.95±3.26
BreastCancer	277	77	9	26.04±4.74

注: 平均错误率指下载数据的网站列出的 RBF 核一次损失函数支持向量机在 100 组数据上的测试结果。

对于二次损失函数支持向量机和一次损失函数支持向量机, 分别利用基于 RM 界的梯度方法和 10 重交叉检验方法优化每种数据集前 10 组数据的参数, 取其平均值作为所有 100 组数据的最优参数, 实验结果见表 3。表中“#SV”表示支持向量数。

表 3 测试结果

数据集	损失函数	优化参数		测试错误率(%)	训练错误率(%)	#SV(个)
		C	σ			
Thyroid	一次	487.00	3.30	2.83±1.89	0.45±0.39	39.00
	二次	0.89	1.14	3.47±1.87	1.19±0.49	91.20
Banana	一次	562.00	1.57	10.54±0.52	8.33±1.22	98.40
	二次	0.35	0.46	10.42±0.44	8.04±1.36	339.60
Heart	一次	22.22	34.40	15.59±3.31	13.85±1.97	69.00
	二次	0.42	4.39	16.00±3.18	12.14±1.86	156.54
Breast Cancer	一次	91.63	3.16	23.00±4.37	8.04±0.91	159.00
	二次	0.25	7.18	26.65±4.61	25.56±1.71	212.20

实验中观察到的第二个现象是二次损失函数支持向量机往往具有更小的优化参数。由于二次损失函数支持向量机中错分向量对其性能有很大影响且影响力大小和 C 成正比, 因此二次损失函数支持向量机有使得经验误差更小(错分支持向量尽量减少)和使得参数 C 更小以减少上述不利影响的趋势。太大的 C 将导致对错分样本过适应而使其泛化能力变差, 因此二次损失函数 SVM 的优化参数中 C 相对较小; 同时, 依据结构风险最小化的原理, 为了得到好的泛化能力, 即使在 C 相对小的情况下, 二次损失函数支持向量机还必须保证较小的经验风险, 这种情况只有在更复杂的特征空间才可能出现, 对于 RBF 核函数, 参数 σ 越小, 其对应的特征空间的 VC 维越高, 特征空间越复杂, 因此一般情况下, RBF 核二次损失函数支持向量机的参数 σ 也较小。另一方面, RM 界的极小值点对应的参数比支持向量机

实验中, 利用基于 RM 界的梯度方法优化参数的二次损失函数支持向量机在前三种数据集上获得了和一次损失函数支持向量机相似的分类效果, 这说明大部分情况下, 二次损失函数和一次损失函数支持向量机具有相似的分类能力, 并且基于 RM 界的梯度方法是一种很有效的参数优化方法。尽管对 BreastCancer 数据集, 表 3 显示二次损失函数支持向量机的测试错误率比一次损失函数支持向量机增加了大约 3%, 但在文献[8]中, 采用一种新的方法重新优化参数后($C=7.6, \sigma=4.1$), 二次损失函数支持向量机在 BreastCancer 数据集仍然获得了 23.54%±4.52% 的分类错误率, 这一方面进一步说明两种支持向量机具有相同的分类效果, 另一方面说明基于 RM 界的梯度方法对有些数据不适用, 这主要是由于 RM 界的不适用引起的: 如果数据在特征空间中的分布形状接近球状时, RM 界比较准确; 相反, 如果数据分布形状和球状相差得越远, RM 界就越不准确^[9], 从而导致按照 RM 界最小化标准选择的优化参数不准确。

实际测试错误率极小值点对应的参数小^[6], 这也是导致由基于 RM 界的梯度方法得到的优化核参数偏小的原因之一。

为了更清楚了解二次损失函数支持向量机中核参数的影响, 图 2 给出了在固定 C 或者 σ 时(取表 3 中的值), 两种损失函数支持向量机实际泛化能力随核参数变化的趋势(在 Banana 第 9 组数据上实验的结果, 由于测试样本的数目远大于训练样本, 可以把测试结果看作是支持向量机的实际泛化能力)。主要目的是比较变化趋势, 因此对结果进行归一化处理。由图中可以看出, 两种损失函数支持向量机泛化能力都随 σ 的增加而减小(测试错误率增加), 趋势大致相同, 这主要是 σ 增加导致特征空间的维数降低, 因而 RBF 核支持向量机的学习能力(复杂性)过低,

① <http://ida.first.gmd.de/~raetsch/data/benchmarks.htm>

从而使其泛化能力降低;而对于 C ,从整体上看,随 C 的增加,一次损失函数支持向量机泛化能力增大(测试错误率降低),这主要是由于其经验风险逐渐减小,导致结构风险趋向最小的结果;而对二次损失函数支持向量机,随 C 的增加,尽管经验风险也在减小,但对错分支持向量的过适应现象越来越严重,分类面受错分样本的影响大大增加,导致泛化能力逐渐减小,当 C 大于一定值时,两种支持向量机的复杂性都达到了特征空间允许的最大值,因而 C 的变化对其泛化能力几乎不再有影响,测试错误率保持在一个比较稳定的水平。

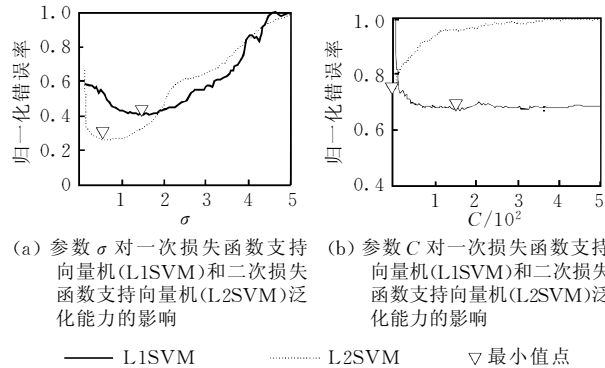


图 2

实验中观察到的第三个现象是二次损失函数支持向量机支持向量数远大于一次损失函数支持向量机.这主要是二次损失函数支持向量机引入二次损失函数,导致了更多的具有误差项 ξ_i 的样本的出现.由于太多的支持向量将导致分类速度变慢,这也正是二次损失函数支持向量机的一个主要缺陷。

3.3 二次损失函数支持向量机对不平衡数据的分类实验

不平衡数据是指训练样本集合中,不同类别样本数目相差较大的情况.由于二次损失函数支持向量机的支持向量占训练样本很大比例,当不同类别样本数目相差较大时,不同类别支持向量的数目就会存在明显差异,这种差异将导致分类面发生很大变化。

本文用 Banana 数据集第 4 组训练数据来构造不平衡数据,方法是保持其中一类样本的数目(用 $\#Samples_2$ 表示)不变,随机选择一定数目的另外一类样本(用 $\#Samples_1$ 表示),两者共同组成新的训练样本.定义

$$Unbalance = 1 - \frac{\#Samples_1}{\#Samples_2} \quad (20)$$

则 $Unbalance$ 表示了两类样本的不平衡程度, $Un-$

$balance$ 越大,样本的不平衡程度越高.本文分别取 $Unbalance = 0.1, 0.2, \dots, 0.9$ 九组不平衡数据实验,对每一组数据仍然采用基于 RM 界的梯度方法和 10 重交叉检验方法分别优化两种损失函数支持向量机的参数然后训练,最后用 Banana 数据集第 4 组测试数据测试,测试错误率(图中用“ Re ”表示)见图 3。

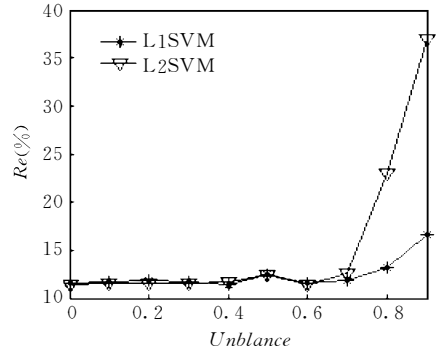


图3 二次损失函数 RBF 核 SVM(L2SVM)和一次损失函数 RBF 核 SVM(L1VM)对不平衡数据分类错误率

由图 3 可以看出,在不平衡程度较小时,一次损失函数支持向量机和二次损失函数支持向量机有比较接近的分类错误率,但当不平衡度增加时,二次损失函数 SVM 的分类错误率急剧增加,远远大于一次损失函数 SVM 的分类错误率,这说明二次损失函数 SVM 不适合不平衡程度高的数据。

图 4 绘出了 $Unbalance = 0.9$ 时 RBF 核一次损失函数支持向量机(L1SVM)和二次损失函数支持向量机(L2SVM)分类面的情况.图(a)显示,即使对高度不平衡的数据,一次损失函数支持向量机的分类面仍然保持了比较理想的形状,这是因为对于一次损失函数支持向量机,其最优分类面只需用靠近分类面、数目很少的支持向量就可以决定,即使对于高度不平衡的训练样本集,只要原来靠近分类面的样本大部分被包含在训练样本中,则分类面的形状仍然可以被较好地保持;对于二次损失函数 SVM,图(b)显示,在数据高度不平衡的情况下,它的分类面向数目少的样本的中心靠近,这种变化的原因可以用约束条件式(9)来解释:由于两类支持向量数目相差很大,为了保证式(9)成立,数目少的支持向量应该有更大的拉格朗日系数,而式(19)表明拉格朗日系数只有在参数 C 较大或者发生错分的情况下才可能大,由于对两类样本使用了相同的 C ,因此数目少的样本必须有更多的错分情况,导致二次损失函数支持向量机的分类面向数目少的样本中心收缩。

这一现象同时表明,对不平衡数据,通过对两类样本采取不同的误差惩罚参数 C ,即对数目少的样

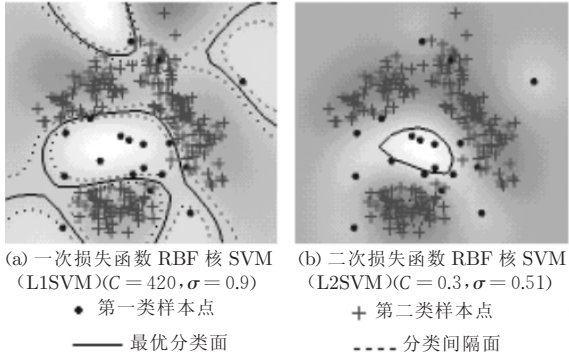


图 4 L1SVM 和 L2SVM 对 $Unbalance=0.9$ 的不平衡数据分类效果图

本采用较大的 C 可以改善分类的效果. 限于篇幅, 本文对此未多加讨论.

4 缩减二次损失函数支持向量机支持向量的方法

二次损失函数支持向量机一个明显的缺点是有更多的支持向量, 这使得它的分类速度比一次损失函数支持向量机更慢. 针对此问题, 许多学者展开了广泛的研究^[10,11]. 其中 Suykens 等^[11]提出一种针对最小二乘支持向量机(Least Squares Support Vector Machines)的方法, 其原理是将反映支持向量重要程度的拉格朗日乘子 α_i 按照从大到小的次序排序, 然后以一定的比例将排在最后的支持向量剔除, 用剩下的支持向量训练支持向量机, 然后再剔除, 重复此过程直至泛化能力下降到预先设定的水平.

由于二次损失函数的支持向量机和最小二乘支持向量机在优化问题上的相似性, Suykens 的方法也可以有效地在二次损失函数支持向量机中应用. 但这种方法移去的支持向量的比例难于确定, 比例大时, 剩余的支持向量必然有较大的冗余; 比例过小, 需要的循环次数将大大增加, 为此本文引入优化问题中的“黄金分割”方法, 实现支持向量的变步长剔除, 该算法最后冗余的支持向量数不超过一个. 算法如下:

1. 用给定的训练样本训练二次损失函数支持向量机, 得到 N 个支持向量, 令 $N_{old} = N$.
2. 将支持向量按照其拉格朗日乘子 α_i 的大小降序排列, 取前 $N_{new} = I(0.618 \times N_{old})$ 个支持向量为新的训练集训练和测试, 其中 $I(x)$ 为取整函数;
3. 如果得到的支持向量机的性能优于预定的目标值, 则令 $N_{old} = N_{new}$, 返回步 2, 否则, 执行步 4.
4. 若新支持向量机的性能低于预定目标值, 则令 $N_{new} = N_{new} + I(0.618 \times (N_{old} - N_{new} + 1))$, 用前 N_{new} 个支持向量训

练和测试, 返回步 3.

5. 当 $(N_{old} - N_{new}) < 1$, 终止循环.

对于 BreastCancer 数据, 采用文献[8]中的优化参数($C=7.6, \sigma=4.1$), 其它 3 种数据采用表 3 中的优化参数, 然后允许二次损失函数支持向量机的性能下降 1%~30%, 采用上述的缩减方法缩减其支持向量数. 由于在不同的错误率阈值时可能得到相同的实际测试错误率和相同的支持向量数, 因此表 4 列出了每个数据集在不同的实际测试错误率或者不同支持向量数时的结果. 表中, “#SV”表示支持向量数, “Re”表示实际测试错误率, 数据集名称后面的数字表示采用该数据集第几组数据进行实验.

表 4 支持向量缩减结果

	Breast Cancer(9)		Heart (7)		Banana (1)		Thyroid (3)	
	#SV	Re(%)	#SV	Re(%)	#SV	Re(%)	#SV	Re(%)
1	174	22.08	154	18	335	11.33	92	4.00
2	139	22.08	73	18	156	11.12	26	4.00
3	125	24.68	72	23	151	12.29	19	6.67
4	121	25.97	69	25	147	13.67	18	12.00
5	117	23.38	90	20	90	10.73	15	5.33

表的第 1 行是进行缩减前的支持向量数和测试错误率, 第 2 行是允许错误率下降 1% 时的情况. 第 1 行和第 2 行的结果, 可以发现 4 种数据集在测试错误率没有增加的情况下 (Banana 数据集反而有所减小), 支持向量数分别下降了 20.11%, 52.6%, 53.4%, 71.74%, 这说明二次损失函数支持向量机包含很多的冗余支持向量, 而且冗余程度随着样本重叠度的减小而增加. 利用本文提出的方法, 可以在对性能几乎没有影响的情况下, 大大减少冗余支持向量; 而第 3 和第 4 行的试验结果则说明在冗余支持向量被剔除后, 剩余支持向量对二次损失函数支持向量机的性能有很大的影响, 即使被剔除很少的几个, 也会大大增加分类错误率, 比如对于 Heart 数据集, 当支持向量数由 73 减少到 72 时, 其分类错误率由 18% 增加到了 23%.

表中第 5 行是利用表 3 的优化参数对一次损失函数支持向量机训练和测试的结果, 和前面各组结果比较, 可以发现即使经过上述的缩减程序, 要达到和一次损失函数支持向量机相同的分类错误率, 二次损失函数支持向量机仍然需要更多的支持向量. 因此寻找更好的缩减支持向量的方法仍然是二次损失函数支持向量机研究的一个主要内容.

5 结 论

本文在介绍了二次损失函数支持向量机“硬间

隔”形式以及基于 RM 界的参数优化方法的基础上,通过在几个标准数据集上的实验重点研究了二次损失函数支持向量机的特性,得到的主要结论有:

(1)对于平衡数据,二次损失函数支持向量机具有和一次损失函数支持向量机相同的分类效果,而对于不平衡数据,二次损失函数支持向量机的分类能力随着不平衡度的增加而急剧降低.

(2)一般情况下二次损失函数支持向量机具有更小的优化参数和更多的支持向量.

(3)基于 RM 界的梯度算法在大部分情况下是一种有效的参数优化方法,但对某些数据并不适用.

(4)本文提出的缩减二次损失函数支持向量机支持向量的方法,可以在对性能影响很小的情况下,最大程度减少支持向量的冗余程度.

参 考 文 献

- 1 Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1998
- 2 Zhang Xue-Gong. Introduction to statistical learning theory and support vector machines. Acta Automatica Sinica, 2000, 26(1): 32~42 (in Chinese)
(张学工. 关于统计学习理论与支持向量机. 自动化学报, 2000, 26(1): 32~42)
- 3 Chapelle O, Vapnik V N, Bousquet O *et al.* Choosing multiple parameters for support vector machines. Machine Learning, 2002, 46(1): 131~159
- 4 Duan K, Keerthi S S, AN Poo. Evaluation of simple perform-

ance measures for tuning SVM hyperparameters. Department of Mechanical Engineering, National University of Singapore; Technical Report, Control Division Technical Report CD-01-11, 2001

- 5 Keerthi S S. Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. Department of Mechanical Engineering, National University of Singapore; Technical Report, Control Division Technical Report CD-01-12, 2001
- 6 Vapnik V N, Chapelle O. Bounds on error expectation for support vector machine. Neural Computation, 2000, 12(9): 2013~2036
- 7 Burges C J C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2(2): 121~167
- 8 Zhu Yong-Sheng. Support vector machines and its application in faults pattern recognition of mechanical equipments [Ph D dissertation]. Xi'an Jiaotong University, Xi'an, 2003 (in Chinese)
(朱永生. 支持向量机及其在机械故障模式识别中的应用 [博士学位论文]. 西安交通大学, 西安, 2003)
- 9 Scholkopf B, Mika S, Burges C J C *et al.* Input spaces vs. feature space in kernel-based methods. IEEE Transactions on Neural Networks, 1999, 10(5): 1000~1017
- 10 Burges C J C, Scholkopf B. Improving the accuracy and speed of support vector machines. Neural Information Processing Systems, 1997, 9(7): 375~381
- 11 Suykens J A K, Lukas L, Vandewalle J. Sparse least squares support vector machine classifiers. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2000), Geneva, Switzerland, 2000. 2757~2760



ZHU Yong-Sheng, born in 1973, Ph. D. candidate. His research interests mainly include the method of mechanical equipment faults diagnosis, pattern recognition and related area.

WANG Cheng-Dong, born in 1972, Ph. D. candidate. His research interests mainly include the mechanical equipment faults diagnosis, signal processing.

ZHANG You-Yun, born in 1947, Ph. D., professor and Ph. D. supervisor. Her research interests include condition monitoring and faults diagnosis of machinery, rotor-bearing system dynamics, theory of lubrication and modern design.