SVM最优分类面相对位置的修正

周 皓 李少洪

(北京航空航天大学 电子信息工程学院, 北京 100191)

摘 要: 通过放宽标准支持向量机(SVM, Support Vector Machines)中类别边界至分类面等间隔的约束,保持两类函数间隔之和不变的条件,在支持向量机思想框架下得出分类面依样本分布进行调整的新型支持向量机,其对偶形式与标准支持向量机完全相同,从而在理论上进一步完善了支持向量机. 在此基础上,提出使类别的函数间隔正比于样本标准差的具体算法——方差修正法,达到最优分类面的相对位置依样本方差而调整之目的. 从统计意义上来说,方差修正法在分类精度上有所提高,但计算量增加不多.

关 键 词: 支持向量机; 修正; 计算复杂度; 分类器

中图分类号: TP 181; TP 391. 4

文献标识码: A **文 章 编 号**: 1001-5965(2009)11-1302-04

Relative position modification of SVM's optimal hyperplane

Zhou Hao Li Shaohong

(School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Through releasing the equal-margin constraint in the standard support vector machines (SVM), keeping the sum of the binary-class function margins, a new SVM was gotten within the framework of SVM. The separating hyperplane of the new SVM can be adjusted as per the distribution of the binary-class samples, and its dual express is same as the standard SVM. Thus, the SVM was further improved theoretically. On the basis of the new SVM, a concrete algorithm, variance modification algorithm, was proposed. In the variance modification algorithm, the binary-class margins are in proportion to the standard deviation of binary-class samples. The goal of adjusting the optimal separating hyperplane as per sample's variance is attained through the variance modification algorithm. Statistically, errors are reduced through the variance modification algorithm, while the computational complexity is not increased much.

Key words: support vector machines; modification; computational complexity; classifiers

对于一个分类面位于两类边界中央的二类线性分类器而言,其分类性能并非最优[1],而当前的支持向量机(SVM, Support Vector Machines)恰属于这种情况.

SVM问世以后,虽然有学者发现了SVM的分类面问题并展开了相关的研究,如文献[2-4],但他们都没能从SVM自身上给出解释和解决,而都是从大间隔分类器的角度出发,在不同程度上跳出SVM思想框架,重新设计学习机. 这样得到的分类器在一定程度上丧失了SVM的一些优良性质,如解的稀疏性,使分类器判决输出的计算量较标准SVM增加很多,而且这些算法训练过程更为复杂,从性价比的角度来看并不可取.

本文在SVM思想框架下,对SVM分类面问题加以解释和解决,给出修正SVM分类面的总体算法,并提出具体的方差修正算法,从性价比的角度来看,本文方法以较小的代价实现了SVM性能的提升.

1 SVM分类面问题的起因

图1清楚地说明了SVM分类面问题.

收稿日期: 2008-10-20

作者简介: 周 皓(1970-),男,辽宁沈阳人,博士生, zhouhao@ee.buaa.edu.cn.

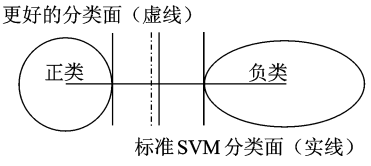

图 1　SVM分类面问题示意图

这是一个很常见的问题. 产生原因在于标准SVM是以与分布无关的 PAC (Probably Approximately Correct)模型为基础,其 VC泛化性理论的基础概念函数间隔[5]没有考虑样本分布,而直接使用欧氏测度来确定最优分类面.

Vapnik的最优分类面[6]要求满足两个条件: ①无误差分类; ②到类边界点距离最大. 对于第 2个条件,不考虑样本分布仅以欧氏距离作为距离测度时,有两个结论:最优分类面必然位于两类边界的正中;两类边界间的距离最大 (即 SVM的分类间隔最大). 因此对于分类面 $(\omega \cdot x) + b = 0$,有

$$\gamma^+ = y^+ ((\omega \cdot x^+) + b) = (\omega \cdot x^+) + b = 1 \quad (1)$$

$$\gamma^- = y^- ((\omega \cdot x^-) + b) = -((\omega \cdot x^-) + b) = 1 \quad (2)$$

γ^+, γ^- 分别为正负类样本临界点 x^+, x^- 的函数间隔,此处为 1; y^+, y^- 分别为类别标签. 注意:两类的函数间隔之和为 2

再按解析几何知识,得分类间隔为 $\frac{2}{\|\omega\|_2}$,将其最大化即可得出 SVM的优化问题.

式 (1)、式 (2)表明,两类边界点到最优分类面等函数间隔 (为 1)的约束条件事先固定了最优分类面相对于两类边界的位置 (居中). 这恰是导致 SVM的分类面问题的原因.

2　*SVM* 分类面的修正

2 1　*SVM* 分类面修正的总体思想

修改 Vapnik最优分类面的第 2个条件为:使两类的分类间隔最大. 即可以将各类函数间隔为 1的约束条件调整为与样本具体情况有关的量,在保持两类的函数间隔之和为 2的总体约束下,利用标准 SVM的推导方法得出分类面相对位置可调的修正 SVM. 保证分类间隔的最大化正是 SVM的最优分类面所要反应的实质 ——结构风险最小化思想.

VC泛化性理论中分类器推广能力的上界[5]依赖于"目标间隔 γ".当把整个样本集分为两类考虑时,对于不同的类别给出不同的目标间隔,如 γ^+, γ^-,若取泛化性定理中的 $\gamma = \min(\gamma^+, \gamma^-)$,并不违背泛化性理论有关目标间隔的规定,自然泛化性理论中有关推广能力上界的结论在这种情况下依然适用,但此时却综合考虑了样本集的分布情况,为进一步缩小泛化误差创造了可能性.

2 2　*SVM* 分类面修正思想的实现

设 γ^+, γ^- 是与样本分布情况有关的量,由上节讨论,限定

$$\gamma^+ + \gamma^- = 2 \quad (3)$$

重写式 (1)、式 (2)后分类间隔仍为

$$2r = \frac{|(\omega \cdot x^+) + b|}{\|\omega\|_2} + \frac{|(\omega \cdot x^-) + b|}{\|\omega\|_2} = \frac{\gamma^+ + \gamma^-}{\|\omega\|_2} = \frac{2}{\|\omega\|_2}$$

由此得到的优化问题为

$$\min \quad \frac{1}{2}\|\omega\|_2^2 = \frac{1}{2}(\omega \cdot \omega) \quad (4)$$

$$\text{s t} \quad y_i \times ((\omega \cdot x_i) + b) \geq \gamma^+, \ i = 1, 2, \cdots, N^+$$

$$y_j \times ((\omega \cdot x_j) + b) \geq \gamma^-, \ j = 1, 2, \cdots, N^-$$

N^+、N^- 分别为正、负类样本数,样本总数设为 l, (x_n, y_n) 为样本点, $y_n \in \{1, -1\}$ 为 x_n 的类别标签.

当 $\gamma^+ = \gamma^- = 1$ 时,式 (4)就成为标准 SVM.

下面看一下式 (4)的对偶问题. 式 (4)的拉格朗日函数为

$$L(\omega, b, \alpha^+, \alpha^-) = \frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{N^+}\alpha_i^+ ((\omega \cdot x_i^+) + b - \gamma^+) + \sum_{i=1}^{N^-}\alpha_i^- ((\omega \cdot x_i^-) + b + \gamma^-)$$

令

$$\frac{\partial}{\partial b} = -\sum_{i=1}^{N^+}\alpha_i^+ + \sum_{i=1}^{N^-}\alpha_i^- = 0 \quad (5)$$

(注:式 (5)的成立要求 $\frac{\partial \gamma^+}{\partial b} \neq 1$,这一条件不难实现.)

再令 $\frac{\partial}{\partial \omega} = 0$,将式 (5)代入得

$$\omega = \sum_{i=1}^{N^+}\alpha_i^+ x_i^+ - \sum_{i=1}^{N^-}\alpha_i^- x_i^- \quad (6)$$

另外,由式 (3)、式 (5)可得

$$\gamma^+ \sum_{i=1}^{N^+}\alpha_i^+ + \gamma^- \sum_{i=1}^{N^-}\alpha_i^- = 2\sum_{i=1}^{N^-}\alpha_i^- = \sum_{i=1}^{N^+}\alpha_i^+ + \sum_{i=1}^{N^-}\alpha_i^- = \sum_{i=1}^{l}\alpha_i$$

将式 (5)、式 (6)及上式代回式 (4)的拉格朗日函数式,并引入核函数,最终得到式 (4)的对偶表示式为

$$\min f(\alpha) = \frac{1}{2}\alpha^{\mathrm{T}}(YKY)\alpha - \alpha^{\mathrm{T}}e \quad (7)$$

$$\text{s.t. } \alpha^{\mathrm{T}}Ye = 0$$

$$\alpha \geq \mathbf{0}$$

其中，$\alpha = (\alpha_1^+, \cdots, \alpha_{N^+}^+, \alpha_1^-, \cdots, \alpha_{N^-}^-)^{\mathrm{T}}$；$Y = \mathrm{diag}(1, \cdots, 1, -1, \cdots, -1)$，即正负类类别标签构成的对角阵；$e = (1)_{l\times 1}$；K为核矩阵.

由式(7)可见式(4)的对偶问题与标准SVM完全相同，则由式(4)确定的分类超平面与标准SVM具有相同的法向量ω，只是二者的偏置b有所不同.

式(7)是一个硬间隔SVM的对偶式，对于软间隔SVM，也可得出同样结论，此处不再赘述. 相对于标准SVM，式(7)并没有增加计算复杂度.

3 方差修正法

在确定分类面的偏置b时需要引入对γ^+，γ^-的新的约束. 标准SVM是让它们比例为1，再加上式(3)而得到. 不同的约束会衍生出不同的SVM分类面的修正方法，本文提出依样本方差修正SVM分类面的方法——方差修正法.

3.1 方差约束的加入

对于线性可分样本采用距离分类法时，分类面$(\omega \cdot x) + b = 0$至样本均值点的马氏距离应相等. 设在ω方向上正、负类样本标准差为：$\sigma^+ = \sqrt{\omega^{\mathrm{T}}\Sigma^+\omega}$，$\sigma^- = \sqrt{\omega^{\mathrm{T}}\Sigma^-\omega}$，则有$\frac{(\omega \cdot \bar{x}^+) + b}{\sigma^+} = \frac{-[(\omega \cdot \bar{x}^-) + b]}{\sigma^-}$，将此式代入式(4)负类样本的约束条件中，得$((\omega \cdot \bar{x}^+) + b) \geq \frac{\sigma^+}{\sigma^-}\gamma^-$，$\bar{x}^+$，$\bar{x}^-$分别为正负类样本均值点，取正类样本均值点至分类面距离下限相等(因为ω是变化的)，则得出本文为最优分类面引入的新约束条件：两类对最优分类面的函数间隔与其在ω方向上的标准差成正比，即

$$\frac{\gamma^+}{\gamma^-} = \frac{\sigma^+}{\sigma^-} \quad (8)$$

由式(3)、式(8)改写式(4)为

$$\min \frac{1}{2}\|\omega\|_2^2 = \frac{1}{2}(\omega \cdot \omega) \quad (9)$$

$$\text{s.t. } ((\omega \cdot x_i^+) + b) \geq \frac{2\sigma^+}{\sigma^+ + \sigma^-}$$

$$i = 1, 2, \cdots, N^+$$

$$-((\omega \cdot x_i^-) + b) \geq \frac{2\sigma^-}{\sigma^+ + \sigma^-}$$

$$i = 1, 2, \cdots, N^-$$

当$\sigma^+ = \sigma^-$时，式(9)即为标准SVM. 式(9)中γ^+，γ^-与b无关，故其对偶式与式(7)相同.

3.2 σ^+，σ^-的对偶及核函数表示式

将核矩阵分块表达为：$K = \begin{pmatrix} K^+ \\ K^- \end{pmatrix}$，其中$K^+$(或$K^-$)为正(或负)类样本与所有样本的核函数构成的子块，经推导得

$$(\sigma^+)^2 = \omega^{\mathrm{T}}\Sigma^+\omega = \alpha^{\mathrm{T}}M^+\alpha \quad (10)$$

$$(\sigma^-)^2 = \omega^{\mathrm{T}}\Sigma^-\omega = \alpha^{\mathrm{T}}M^-\alpha \quad (11)$$

其中

$$M^+ = Y\left\{\frac{1}{N^+ - 1}(K^+ - e^+k^{+\mathrm{T}})^{\mathrm{T}}(K^+ - e^+k^{+\mathrm{T}})\right\}Y$$

$$M^- = Y\left\{\frac{1}{N^- - 1}(K^- - e^-k^{-\mathrm{T}})^{\mathrm{T}}(K^- - e^-k^{-\mathrm{T}})\right\}Y$$

k^+(或k^-)分别为K^+(或K^-)各列均值构成的行向量，$e^+ = (1)_{N^+\times 1}$，$e^- = (1)_{N^-\times 1}$.

式(10)、式(11)实现了σ^+，σ^-的对偶表示. 这一点与文献[2-4]有类似的结论，不同之处在于：两式的计算都是在支持向量确定后进行，运算量已经有较大的下降.

3.3 偏置b的求解

式(7)求解之后，由式(10)、式(11)即可得出σ^+，σ^-，任选一个$\alpha_i \neq C$的支持向量由KT(Kuhn-Tucker)条件即可解出偏置b.

4 仿真试验

通过3种样本分布情况下的仿真，比较方差修正SVM(本文)，标准SVM，M^4(Maxi-Min Margin Machine)算法(文献[3])及MPM(Minimax Probability Machine)算法(文献[2])的性能情况. 以matlab优化工具箱的优化函数编制各算法仿真程序.

例1 仿照文献[3]的仿真数据：正类为均值[-3, 0]，协方差阵[0.5, 0; 0, 8]的高斯分布；负类为均值[4, 0]，协方差阵[6, 0; 0, 1]的高斯分布. 各生成20个样本为训练样本，以60, 100, 500, 1 000个测试样本分别测试，结果见表1.

例2 正类[0, 0]为中心点，30点，协方差阵[0.1, 0; 0, 0.1]；负类[2, 2]为中心，30点，协方差阵[1, 0; 0, 1]. 以100, 500, 1 000个测试样本分别测试，结果见表2.

例3 正类[0, 0]为中心点，50点，协方差阵[1, 0; 0, 1]；负类[2, 2]为中心，50点，协方差阵[1, 0; 0, 1]. 以100, 500, 1 000个测试样本分别测试，结果见表3.

表 1　例 1的仿真结果

样本数	算法			
	方差修正 SVM	标准 SVM	M^4算法	MPM算法
60	1.6667 ($n_1=2$, $n_2=0$)	1.6667 ($n_1=2$, $n_2=0$)	1.6667 ($n_1=2$, $n_2=0$)	3.3333 ($n_1=2$, $n_2=2$)
100	1.5 ($n_1=2$, $n_2=1$)	2.5 ($n_1=5$, $n_2=0$)	2.5 ($n_1=5$, $n_2=0$)	4 ($n_1=7$, $n_2=1$)
500	2.6 ($n_1=20$, $n_2=6$)	3 ($n_1=24$, $n_2=6$)	3.2 ($n_1=26$, $n_2=6$)	3.7 ($n_1=30$, $n_2=7$)
1 000	2.3 ($n_1=31$, $n_2=15$)	3.05 ($n_1=48$, $n_2=13$)	3.2 ($n_1=49$, $n_2=15$)	3.75 ($n_1=51$, $n_2=24$)

注:表中内容为各算法的误差,包含 2项内容, ①测试误差百分率; ②正类 (负类)的误分样本数. 如: 1.6667 ($n_1=2$, $n_2=0$)表示测试误差 1.6667%,正类误分 2个样本 ($n_1=2$),负类误分 0个样本 ($n_2=0$)

表 2　例 2的仿真结果

样本数	算法			
	方差修正 SVM	标准 SVM	M^4算法	MPM算法
100	0.5 ($n_1=0$, $n_2=1$)	2 ($n_1=0$, $n_2=4$)	2.5 ($n_1=1$, $n_2=4$)	1.5 ($n_1=0$, $n_2=3$)
500	1 ($n_1=3$, $n_2=7$)	0.9 ($n_1=0$, $n_2=9$)	1.1 ($n_1=1$, $n_2=10$)	1 ($n_1=0$, $n_2=10$)
1 000	0.75 ($n_1=3$, $n_2=12$)	1.15 ($n_1=0$, $n_2=23$)	1.4 ($n_1=3$, $n_2=25$)	1 ($n_1=0$, $n_2=20$)

表 3　例 3的仿真结果

样本数	算法			
	方差修正 SVM	标准 SVM	M^4算法	*MPM*算法
100	5.5 ($n_1=6$, $n_2=5$)	5.5 ($n_1=9$, $n_2=2$)	16.5 ($n_1=15$, $n_2=18$)	5.5 ($n_1=8$, $n_2=3$)
500	7.9 ($n_1=38$, $n_2=41$)	9.6 ($n_1=73$, $n_2=23$)	17 ($n_1=87$, $n_2=83$)	8.6 ($n_1=51$, $n_2=35$)
1 000	7.7 ($n_1=63$, $n_2=91$)	8.55 ($n_1=122$, $n_2=49$)	16.25 ($n_1=156$, $n_2=169$)	7.9 ($n_1=89$, $n_2=69$)

5　结束语

SVM分类面并不是一个分类器的最优分类面,其性能存在着提升的空间. 本文考虑到样本的分布情况,依结构风险最小化原则修改了 Vapnik最优分类面定义,进而修改了标准 SVM的约束条件,在 SVM框架下得出修正的 SVM,从理论上进一步完善了 SVM. 在此思想基础上,提出方差修正 SVM算法,仿真结果表明方差修正法在性能上有所提高,但计算复杂度增加较小.

本文主旨在于由式 (4)指出标准 SVM中存在的问题,其性能存在提高的空间. 方差修正法只是解决这一问题的方法之一. 寻求其它更好的方法是本文下一步的工作,如以样本先验概率估计的比值做为 γ^+, γ^-的比例因子,或者考察 SVM分离区样本情况来调整 γ^+, γ^-等方法. 文献 [7]虽通过考察分离区样本情况得出 SVM分类面的修正算法,但仍未能从 SVM框架下给出该算法的依据.

考虑样本分布的学习机在性能上的提高是具有统计意义的,如果训练样本与测试样本分布吻合较差,所得分类器效果并不见得优于标准 SVM. 仿真中各算法性能因样本集的变化具有一定的波动,即是出于这个原因.

参考文献 (References)

[1] 边肇祺,张学工. 模式识别 [M]. 北京:清华大学出版社, 2000
Bian Zhaoqi, Zhang Xuegong. Pattern recognition [M]. Beijing: Tsinghua University Press, 2000 (in Chinese)

[2] Gert R G Lanckriet, Laurent El Ghaoui, Chiranjib Bhattacharyya, et al. A robust minimax approach to classification[J]. Journal of Machine Learning Research, 2002, 3: 555 - 582

[3] Huang Kaizhu, Yang Haiqin, King Irwin, et al. Maxi-min margin machine: learning large margin classifiers locally and globally[J]. IEEE Transactions on Neural Networks, 2008, 19(2): 260 - 272

[4] Wang Defeng, Yeung D S, Tsang E C C. Probabilistic large margin machine[C]// Proceeding of the Fifth International Conference on Machine learning and Cybernetics. Dalian: [s.n.], 2006: 2190 - 2195

[5] Nello Cristianini, John Shawe-Taylor. 支持向量机导轮 [M]. 李国正,王猛,曾华军,译. 北京:电子工业出版社, 2004
Nello Cristianini, John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods[M]. Translated by Li Guozheng, Wang Meng, Zeng Huajun. Beijing: Publishing House of Electronic Industry, 2004 (in Chinese)

[6] Vladimir N Vapnik. An overview of statistical learning theory [J]. IEEE Transactions on Neural Networks, 1999, 10(5): 988 - 999

[7] 文传军,詹永照. 基于自调节分类面 SVM的平衡不平衡数据分类 [J]. 系统工程, 2009, 27(3): 110 - 114
Wen Chuanjun, Zhan Yongzhao. Balance and imbalance data set classification based on self-adjusting classification-plane SVM [J]. Systems Engineering, 2009, 27(3): 110 - 114 (in Chinese)

(编　辑:娄　嘉)