计算机研究与发展 ISSN 1000-1239/CN 11-1777/TP
Journal of Computer Research and Development 45(8): 1346-1353, 2008

多项式光滑的支持向量机一般模型研究

熊金志[1] 袁华强[1] 彭 宏[2]
[1](东莞理工学院软件学院 广东东莞 523106)
[2](华南理工大学计算机科学与工程学院 广州 510641)
(dgxiongjz@126.com)

A General Formulation of Polynomial Smooth Support Vector Machines

Xiong Jinzhi[1], Yuan Huaqiang[1], and Peng Hong[2]
[1](*Software College, Dongguan University of Technology, Dongguan, Guangdong* 523106)
[2](*College of Computer Science and Engineering, South China University of Technology, Guangzhou* 510641)

Abstract Yuan *et al*. used a polynomial function as smoothing function, and proposed a polynomial smooth support vector machine (PSSVM) in 2005, which improved the performance and efficiency of SVM for classification. Using the technique of interpolation functions, Xiong *et al*. developed a recursive formula to obtain a new class of smoothing functions, and solved the problems of existence and seeking better smoothing functions in 2007. However, problems still exist in looking for other smooth models and the general formulation of the polynomial smooth support vector machines. A class of polynomial functions is applied as new smoothing functions, and a dth-order polynomial smooth support vector machine (dPSSVM) is proposed using the smoothing technique, which is a general formulation of polynomial smooth support vector machines. The global convergence of dPSSVM is proved by a mathematical inductive method, and experiments are carried out to evaluate dPSSVM. The numerical results show that the performance and efficiency of dPSSVM are best, and better than that of the PSSVM when its smooth order is 3, but after its smooth order is greater than 3, the performance of classification is almost the same while the efficiency becomes worse. The problem of general formulation is successfully solved for polynomial smooth support vector machines.

Key words classification; support vector machine; Newton-Armijo algorithm; smoothing function; general formulation

摘 要 2005年袁玉波等人用一个多项式函数作为光滑函数,提出了一个多项式光滑的支持向量机模型PSSVM(polynomial smooth support vector machine),使分类性能及效率得到了一定提高. 2007年熊金志等人用插值函数的方法导出了一个递推公式,得到了一类新的光滑函数,解决了关于是否存在以及如何寻求性能更好的光滑函数的问题. 然而,支持向量机是否存在其他多项式光滑模型,以及多项式光滑模型的一般形式是什么等问题依然存在. 为此,将一类多项式函数作为新的光滑函数,使用光滑技术,提出了多项式光滑的支持向量机一般模型 dPSSVM (dth-order polynomial smooth support vector machine). 用数学归纳法证明了该一般模型的全局收敛性,并进行了数值实验. 实验结果表明,当光滑阶数等于3时,一般模型的分类性能及效率为最好,并优于PSSVM模型;当光滑阶数大于3后,分类性能基本不变,效率会有所降低. 成功解决了多项式光滑的支持向量机的一般形式问题.

关键词 分类; 支持向量机; Newton-Armijo 法; 光滑函数; 一般形式

中图法分类号 TP18

收稿日期:2007-11-19;**修回日期**:2008-03-25

基金项目:国家自然科学基金项目(60573029,60773050)

支持向量机 SVM(support vector machine)是近年来在统计学习理论的基础上发展起来的一种机器学习方法,是数据挖掘的一种新方法[1].它广泛地用于分类问题和回归问题,特别在文本分类、语音识别手写数字识别、数字水印等领域有大量应用[2-3].对分类问题的 SVM 算法,有很多学者开展了深入研究.1999 年,Platt 提出了 SMO 算法[4],Joachim 提出了 $\mathrm{SVM}^{\mathrm{light}}$ 算法[5],Mangasarian 等人提出了 SOR 算法[6].2001 年,Lee 和 Mangasarian 使用光滑技术,用 Sigmoid 函数的积分函数作为光滑函数,提出了光滑的支持向量机模型 SSVM[7],其实验结果表明,SSVM 的分类性能和计算效率比 SOR 算法、SMO 算法以及 $\mathrm{SVM}^{\mathrm{light}}$ 等算法要好.2004 年,Lu 等人对分类方法进行了比较研究,指出了光滑模型 SSVM 的优点和需继续做的工作[8].2005 年,袁玉波、严杰和徐成贤用多项式函数作为光滑函数,也使用光滑技术,提出了多项式光滑的支持向量机模型 PSSVM[9],其结果表明,PSSVM 的效果比 SSVM 好,从而开辟了研究支持向量机的一个新方向.2007 年,我们提出了一类逼近精度更好的多项式光滑函数[10].然而,多项式光滑的支持向量机作为一个新的研究方向,还存在如下 4 个尚未解决的问题:1)是否存在其他的多项式光滑模型;2)如果存在,有多少个,如何寻找它们;3)如果有若干个多项式光滑模型,是否存在一个一般形式,即是否可以用一个一般模型来表示所有的多项式光滑模型;4)如果存在这样一个一般模型,其收敛性如何,分类性能及效率又如何.

针对这些问题,本文应用我们在文献[10]中得到的一类多项式函数作为新的光滑函数,也使用光滑技术,提出多项式光滑的支持向量机一般模型 dPSSVM(dth-order polynomial smooth support vector machine),把多项式光滑的支持向量机模型推广到一般情形;还用数学归纳法证明一般模型的全局收敛性;最后用数值实验分析它的分类性能和效率.

1 多项式光滑的支持向量机一般模型的建立

1.1 分类问题[7,9]

给定$\mathbb{R}^n$中 m 个点 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_m$,用矩阵 $\boldsymbol{A}_{m\times n}$ 表示,目的是将 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_m$ 分为 A^+ 和 A^- 两类.若 A_i 属于类 A^+,记为 1;若属于 A^-,记为 -1,则可以用一个 $m\times m$ 的对角阵 $\boldsymbol{D}$ 来表示分类情形,其中 $\boldsymbol{D}$ 的对角元素为 1 或 -1.该分类问题的标准 SVM 模型为[7,9]:对某个 $v>0$,有

$$\min_{(\boldsymbol{\omega},\gamma,\boldsymbol{y})\in\mathbb{R}^{n+1+m}} v\boldsymbol{e}^{\mathrm{T}}\boldsymbol{y}+\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} \tag{1}$$
$$\text{s.t.}\quad \boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma)+\boldsymbol{y}\geqslant\boldsymbol{e},\ \boldsymbol{y}\geqslant\boldsymbol{0},$$

其中,$\boldsymbol{e}$ 表示分量全部为 1 的列向量.$\boldsymbol{\omega}$ 是如下的边界面的 n 维法线向量:

$$\begin{cases}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma=1,\\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma=-1,\end{cases} \tag{2}$$

而 γ 决定了边界面到原点的距离,$\boldsymbol{y}$ 为松弛变量,当类 A^+ 和 A^- 是严格线性可分时,$\boldsymbol{y}=\boldsymbol{0}$. $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma=1$ 和 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma=-1$ 分别是类 A^+ 和 A^- 的边界,此时分类的超平面为

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}=\gamma, \tag{3}$$

该平面处于式(2)的两平面之间,如图 1 所示.如果类 A^+ 和 A^- 是线性不可分的,则可以引入一个松弛变量 $\boldsymbol{y}(\boldsymbol{y}\geqslant\boldsymbol{0})$,使得:若 $\boldsymbol{x}^{\mathrm{T}}=\boldsymbol{A}_i$,$D_{ii}=1$,则 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma+y_i\geqslant1$;若 $\boldsymbol{x}^{\mathrm{T}}=\boldsymbol{A}_i$,$D_{ii}=-1$,则 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}-\gamma+y_i\leqslant-1$.

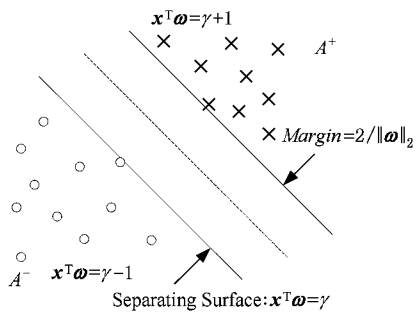

Fig. 1 Classification by SVM for the linearly separable case.

图 1 线性可分的 SVM 分类示意图

在标准的 SVM 模型式(1)中,目标函数对松弛变量 $\boldsymbol{y}$ 取 1-范数,并用 v 进行加权.如对目标函数进行修正,则可以得到如下的修正模型[6-7,9]:

$$\min_{(\boldsymbol{\omega},\gamma,\boldsymbol{y})\in\mathbb{R}^{n+1+m}} \frac{v}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2) \tag{4}$$
$$\text{s.t.}\ \boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma)+\boldsymbol{y}\geqslant\boldsymbol{e},\ \boldsymbol{y}\geqslant0,$$

令

$$\boldsymbol{y}=(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma))_+, \tag{5}$$

其中,$(\cdot)_+$ 为正号函数,其第 i 个分量定义为 $(x_+)_i=\max\{0,\ x_i\}$.将式(4)中的 $\boldsymbol{y}$ 用式(5)代替,就把有约束的支持向量机模型式(4)变成了无约束的支持向量机模型

$$\min_{\boldsymbol{\omega},\gamma}\frac{v}{2}\ \|\ (\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma))_+\ \|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2). \tag{6}$$

这种修正对原分类问题的影响很小,却能导出优化问题中目标函数的严格凸性,便于数学处理[6-7,9]. 优化问题式(6)的目标函数不具有光滑性,不能用快速的牛顿法求解.

1.2 一般模型的形式

为适用快速的牛顿法求解,我们引入光滑技术,用多项式光滑函数的一般形式[10]$\{p_d(x,k),\ d=1,2,3,\cdots\}$来逼近原支持向量机模型式(6)的正号函数$(\cdot)_+$,得到 d 阶多项式光滑的支持向量机模型 dPSSVM:

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_d(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2). \tag{7}$$

此模型即为多项式光滑的支持向量机一般模型. 该一般模型具有以下特点:

1) 目标函数具有光滑性,其光滑阶数与光滑函数 $p_d(x,k)$相同,即一般模型的光滑阶数与光滑函数 $p_d(x,k)$相同;

2) 任意给定一个 $d(d=1,2,3,\cdots)$,就有一个模型与之对应,显然该一般模型包含了无数个多项式光滑模型;

3) $\{p_d(x,k),\ d=1,2,3,\cdots\}$包含了全部的多项式光滑函数[10],因而一般模型式(7)也包含了全部多项式光滑的支持向量机模型.

下面我们以 $d=1,2,3,4,5$ 为例,列出一般模型的前 5 个模型:

1 阶多项式光滑模型 1PSSVM

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_1(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2); \tag{8}$$

2 阶多项式光滑模型 2PSSVM

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_2(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2); \tag{9}$$

3 阶多项式光滑模型 3PSSVM

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_3(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2); \tag{10}$$

4 阶多项式光滑模型 4PSSVM

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_4(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2); \tag{11}$$

5 阶多项式光滑模型 5PSSVM

$$\min_{(\boldsymbol{\omega},\gamma)\in\mathbf{R}^{n+1}} \frac{v}{2}\|p_5(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-\boldsymbol{e}\gamma),k)\|_2^2+\frac{1}{2}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma^2). \tag{12}$$

由式(8)和式(9)可知一般模型 dPSSVM 与文献[9]中模型 PSSVM 的关系:PSSVM 与以上两个模型 1PSSVM 和 2PSSVM 相同,它是一般模型 dPSSVM 的特殊情形,而一般模型 dPSSVM 把 PSSVM 推广到了一般情形.

2 一般模型的收敛性

我们先研究光滑函数 $p_d(x,k)$的一个重要性质:

定理 1. 当 x 和 k 一定时,正号函数 x_+ 的 d 阶多项式光滑函数 $p_d(x,k)$关于光滑阶数 d 是单调减函数:

1) $p_d(x,k)\leqslant p_{d-1}(x,k),\ d=2,3,4,\cdots$;

2) $0\leqslant p_d^2(x,k)-x_+^2\leqslant p_{d-1}^2(x,k)-x_+^2,\ d=2,3,4,\cdots$.

证明见附录 A.

证明一般模型的收敛性,就是要证明对任意正整数 d,当 k 趋于正无穷时,一般模型式(7)的解收敛于原支持向量机模型式(6)的解. 为便于证明,我们沿用文献[9]的一套符号:

设 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$, $\boldsymbol{b}\in\mathbf{R}^{m\times 1}$;实函数 $h(x):\mathbf{R}^n\to\mathbf{R}$; $g_d(x,k):\mathbf{R}^n\times\mathbf{N}\to\mathbf{R}$. 将原支持向量机模型式(6)的目标函数写成

$$h(x)=\frac{v}{2}\|(\boldsymbol{A}x-\boldsymbol{b})_+\|_2^2+\frac{1}{2}\|x\|_2^2, \tag{13}$$

将一般模型式(7)的目标函数写成

$$g_d(x,k)=\frac{v}{2}\|p_d(\boldsymbol{A}x-\boldsymbol{b},k)\|_2^2+\frac{1}{2}\|x\|_2^2. \tag{14}$$

显然,用不同的光滑函数 $p_d(x,k)$代入式(14),可得不同的目标函数 $g_d(x,k)$,从而可得不同的优化问题 $\min\limits_x g_d(x,k)$.

定理 2. $h(x)$和 $g_d(x,k)$分别定义如式(13)和式(14),则对任意正整数 d,当 k 趋于正无穷时,优化问题 $\min\limits_x g_d(x,k)$的解收敛于优化问题 $\min\limits_x h(x)$的解,即 $\lim\limits_{k\to+\infty}\bar{x}_{dk}=\bar{x}, d=1,2,3,4,\cdots$.

证明见附录 B. 注意到一般模型式(7)对应于 $g_d(x,k)$,原支持向量机模型式(6)对应于 $h(x)$,由定理 2 可知,对任意给定的光滑阶数 d,当 k 趋于正无穷时,一般模型式(7)的解都收敛于原支持向量机模型式(6)的解,即一般模型是全局收敛的.

为便于下一节研究一般模型的分类性能和效率,我们先研究它的前 1~5 阶光滑模型的收敛性及

其收敛上界.为此,引入下面的引理:

引理 1. 多项式光滑函数[10] $p_1(x,k)$,$p_2(x,k)$,$p_3(x,k)$,$p_4(x,k)$和 $p_5(x,k)$对任意的 x 和 k,有:

1) $p_1^2(x,k)-x_+^2\leqslant 0.0909/k^2$;

2) $p_2^2(x,k)-x_+^2\leqslant 0.0562/k^2$;

3) $p_3^2(x,k)-x_+^2\leqslant 0.03578/k^2$;

4) $p_4^2(x,k)-x_+^2\leqslant 0.02763/k^2$;

5) $p_5^2(x,k)-x_+^2\leqslant 0.02243/k^2$.

引理 1 的 1)和 2),文献[9]已有证明,引理 1 的 3)和 4),文献[10]也有证明.引理 1 的 5)可类似证明,证明过程予以省略.

当 $d=1,2,3,4,5$ 时,定理 2 中相应的优化问题 $\min\limits_x g_d(x,k)$ 可分别记为$\min\limits_x g_1(x,k)$,$\min\limits_x g_2(x,k)$,$\min\limits_x g_3(x,k)$,$\min\limits_x g_4(x,k)$ 和$\min\limits_x g_5(x,k)$,它们的解可分别记为 $\bar{x}_{1k}$,$\bar{x}_{2k}$,$\bar{x}_{3k}$,$\bar{x}_{4k}$ 和 $\bar{x}_{5k}$.于是,我们用定理 2 分别结合引理 1 的 1)2)3)4)5),可推得1～5 阶多项式光滑模型的收敛性及其收敛上界,即下面的推论:

推论 1. 1 阶多项式光滑模型 1PSSVM 收敛,且 $\|\bar{x}_{1k}-\bar{x}\|_2^2\leqslant 0.0909m/2k^2$;

推论 2. 2 阶多项式光滑模型 2PSSVM 收敛,且 $\|\bar{x}_{2k}-\bar{x}\|_2^2\leqslant 0.0526m/2k^2$;

推论 3. 3 阶多项式光滑模型 3PSSVM 收敛,且 $\|\bar{x}_{3k}-\bar{x}\|_2^2\leqslant 0.03578m/2k^2$;

推论 4. 4 阶多项式光滑模型 4PSSVM 收敛,且 $\|\bar{x}_{4k}-\bar{x}\|_2^2\leqslant 0.02763m/2k^2$;

推论 5. 5 阶多项式光滑模型 5PSSVM 收敛,且 $\|\bar{x}_{5k}-\bar{x}\|_2^2\leqslant 0.02243m/2k^2$.

其中 m 是训练样本的数目,k 为光滑因子.

该 5 个光滑模型的收敛性皆可由定理 2 直接得到.推论 1 的收敛上界,已在附录 B 中得到证明,推论 2～5 的收敛上界可类似证明,证明过程予以省略.

有了这些收敛上界,我们就可以讨论下一节实验时光滑因子 k 应如何选取的问题.由定理 2 知,当 k 趋于正无穷时,一般模型的解收敛于原支持向量机模型的解,但实验时 k 不可能也无必要取正无穷大,那么 k 最小要多大才能保证解的精度呢,这个最小的 k 被称为最优光滑因子,记为 k_{opt}.文献[9]已给出一般模型当 $d=1,2$ 时的最优光滑因子,下面的定理 3 给出当 $d=3,4,5$ 时的最优光滑因子.

定理 3. 对任意给定的 m 个训练样本,任意给定一般模型最优解 $\bar{x}_{dk}$ 与原分类模型最优解 $\bar{x}$ 的逼近精度 ε,则:

1) 当 $d=3$ 时,$k_{\mathrm{opt}}=\sqrt{0.03578m/2\varepsilon}$;

2) 当 $d=4$ 时,$k_{\mathrm{opt}}=\sqrt{0.02763m/2\varepsilon}$;

3) 当 $d=5$ 时,$k_{\mathrm{opt}}=\sqrt{0.02243m/2\varepsilon}$.

证明. 由引理 1 的 3)知,$\|\bar{x}_{3k}-\bar{x}\|_2^2\leqslant 0.03578m/2k^2$,若 $0.03578m/2k^2\leqslant\varepsilon$,即只要 $k\geqslant\sqrt{0.03578m/2\varepsilon}$,就有 $\|\bar{x}_{3k}-\bar{x}\|_2^2\leqslant\varepsilon$,就可保证求解精度.因此 $k_{\mathrm{opt}}=\sqrt{0.03578m/2\varepsilon}$.

类似地,由引理 1 的 4),可证明 2);由引理 1 的 5),可证明 3),证明过程予以省略. 证毕.

为保证求解精度,由定理 3 可计算光滑因子 k 的一个下界,即最优光滑因子.在求解一般模型 dPSSVM 时,只需令光滑因子取这个下界的值即可.由文献[9]知,当 $d=1$, 2 时,一般模型 dPSSVM 就是 PSSVM 模型,其最优光滑因子$k_{\mathrm{opt}}=\sqrt{0.0909m/2\varepsilon}$或 $k_{\mathrm{opt}}=\sqrt{0.0526m/2\varepsilon}$.一般地,任意给定一个光滑阶数 d,一般模型 dPSSVM 的最优光滑因子皆可仿定理 3 求出.

3 数值实验

3.1 实验设计

我们先讨论一般模型如何与有关的光滑模型进行比较的问题.

1) 文献[9]已指出,PSSVM 模型的分类性能和效率比 SSVM 模型好,因此我们只需将一般模型 dPSSVM 与 PSSVM 模型进行比较,不必与 SSVM 模型比较.

2) 为同文献[9]的 PSSVM 模型进行公平比较,我们在实验中使用相同的计算机配置.

3) 采用的算法也相同,模型 1PSSVM 用 BFGS-Armijo 法求解,模型 3PSSVM,4PSSVM 和 5PSSVM 用 Newton-Armijo 法求解[1].

4) 采用相同的实验数据,给定数据库规模为 $(m+m_1)n$,其中 m 为训练样本数,m_1 为测试样本数,n 为每个样本的属性个数.令 $\boldsymbol{w},\boldsymbol{x}\in\mathbb{R}^n$,且 $\boldsymbol{w}$, $\boldsymbol{x}$ 是列向量,$\boldsymbol{w}$ 的分量全部为 1,$\boldsymbol{x}\in[a,b]^n$,a, b 的值可以任意选取,如 $a=-100$, $b=100$.根据样本点的取值,可以分 3 种情形:1)若 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}\geqslant 1$,将 $\boldsymbol{x}$ 归为第 1 类;2)若 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}\leqslant -1$,将 $\boldsymbol{x}$ 归为第 2 类;3)若 $-1<\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}<1$,将该点以 p_1 的概率保留(如 $p_1=0.7$),然后,若 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}<0$,以 p_2 的概率(如 $p_2=0.8$)归为第 2 类,$1-p_2$ 的概率归为第 1 类,若 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}\geqslant 0$,以 p_2 的概率(如 $p_2=0.8$)归为第 1 类,$1-p_2$ 的概率归为

第 2 类. 若不保留该点,则重新产生新的数据点,直到产生 $m+m_1$ 个样本为止. 将前 m 个样本用来训练,后 m_1 个样本用来测试分类器性能.

因为一般模型含有无数个光滑模型,不可能也无必要对每一个光滑模型都进行实验,所以我们先考虑一般模型中前 5 个模型的实验问题,然后从中分析出一般模型的分类规律. 为此,我们先对 1 阶光滑模型 1PSSVM、2 阶光滑模型 2PSSVM、3 阶光滑模型 3PSSVM、4 阶光滑模型 4PSSVM 和 5 阶光滑模型 5PSSVM 进行实验,然后将它们进行比较. 值得注意的是,3PSSVM,4PSSVM 和 5PSSVM 在与 1PSSVM,2PSSVM 比较的同时,也就是与 PSSVM 模型比较.

3.2 实验结果与分析

我们按训练样本数的规模做了 6 个实验,数据规模分别为(100,10),(400,10),(1000,10),(200,10),(200,40),(200,100). 这 6 个实验分为两组:一组是样本属性个数不变而训练样本数变化时的实验;另一组是训练样本数不变而样本属性个数变化时的实验.

实验中的参数:一般模型的解与原分类模型的解的逼近精度 $\varepsilon=1.0\mathrm{E}-3$,BFGS-Armijo 法和 Newton-Armijo 法结束时下降方向的模为 $\varepsilon_1=1.0\mathrm{E}-3$,测试样本数 $m_1=1000$. 模型 1PSSVM, 2PSSVM, 3PSSVM,4PSSVM 和 5PSSVM 按定理 3 取各自的最优光滑因子.

实验结果见表 1 和表 2,其中第 1 列是训练样本数的规模,样本数为 m,样本属性为 n,其余各列是取不同模型的实验结果. 结果包括 4 个指标,测试样本分类正确率和 CPU 时间是国际通用的指标,引入训练样本正确率和目标函数最优值 opt 这两个指标,有助于分析实验结果.

Table 1 Experimental Results by Varying the Data Size

表 1 训练样本数变化时取不同模型的实验结果

Dataset (m,n)	1PSSVM	2PSSVM	3PSSVM	4PSSVM	5PSSVM
100,10	100	100	100	100	100
	94	94	95	95	95
	6.6	1.26	1.20	1.8	2.12
	0.0097	0.0097	0.0069	0.0079	0.0099
400,10	100	100	100	100	100
	99	99	99	99	99
	28.84	6.26	6.05	7.56	10.56
	0.1077	0.1077	0.0877	0.0678	0.2340
1000,10	99.70	99.70	99.70	99.70	99.70
	99	99	99	99	99
	140.77	29.22	28.01	32.13	47.25
	6.2841	6.2841	4.1328	8.3148	7.2587

Note: From top to bottom, the data in every cell refer to training correct rate(%), test correct rate(%), CPU time (s), opt.

Table 2 Experimental Results by Varying the Data Dimension

表 2 样本属性个数变化时取不同模型的实验结果

Dataset (m,n)	1PSSVM	2PSSVM	3PSSVM	4PSSVM	5PSSVM
200,10	100	100	100	100	100
	98.5	98.5	99	99	99
	11.10	3.41	1.98	4.52	6.15
	0.0458	0.0458	0.0235	0.0426	0.0142
200,40	100	100	100	100	100
	90.4	90.4	91.5	91.5	91.5
	72.39	10.44	9.23	18.65	23.16
	0.0045	0.0045	0.0045	0.0056	0.0078
200,100	100	100	100	100	100
	80.4	80.5	82.5	82.6	82.6
	233.43	14.5	12.6	21.8	32.5
	5.8131E−4	5.8216E−4	5.7123E−4	3.2246E−4	2.5562E−4

Note: From top to bottom, the data in every cell refer to training correct rate(%), test correct rate(%), CPU time (s), opt.

为了分析一般模型的分类效果,我们结合表 1 和表 2 将它同 PSSVM 模型从以下 3 方面进行比较.

1) 模型的有效性

模型是否有效首先要看目标函数最优值 opt 的大小,opt 取值越小,模型越有效. 从表 1 和表 2 可以看出,这 5 个模型的 opt 值都很小,说明它们对原分类问题都是有效的.

2) 计算效率

从表 1 和表 2 可以看出,2PSSVM 的效率比 1PSSVM 高,3PSSVM 的效率比 2PSSVM 高,也就是说,3PSSVM 的效率比文献[9]的 PSSVM 高;4PSSVM 的效率比 3PSSVM 低,5PSSVM 的效率比 4PSSVM 低,可见一般模型当光滑阶数大于 3 后,效率会降低,以光滑阶数等于 3 的分类效率为最好.

3) 分类正确率

从表 1 和表 2 还可以看出,当数据规模相同时,1PSSVM 和 2PSSVM 模型的训练样本正确率以及测试样本正确率基本相同;3PSSVM 模型的训练样本正确率以及测试样本正确率高于 1PSSVM 和 2PSSVM 模型,即高于 PSSVM 模型;4PSSVM 和 5PSSVM 模型与 3PSSVM 模型相比,训练样本正确率以及测试样本正确率基本相同. 可见一般模型当光滑阶数大于 3 后,分类性能基本不变.

以上实验结果表明,一般模型的分类性能和效率总体来说,当光滑阶数 d 等于 2 时比等于 1 时好;当光滑阶数 d 等于 3 时比等于 2 时好;当光滑阶数大于 3 后,一般模型的分类性能基本不变,但效率会降低. 其原因是:随着光滑阶数 d 的增大,一方面一般模型的收敛上界在缩小,使一般模型的解与原支持向量机模型的解更接近;另一方面,一般模型中所用的光滑函数 $p_d(x,k)$ 是 $2d$ 次多项式函数,计算一次 $p_d(\boldsymbol{e}-\boldsymbol{D}(\boldsymbol{A\omega}-\boldsymbol{e}\gamma),k)$ 的乘法次数约为 $(2dn)$ 次,计算量增大较快.

4 结　　论

本文用我们在文献[10]中得到的一类多项式函数作为新的光滑函数,使用光滑技术,提出了多项式光滑的支持向量机一般模型. 该一般模型包含有无穷多个多项式光滑的支持向量机模型,是一个包含 PSSVM 模型在内的多项式光滑的支持向量机的一般形式. 我们首先证明了该类多项式光滑函数关于光滑阶数是单调减函数,然后用数学归纳法证明了一般模型的全局收敛性,并做了数值实验. 结合引言中提到的 4 个尚未解决的问题,本文归纳起来可得出以下结论:

1) 多项式光滑的支持向量机除 PSSVM 模型外,还存在其他的多项式光滑模型.

2) 这种多项式光滑模型有无穷多个. 用一个多项式光滑函数对支持向量机作光滑处理,就可得到一个多项式光滑模型.

3) 这无穷多个光滑模型存在一个一般形式,即可以用一个一般模型来表示.

4) 一般模型是全局收敛的. 实验结果表明,一般模型的分类性能和效率当光滑阶数等于 2 时比等于 1 时好,当光滑阶数等于 3 时比等于 2 时好;当光滑阶数大于 3 后,分类性能基本不变,效率会降低. 一般模型的分类性能和效率总体上以光滑阶数等于 3 时为最好,且好于 PSSVM 模型.

于是,关于支持向量机是否存在和如何寻找其他的多项式光滑模型,以及求它们的一般形式等问题,就基本得到了解决.

致谢　在本文的研究过程中,Lee Yuh-Jye 博士和袁玉波博士提供了部分数据和程序;余龙才硕士参与了实验. 在此一并表示感谢!

参　考　文　献

[1] Deng Naiyang, Tian Yingjie. New Method in Data Mining: Support Vector Machine [M]. Beijing: Science Press, 2004 (in Chinese)

(邓乃扬,田英杰. 数据挖掘的新方法——支持向量机[M]. 北京:科学出版社,2004)

[2] Li Chunhua, Lin Hefei, Lu Zhengding. Adaptive spatial domain image watermarking based on support vector machine [J]. Journal of Computer Research and Development, 2007, 44(8): 1399-1405 (in Chinese)

(李春花,凌贺飞,卢正鼎. 基于支持向量机的自适应图像水印技术[J]. 计算机研究与发展,2007,44(8):1399-1405)

[3] Wang Jian, Lin Fuzong. Digital audio watermarking based on support vector machine (SVM) [J]. Journal of Computer Research and Development, 2005, 42(9): 1605-1611 (in Chinese)

(王剑,林福宗. 基于支持向量机(SVM)的数字音频水印[J]. 计算机研究与发展,2005,42(9):1605-1611)

[4] Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines [G] //Scholkopf B, *et al*. eds. Advances in Kernel Methods—Support Vector Learning. Cambridge, MA: MIT Press, 1999: 185-208

[5] Joachims T. Making large-scale support vector machine learning practical [G] //Scholkopf B, *et al*. eds. Advances in Kernel Methods—Support Vector Learning. Cambridge, MA: MIT Press, 1999: 169-184

[6] Mangasarian O L, Musicant D R. Successive overrelaxation for support vector machines [J]. IEEE Trans on Neural Networks, 1999, 10(8): 1032-1037

[7] Lee Y J, Mangasarian O L. SSVM: A smooth support vector machine for classification [J]. Computational Optimization and Applications, 2001, 22(1): 5-21

[8] Lu S, Wang X. A comparison among four SVM classification methods: LSVM, NLSVM, SSVM and NSVM [C] //Proc of the 3rd Int'l Conf on Machine Learning and Cybernetics. New York: IEEE, 2004: 4277-4282

[9] Yuan Yubo, Yan Jie, Xu Chengxian. Polynomial support vector machine [J]. Chinese Journal of Computers, 2005, 28(1): 9-17 (in Chinese)
(袁玉波, 严杰, 徐成贤. 多项式光滑的支撑向量机[J]. 计算机学报, 2005, 28(1): 9-17)

[10] Xiong Jinzhi, Hu Jinlian, Yuan Huaqiang, *et al*. A new class of smoothing functions for support vector machines [J]. Acta Electronica Sinica, 2007, 35(2): 366-370 (in Chinese)
(熊金志, 胡金莲, 袁华强, 等. 一类光滑支持向量机新函数的研究[J]. 电子学报, 2007, 35(2): 366-370)

Xiong Jinzhi, born in 1964. Associate professor in the Software College, Dongguan University of Technology. Received his M. S. degree from North-West Polytechnic University in 1988. He studied at Tsinghua University in 2005, and was a senior visiting scholar at the South China University of Technology in 2006. His main research interests include artificial intelligence and data mining.

熊金志,1964 年生,工学硕士,副教授,2005 年在清华大学学习,2006 年为华南理工大学高级访问学者,主要研究方向为人工智能、数据挖掘.

Yuan Huaqiang, born in 1966. Ph. D. Professor in the Software College, Dongguan University of Technology. His main research interests include artificial intelligence.

袁华强,1966 年生,博士,教授,主要研究方向为人工智能.

Peng Hong, born in 1956. Professor and Ph. D. supervisor in the College of Computer Science and Engineering, South China University of Technology. His main research interests include intelligent computing.

彭 宏,1956 年生,教授,博士生导师,主要研究方向为智能计算.

Research Background

Based on statistical learning theory, support vector machine (SVM) is now attracting increasing attention and has been widely used in data miming. It has been successfully applied for both classification and regression problems, such as handwritten digit recognition, speaker identification and face detection. The smooth support vector machine (SSVM) is a new reformulation of SVM and shows its advantages. However, as a new classification model, there are many theoretical problems. By designing better smooth functions, this paper proposes new models of SSVM, which have higher convergence speed and thus better classification performance and efficiency. This paper focuses on polynomial smooth support vector machines and solves an important problem in its general formulation. This work is supported by the National Natural Science Foundation under grant No. 60573029 and No. 60773050.

附录 A. 定理 1 的证明.

证明. 在区间$\left(-\frac{1}{k},\frac{1}{k}\right)$以外,由 d 阶光滑函数 $p_d(x,k)$的定义[A1],显然该定理成立;在区间$\left(-\frac{1}{k},\frac{1}{k}\right)$上,以下我们对定理进行证明:

1) 记 $I_i=\int\left(x^2-\frac{1}{k^2}\right)^i\mathrm{d}x$, $i=1,2,3,\cdots$, 将递推公式[A1]

$$I_{d-1}=\frac{x\left(x^2-\frac{1}{k^2}\right)^{d-1}}{2d-1}-\frac{2(d-1)}{(2d-1)k^2}I_{d-2}$$

两端乘以 a_d,然后积分,由公式[A1] $p_d(x,k)=a_d\int I_{d-1}\mathrm{d}x$ 和 $p_{d-1}(x,k)$ 的光滑性条件,可得

$$p_d(x,k)=-\frac{2(d-1)}{(2d-1)k^2}\frac{a_d}{a_{d-1}}p_{d-1}(x,k)+\frac{(-1)^d a_d}{2d(2d-1)}\left(\frac{1}{k^2}-x^2\right)^d. \tag{A1}$$

再用 $p_d\left(\frac{1}{k},k\right)=\frac{1}{k}$, $p_{d-1}\left(\frac{1}{k},k\right)=\frac{1}{k}$代入式(A1),得

$$a_d=-\frac{(2d-1)k^2}{2(d-1)}a_{d-1},\ d=2,3,4,\cdots, \tag{A2}$$

由式(A2)易得

$$a_d=(-1)^{d-1}\frac{3\cdot5\cdot7\cdots(2d-1)k^{(2d-2)}}{2\cdot4\cdot6\cdots(2d-2)}a_1, \tag{A3}$$

记

$$a'_d=\frac{3\cdot5\cdot7\cdots(2d-1)k^{(2d-2)}}{2\cdot4\cdot6\cdots(2d-2)}a_1,$$

式(A3)变成

$$a_d = (-1)^{d-1} a'_d, \tag{A4}$$

将式(A4)代入式(A1)，得

$$p_d(x,k) = p_{d-1}(x,k) - \frac{a'_d}{2d(2d-1)}\left(\frac{1}{k^2} - x^2\right)^d.$$

由文献[A1]知 $a_1 > 0$，则 $a'_d > 0$. 因 x 在区间 $\left(-\frac{1}{k}, \frac{1}{k}\right)$ 上，显然由上式知

$$p_d(x,k) < p_{d-1}(x,k),\ d = 2,3,4,\cdots.$$

2) 由光滑函数 $p_d(x,k)$ 的定义，显然 $x_+ < p_d(x,k)$. 结合上式，有

$$0 < p_d^2(x,k) - x_+^2 < p_{d-1}^2(x,k) - x_+^2,$$
$$d = 2,3,4,\cdots,$$

综上，在整个 x 轴上该定理成立. 证毕.

参 考 文 献

[A1] Xiong Jinzhi, Hu Jinlian, Yuan Huaqiang, *et al*. A new class of smoothing functions for support vector machines [J]. Acta Electronica Sinica, 2007, 35(2): 366-370 (in Chinese)
(熊金志，胡金莲，袁华强，等. 一类光滑支持向量机新函数的研究[J]. 电子学报，2007，35(2)：366-370)

附录 B. 定理 2 的证明.

证明. 由文献[B1]知，$h(x)$ 和 $g_d(x,k)$ 为严格凸函数，优化问题 $\min\limits_x h(x)$ 和 $\min\limits_x g_d(x,k)$ 的解是存在且唯一的(当 d 一定时). 下面我们用数学归纳法证明 $\lim\limits_{k\to+\infty} \bar{x}_{dk} = \bar{x}$，$d = 1, 2, 3, \cdots$.

1) $d=1$ 的情形，文献[B1]已有证明，为本文的完整性，我们重述如下：

由式(14)知

$$g_1(x,k) = \frac{v}{2}\| p_1(\boldsymbol{A}x - \boldsymbol{b},k)\|_2^2 + \frac{1}{2}\|x\|_2^2, \tag{B1}$$

由一阶最优性条件 $h'(\bar{x})=0$ 和 $h(x)$ 的严格凸性，可得

$$h(\bar{x}_{1k}) - h(\bar{x}) \geqslant \frac{v}{2}\|\bar{x}_{1k} - \bar{x}\|_2^2. \tag{B2}$$

同理，由一阶最优性条件和 $g_1(x,k)$ 的严格凸性，可得

$$g_1(\bar{x},k) - g_1(\bar{x}_{1k},k) \geqslant \frac{v}{2}\|\bar{x} - \bar{x}_{1k}\|_2^2. \tag{B3}$$

式(B2)加式(B3)，得

$$v\|\bar{x}_{1k} - \bar{x}\|_2^2 \leqslant (g_1(\bar{x},k) - h(\bar{x})) - (g_1(\bar{x}_{1k},k) - h(\bar{x}_{1k})), \tag{B4}$$

式(B1)减式(13)，并用 $x=\bar{x}_{1k}$ 代入，得

$$g_1(\bar{x}_{1k},k) - h(\bar{x}_{1k}) = \frac{v}{2}(\| p_1(\boldsymbol{A}\bar{x}_{1k} - \boldsymbol{b},k)\|_2^2 - \|(\boldsymbol{A}\bar{x}_{1k} - \boldsymbol{b})_+\|_2^2).$$

文献[B1]已证明 $p_1(x,k) \geqslant x_+$，代入上式可得 $g_1(\bar{x}_{1k},k) - h(\bar{x}_{1k}) \geqslant 0$. 由此式结合式(B4)，可得

$$\|\bar{x}_{1k} - \bar{x}\|_2^2 \leqslant \frac{1}{2}\sum_{i=1}^m [p_1^2(\boldsymbol{A}_i\bar{x} - b_i) - (\boldsymbol{A}_i\bar{x} - b_i)_+^2],$$

其中 $\boldsymbol{A}_i$ 是 $\boldsymbol{A}$ 的第 i 个行向量，b_i 是 $\boldsymbol{b}$ 的第 i 个分量.

由引理 1，有 $p_1^2(\boldsymbol{A}_i\bar{x} - b_i) - (\boldsymbol{A}_i\bar{x} - b_i)_+^2 \leqslant 0.0909/k^2$，代入上式可得

$$\|\bar{x}_{1k} - \bar{x}\|_2^2 \leqslant 0.0909m/2k^2, \tag{B5}$$

其中 m 为训练样本数目，k 为最优光滑因子. 显然 $\lim\limits_{k\to+\infty}\bar{x}_{1k} = \bar{x}$.

2) 当 $d=j-1$ 时，定理中的光滑函数 $p_d(x,k)$、优化问题 $\min\limits_x g_d(x,k)$ 及其解 $\bar{x}_{dk}$ 就分别表示为 $p_{j-1}(x,k)$，$\min\limits_x g_{j-1}(x,k)$ 和 $\bar{x}_{(j-1)k}$. 假设此时一般模型是收敛的，即

$$\|\bar{x}_{(j-1)k} - \bar{x}\|_2^2 \leqslant 0.0909m/2k^2.$$

3) 当 $d=j$ 时，由式(14)知

$$g_j(x,k) = \frac{v}{2}\| p_j(\boldsymbol{A}x - \boldsymbol{b},k)\|_2^2 + \frac{1}{2}\|x\|_2^2, \tag{B6}$$

类似于式(B2)，得

$$h(\bar{x}_{jk}) - h(\bar{x}) \geqslant \frac{v}{2}\|\bar{x}_{jk} - \bar{x}\|_2^2. \tag{B7}$$

注意到当 $d=1$ 的情形：$g_1(x,k)$ 在最优解 $\bar{x}_{1k}$ 处得到了式(B3)，同理，$g_j(x,k)$ 在最优解 $\bar{x}_{jk}$ 处可得

$$g_j(\bar{x},k) - g_j(\bar{x}_{jk},k) \geqslant \frac{v}{2}\|\bar{x} - \bar{x}_{jk}\|_2^2. \tag{B8}$$

式(B7)加式(B8)，得

$$v\|\bar{x}_{jk} - \bar{x}\|_2^2 \leqslant (g_j(\bar{x},k) - h(\bar{x})) - (g_j(\bar{x}_{jk},k) - h(\bar{x}_{jk})). \tag{B9}$$

式(B6)减式(13)，并用 $x=\bar{x}_{jk}$ 代入，得

$$g_j(\bar{x}_{jk},k) - h(\bar{x}_{jk}) = \frac{v}{2}(\| p_j(A\bar{x}_{jk} - \boldsymbol{b},k)\|_2^2 - \|(A\bar{x}_{jk} - \boldsymbol{b})_+\|_2^2).$$

由光滑函数 $p_j(x,k)$ 的定义[B2]，显然有 $p_j(x,k) \geqslant x_+$，所以上式有

$$g_j(\bar{x}_{jk},k) - h(\bar{x}_{jk}) \geqslant 0. \tag{B10}$$

结合式(B10)，由式(B9)可得

$$\|\bar{x}_{jk} - \bar{x}\|_2^2 \leqslant \frac{1}{2}(\| p_j(\boldsymbol{A}\bar{x} - \boldsymbol{b},k)\|_2^2 - \|(\boldsymbol{A}\bar{x} - \boldsymbol{b})_+\|_2^2). \tag{B11}$$

由定理 1，显然有

$$\| p_j(\boldsymbol{A}\bar{x} - \boldsymbol{b},k)\|_2^2 - \|(\boldsymbol{A}\bar{x} - \boldsymbol{b})_+\|_2^2 \leqslant \| p_{j-1}(\boldsymbol{A}\bar{x} - \boldsymbol{b},k)\|_2^2 - \|(\boldsymbol{A}\bar{x} - \boldsymbol{b})_+\|_2^2,$$

因此由式(B11)可得

$$\|\bar{x}_{jk} - \bar{x}\|_2^2 \leqslant \frac{1}{2}(\| p_{j-1}(\boldsymbol{A}\bar{x} - \boldsymbol{b},k)\|_2^2 - \|(\boldsymbol{A}\bar{x} - \boldsymbol{b})_+\|_2^2).$$

结合 $d=j-1$ 时的假设，可得 $\|\bar{x}_{jk} - \bar{x}\|_2^2 \leqslant 0.0909m/2k^2$. 可见当 $d=j$ 时，一般模型也是收敛的.

所以由数学归纳法知，一般模型对任意光滑阶数 d 都是收敛的. 证毕.

参 考 文 献

[B1] Yuan Yubo, Yan Jie, Xu Chengxian. Polynomial support vector machine [J]. Chinese Journal of Computers, 2005, 28(1): 9-17 (in Chinese)
(袁玉波，严杰，徐成贤. 多项式光滑的支撑向量机[J]. 计算机学报，2005，28(1)：9-17)

[B2] Xiong Jinzhi, Hu Jinlian, Yuan Huaqiang, *et al*. A new class of smoothing functions for support vector machines [J]. Acta Electronica Sinica, 2007, 35(2): 366-370 (in Chinese)
(熊金志，胡金莲，袁华强，等. 一类光滑支持向量机新函数的研究[J]. 电子学报，2007，35(2)：366-370)