

一个新的多项式光滑支持向量机

袁华强¹ 涂文根^{1,2} 熊金志^{1,2} 刘婷婷^{1,2}

(东莞理工学院工程技术研究院 东莞 523808)¹ (华南理工大学计算机科学与工程学院 广州 510006)²

摘要 光滑支持向量机(SSVM)是支持向量机(SVM)的快速求解模型,拥有更快的求解速度和训练效果。基于光滑的分段多项式函数和插值思想推导出一个新的光滑函数,从而可以更好地逼近正号函数。通过所得到的新光滑函数改进多项式光滑支持向量机模型(PSSVM),得到了更新的光滑支持向量机模型。还给出了新光滑函数的逼近性能和精度分析以及新模型的收敛性证明和最优解的逼近上限。数值实验表明,所提出的新光滑支持向量机模型性能优于 PSSVM 模型。

关键词 模式识别,多项式光滑支持向量机,样条多项式,光滑技术

New Polynomial Smooth Support Vector Machine

YUAN Hua-qiang¹ TU Wen-gen^{1,2} XIONG Jin-zhi^{1,2} LIU Ting-ting^{1,2}

(Engineering & Technology Institute, Dongguan University of Technology, Dongguan 523808, China)¹

(School of Computer Science & Engineering, South China University of Technology, Guangzhou 510006, China)²

Abstract Smooth Support Vector Machine(SSVM), which has better advantage than Support Vector Machine(SVM), is a model of SVM for quick solving. We got a new smooth function for approximating the plus function by interpolation base on smooth piecewise polynomial functions. We found a new model of Smooth Support Vector Machine by improving the model of Polynomial Smooth Support Vector Machine using the new smooth function. Performances of approaching and approximate error were given for the new smooth function, as well as the study of convergence and the approximation limit of optimum for the new model. Obviously numeric experiment shows that the new model has better performance than PSSVM.

Keywords Pattern recognition, Polynomial smooth support vector machine, Spline polynomial, Smooth technology

1 引言

光滑支持向量机(Smooth Support Vector Machine, SSVM)是支持向量机(Support Vector Machine, SVM)的快速求解模型,由 Lee 等人于 2001 年提出^[1],使用式(1)定义的 Sigmoid 函数的积分函数对正号函数 $(x)_+$ 作光滑处理。

$$p(x, k) = x + \frac{1}{k} \log(1 + e^{-kx}), k > 0 \quad (1)$$

式中, e 是自然对数的底, k 是光滑因子。获得了 SSVM 模型:

$$\min_{(\omega, \gamma) \in R^{n+1}} \frac{\nu}{2} \|p(e - D(A\omega - e\gamma), k)\|_2^2 + \frac{1}{2}(\omega' \omega + \gamma^2) \quad (2)$$

式中, ν 是平衡经验风险和置信区间的因子^[6], 矩阵 $A_{m \times n}$ 表示 m 个 n 维样本的数据集, $D_{m \times m}$ 相应地为 m 个样本所属类型的对角矩阵^[1], e 为任意维的单位向量, (ω, γ) 为所求的分类面。还给出数值实验,证明了使用普通无约束优化算法的 SSVM 比 SOR 算法^[8]、SMO 算法^[9]以及 SVM^{light} 算法^[10]有更好的求解效果。

2005 年,袁玉波等^[2]在 SSVM 的基础上研究了多项式光滑支持向量机(Polynomial Smooth Support Vector Machine, PSSVM),给出了一阶光滑多项式函数以及二阶光滑多项式

函数。一阶光滑多项式函数如式(3)定义:

$$f(x, k) = \begin{cases} x, & x \geq 1/k \\ \frac{1}{4}kx^2 + \frac{1}{2}x + \frac{1}{4k}, & -1/k < x < 1/k \\ 0, & x \leq -1/k \end{cases} \quad (3)$$

式中, $k > 0$ 是光滑因子。获得了 PSSVM 模型为:

$$\min_{(\omega, \gamma) \in R^{n+1}} \frac{\nu}{2} \|f(e - D(A\omega - e\gamma), k)\|_2^2 + \frac{1}{2}(\omega' \omega + \gamma^2) \quad (4)$$

证明了 PSSVM 模型的收敛性;给出了多项式光滑函数的逼近精度,与 Sigmoid 函数的积分函数做了相关比较;用 BFGS^[7]和 Newton-Armijo 算法^[11]分别训练 SSVM 和 PSSVM,得出了 PSSVM 有更好性能的结论。

基于 PSSVM,熊金志等^[3]研究了 PSSVM 的一般模型;用插值方法导出了一个重要的递推公式(5)^[4]:

$$I_{d-1} = \frac{x(x^2 - \frac{1}{k^2})^{d-1}}{2d-1} - \frac{2(d-1)}{(2d-1)k^2} I_{d-2}, d \geq 2$$
$$f_d(x, k) = a_d \int I_{d-1} dx \quad (5)$$

根据式(5)及光滑性能能够得到一类形式(3)的多项式光滑函数;对所获得的多项式光滑函数进行了逼近精度分析和

到稿日期:2010-04-07 返修日期:2010-06-28 本文受国家自然科学基金(60573029,60773048,60773050)资助。

袁华强(1966-),男,博士,教授,CCF 高级会员,主要研究方向为数据挖掘、模式识别。

比较。

以上提到的 SSVM^[1] 和 PSSVM^[2] 分别基于 sigmoid 函数的积分函数和多项式光滑函数,但是都还不能很好地逼近正号函数,还有很大的改进空间^[14-16]。本文用间接的方法获得了一个新的光滑函数:先用插值方法逼近正号函数 $(x)_+$,获得一个多项式函数,再对其进行改进,得到了一个新的可应用于 SSVM 的光滑函数。本文证明了新函数的光滑性;给出了逼近性能分析。通过所得的新光滑函数改进 PSSVM,得到了新模型,证明了新模型的收敛性;收敛上限表明其相比 PSSVM,新模型拥有更好的收敛性,同时数值实验也表明其有更佳的求解效果。

为方便理解给出符号说明,向量没有经转置 T 作用的为列向量,向量 x 和 y 的内积为 $x^T y$, $\|x\|_p$ 为 x 的 p -范数,正号函数 $(x)_+ = \max(0, x)$ 。 $k!$ 是 k 的阶乘, $\{x_i\}_{i=1}^n$ 为点集 $\{x_1, x_2, \dots, x_n\}$ 。 $f \in C^n (n \geq 0)$ 代表函数 f 有 n 阶光滑性, ∇f 为 f 的一阶导函数, $\nabla^m f (m \geq 1)$ 是函数 f 的 m 阶导函数。给定 δ , f 的水平集定义为 $L_\delta(f) = \{x | f(x) \leq \delta\}$ 。

2 样条 PSSVM 模型

2.1 样条多项式

样条函数是函数逼近中一个十分活跃的分支^[12],适用范围广,类型多。一般都假设插值点把区间分成 $\Pi: x_0 < x_1 < \dots < x_{n-1} < x_n$, 去掉两边界点 x_0 和 x_n , 在点集 $\{x_j\}_{j=1}^n$ 上构造 m 次样条基本元函数集合:

$$\left\{ \frac{(x-x_j)_+^m}{m!} \right\}_{j=1}^{n-1} \quad (6)$$

式中,操作符 $(\cdot)_+^m$ 的定义如下:

$$(u)_+^m = \begin{cases} u^m, & u > 0 \\ 0, & u \leq 0 \end{cases} \quad (7)$$

从而得出 k 次样条多项式函数的基本形式^[13]:

$$S_k(x) = a_0 + a_1 x + \dots + a_k x^k + \sum_{j=1}^{n-1} C_j \frac{(x-x_j)_+^k}{k!} \quad (8)$$

结合式(7)可以将式(8)化为分段的形式:

$$S_k(x) = \begin{cases} \sum_{l=0}^k a_l x^l, & x \in (-\infty, x_1] \\ \sum_{l=0}^k a_l x^l + \sum_{j=1}^i \frac{C_j (x-x_j)_+^k}{k!}, & x \in (x_i, x_{i+1}] \\ \sum_{l=0}^k a_l x^l + \sum_{j=1}^{n-1} \frac{C_j (x-x_j)_+^k}{k!}, & x \in (x_{n-1}, +\infty) \end{cases} \quad (9)$$

式中, $1 \leq i \leq n-2$, 样条多项式函数式(9)有如下性质。

性质 1 $k-1$ 次光滑, 即 $S_k(x) \in C^{k-1}$ 。

证明:

(1) 显然, 根据插值点的函数值可以判定 $S_k(x)$ 在 $(-\infty, +\infty)$ 上是连续的。

(2) 对其进行求导, 可以获得在 $n-1$ 个插值点处的 $m (1 \leq m \leq k-1)$ 阶导数都存在的结论。

综上所述, 可知 $S_k(x)$ 在节点 $\{x_j\}_{j=1}^n$ 上都是 $k-1$ 次光滑的, 即 $S_k(x) \in C^{k-1}$ 。

2.2 逼近 $(x)_+$ 的三阶样条多项式函数

对于本文 2.1 节提到的 k 次样条函数, 根据性质 1 可知当 $k=3$ 时就可以得到二阶光滑的样条多项式函数:

$$S_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \sum_{i=1}^{n-1} C_i (x-x_i)_+^3 \quad (10)$$

展开后为:

$$S_k(x) = \begin{cases} \sum_{l=0}^3 a_l x^l, & x \in (-\infty, x_1] \\ \sum_{l=0}^3 a_l x^l + \sum_{j=1}^i \frac{C_j (x-x_j)_+^3}{3!}, & x \in (x_i, x_{i+1}] \\ \sum_{l=0}^3 a_l x^l + \sum_{j=1}^{n-1} \frac{C_j (x-x_j)_+^3}{3!}, & x \in (x_{n-1}, +\infty) \end{cases} \quad (11)$$

式中, $1 \leq i \leq n-2$, 要确定式(10)的 $4 + (n-1) = n+3$ 个系数, 需要相同个数的插值条件。原点两侧的两对称点 $(x_0 = -\frac{1}{k}, y_0 = 0)$, 和 $(x_2 = \frac{1}{k}, y_2 = \frac{1}{k})$, 其中 $k > 0$, 加上原点 $(x_1 = 0, y_1 = \max(0, x) = 0)$, 以 3 个点来插值得到 $S_3(x, k)$ 及一阶光滑条件, 即 $f'(x_0) = S_3'(x_0)$ 和 $f'(x_n) = S_3'(x_n)$, 于是式(10)可以改写为:

$$\begin{aligned} S_3(x, k) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \sum_{i=1}^{2-1} C_i (x-x_i)_+^3 \\ &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + C_1 (x-x_1)_+^3 \\ &= \begin{cases} a_0 + a_1 x + a_2 x^2 + a_3 x^3, & x \leq x_1 \\ a_0 + a_1 x + a_2 x^2 + a_3 x^3 + C_1 x^3, & x > x_1 \end{cases} \end{aligned} \quad (12)$$

通过计算可以得到一阶光滑的样条函数为:

$$S_3(x, k) = \begin{cases} 0, & x \leq -\frac{1}{k} \\ \frac{1}{2} k^2 x^3 + kx^2 + \frac{1}{2} x, & -\frac{1}{k} < x \leq 0 \\ -\frac{1}{2} k^2 x^3 + kx^2 + \frac{1}{2} x, & 0 < x < \frac{1}{k} \\ x, & x \geq \frac{1}{k} \end{cases} \quad (13)$$

式中, $k > 0$ 。

正号函数的逼近效果如图 1 所示。

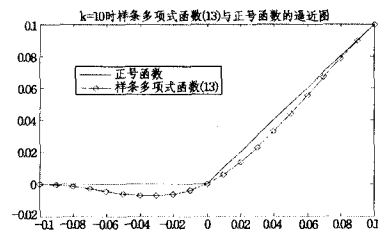


图 1 逼近正号函数的插值函数曲线

性质 2 $S_3(x, k)$ 为式(13)定义的 3 阶样条多项式函数, $(x)_+$ 为正号函数, 则有:

- (1) $S_3(x, k)$ 关于 x 一阶光滑;
- (2) $S_3(x, k) \leq (x)_+$;
- (3) 有 $(x)_+ - S_3(x, k) \leq \frac{2}{27k}$, 在 $x = \pm \frac{1}{3k}$ 相等;
- (4) $\nabla S_3(-\frac{1}{3k}, k) = 0, \nabla S_3(\frac{1}{3k}, k) = 1$ 。

证明:

(1) 可直接验证在 $x=0$ 和 $x = \pm \frac{1}{k}$ 处满足条件:

$$\begin{cases} S_3(-\frac{1}{k}, k) = 0, \nabla S_3(-\frac{1}{k}, k) = 0 \\ S_3(\frac{1}{k}, k) = \frac{1}{k}, \nabla S_3(\frac{1}{k}, k) = 1 \end{cases}$$

同时根据性质 1 也可以直接得到结论。

(2) $S_3(x, k) - (x)_+$ 的分析如下:

$$S_3(x, k) - (x)_+ =$$

$$\left\{ \begin{array}{ll} 0, & x \leq -\frac{1}{k}, x \geq \frac{1}{k} \\ \frac{1}{2}x(kx+1)^2, & -\frac{1}{k} < x \leq 0 \\ -\frac{1}{2}x(kx-1)^2, & 0 < x < \frac{1}{k} \end{array} \right\} \leq 0$$

式中, $k > 0$ 。

(3) 样条多项式(13)在区间 $(-\frac{1}{k}, 0]$ 满足条件 $S_3(x, k) \leq (x_+)_+$ 且先减后增, 在 $x = -\frac{1}{3k}$ 处取得最小, 满足 $(x_+)_+ - S_3(x, k) \leq S_3(x, -\frac{1}{3k}) = \frac{2}{27k}$ 。

在区间 $(0, \frac{1}{k})$ 有 $0 \leq S_3(x, k) \leq (x_+)_+$ 。我们令 $y(x) = (x_+)_+ - S_3(x, k) = \frac{1}{2}k^2x^3 - kx^2 + \frac{1}{2}x$, 对 $y(x)$ 求导 $\nabla y = \frac{3}{2}k^2x^2 - 2kx + \frac{1}{2} = \frac{1}{2}(kx-1)(3k-1)$, 最大值在 $x = \frac{1}{3k}$ 处, $y(\frac{1}{3k}) = (\frac{1}{3k})_+ - S_3(\frac{1}{3k}, k) = \frac{2}{27k}$ 。

(4) 可以直接根据定义求解得证。

2.3 应用于 SSVM 的样条多项式函数

根据性质 2 将 $S_3(x, k)$ 整体上移 $\frac{2}{27k}$, 在区间 $(-\infty, -\frac{1}{3k}]$ 和 $[\frac{1}{3k}, +\infty)$ 依然取正号函数, 可得到一个逼近 $(x_+)_+$ 的新光滑样条多项式函数式(14), 其中 $k > 0$ 。

$$f_3(x, k) = \begin{cases} 0, & x \leq -\frac{1}{3k} \\ \frac{1}{2}k^2x^3 + kx^2 + \frac{1}{2}x + \frac{2}{27k}, & -\frac{1}{3k} < x \leq 0 \\ -\frac{1}{2}k^2x^3 + kx^2 + \frac{1}{2}x + \frac{2}{27k}, & 0 < x < \frac{1}{3k} \\ x, & x \geq \frac{1}{3k} \end{cases} \quad (14)$$

定理 1 $f_3(x, k)$ 为式(14)定义的 3 阶样条多项式函数, $(x_+)_+$ 为正号函数, 则有:

- (1) $f_3(x, k)$ 关于 x 一阶光滑;
- (2) $f_3(x, k) \geq (x_+)_+ \geq 0$;
- (3) 任意 x 和 k 有 $f_3(x, k)^2 - (x_+)_+^2 \leq \frac{1}{122k^2}$ 。

证明:

(1) 可以根据性质 1 直接得到, 也可以直接根据定义验证。

(2) 根据性质 2 有 $(x_+)_+ - S_3(x, k) \leq \frac{2}{27k}$, 而 $f_3(x, k)$ 是 $S_3(x, k)$ 整体上移 $\frac{2}{27k}$, 因此有:

$$(x_+)_+ - f_3(x, k) = ((x_+)_+ - S_3(x, k)) - \frac{2}{27k} \leq 0$$

所以 $f_3(x, k) \geq (x_+)_+$ 。

(3) 根据性质 1, $f_3(x, k)$ 在区间 $(-\frac{1}{3k}, 0]$ 满足: $f_3(x, k)^2 - (x_+)_+^2 \leq f_3(0, k)^2 = (\frac{2}{27k})^2 < \frac{1}{122k^2}$ 。当 $(0, \frac{1}{3k})$ 有 $S_3(x, k)^2 - (x_+)_+^2 = (-\frac{1}{2}k^2x^3 + kx^2 + \frac{1}{2}x + \frac{2}{27k})^2 - x^2$, 令 $a = kx$, 有 $y(a) = \frac{1}{k^2}(-\frac{1}{2}a^3 + a^2 + \frac{1}{2}a + \frac{2}{27})^2 - \frac{a^2}{k^2}$, 可以求得 $a = 0.0751$ 时 y 取得最大值 $\frac{0.0081}{k^2}$, 因此有 $f_3(x, k)^2 - (x_+)_+^2 \leq y$

$$(0.0751) = \frac{0.0081}{k^2} < \frac{1}{122k^2} = \frac{0.0082}{k^2}。$$

2.4 新光滑支持向量机模型

光滑支持向量机的核心内容是使用光滑函数逼近正号函数 $(x)_+ = \max(0, x)$, 从而将模型式(15)中的非光滑部分光滑化。

$$\min_{(\omega, \gamma) \in R^{m+1}} \frac{\nu}{2} \| (e - D(A\omega - e\gamma))_+ \|_2^2 + \frac{1}{2} (\omega' \omega + \gamma^2) \quad (15)$$

式(15)为无约束的强凸性最优化问题。但是, 由于目标函数中存在正号函数, 是非光滑的, 无法使用基于梯度的快速算法来求解, 为此, 文献[1]提出使用光滑函数来逼近目标函数中的正号函数部分。我们使用本文 2.3 节得到的样条多项式函数式(14)代替正号函数, 可以得到新的光滑支持向量机模型:

$$\min_{(\omega, \gamma) \in R^{m+1}} \frac{\nu}{2} \| f_3(e - D(A\omega - e\gamma), k) \|_2^2 + \frac{1}{2} (\omega' \omega + \gamma^2) \quad (16)$$

3 性能分析

3.1 逼近效果分析

样条多项式函数(14)对正号函数的逼近效果可以分析其在区间 $(-1/k, 1/k)$ 的逼近程度。首先可以得出以上提到的几个光滑函数(1)、(3)和(14)在此区间对正号函数 $(x_+)_+$ 的逼近效果比较图。图 2 所示是在光滑因子 $k=30$ 时以上所提到的光滑函数的比较。

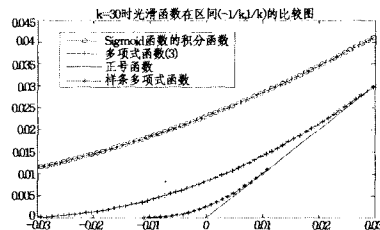


图 2 光滑函数(1)、(3)和(14)的逼近性能比较

引理 1^[2] $f(x, k)$ 是式(3)表示的函数, $(x_+)_+$ 为正号函数, 则有:

- (1) $f(x, k)$ 关于 x 一阶光滑;
- (2) $f(x, k) \geq (x_+)_+$;
- (3) 任意 x 和 k 有 $f(x, k)^2 - (x_+)_+^2 \leq \frac{1}{11k^2} = \frac{0.0909}{k^2}$ 。

对于 Sigmoid 函数的积分函数(1), 文献[2]证明了其明显比式(3)差。这里主要比较多项式函数(3)和样条多项式函数(14)对正号函数的逼近效果。根据定理 1 和引理 1 可以知道两者都是一阶光滑的, 明显样条多项式函数(14)的逼近精度上限 $\frac{0.0082}{k^2}$ 比多项式函数(3)的逼近精度上限 $\frac{0.0909}{k^2}$ 要小一个数量级。另外, 从图 2 也可以明显看出哪个逼近效果要更好。因此可以预测模型(16)的求解效果将优于模型(4)。

3.2 模型的收敛性

可以证明模型(15)的最优解能够逼近原来模型(14)的最优解。定理 2 给出了相关证明。

定理 2 假设 $A \in R^{m \times n}$, $b \in R^{m \times 1}$, 实函数 $h(x): R^n \rightarrow R, g(x, k): R^n \times N \rightarrow R$ 定义如下:

$$h(x) = \frac{1}{2} \| (Ax - b)_+ \|_2^2 + \frac{1}{2} \| x \|_2^2$$

$$g(x, k) = \frac{1}{2} \| f_3(Ax - b, k) \|_2^2 + \frac{1}{2} \| x \|_2^2$$

式中, $S_3(x, k)$ 为函数(13), 则有如下结论:

- (1) $h(x)$ 和 $g(x, k)$ 都是强凸函数;
- (2) 优化问题 $\min_x h(x)$ 和 $\min_x g(x, k)$ 都存在唯一解, 分别记为 x^* 和 x_k^* ;
- (3) 对任意的 k 都有 $\|x_k^* - x^*\| \leq \frac{m}{244k^2}$ 成立。

证明:

(1) 由 $\|\cdot\|_2$ 的强凸性很明显可以得到 $h(x)$ 和 $g(x, k)$ 都是强凸函数的结论;

(2) 证明 $\min_x h(x)$ 和 $\min_x g(x, k)$ 解的存在。根据定理 1 知 $S_3(x, k) \geq (x)_+$, 对任意 v 水平集 $L_v(h(x))$ 和 $L_v(g(x, k))$ 有 $L_v(g(x, k)) \subseteq L_v(h(x)) \subseteq \{x \mid \frac{1}{2} \|x\|_2^2 \leq v\}$, 因此水平集 $L_v(h(x))$ 和 $L_v(g(x, k))$ 都是 R^n 中的紧子集, 所以最优化问题 $\min_x h(x)$ 和 $\min_x g(x, k)$ 的解都存在, 加上(1)的强凸性, 解的唯一性结论得证。

(3) 由于一阶最优条件和(1)的强凸性, 有:

$$\begin{aligned} h(x_k^*) - h(x^*) &\geq \nabla h(x^*)(x_k^* - x^*) + \frac{1}{2} \|x_k^* - x^*\|_2^2 \\ &= \frac{1}{2} \|x_k^* - x^*\|_2^2 \\ g(x^*, k) - g(x_k^*, k) &\geq \nabla g(x_k^*, k)(x^* - x_k^*) + \frac{1}{2} \|x_k^* - x^*\|_2^2 \end{aligned}$$

将两式相加, 并应用定理 1, 有:

$$\begin{aligned} \|x_k^* - x^*\|_2^2 &\leq g(x^*, k) - h(x^*) - (g(x_k^*, k) - h(x_k^*)) \\ &= g(x^*, k) - h(x^*) \\ &= \frac{1}{2} \|f_3(Ax - b, k)\|_2^2 - \frac{1}{2} \|(Ax - b)_+\|_2^2 \\ &\leq \frac{m}{244k^2} \end{aligned}$$

3.3 模型最优解分析

引理 2^[2] 假设 $A \in R^{m \times n}$, $b \in R^{m \times 1}$, 实函数 $h(x): R^n \rightarrow R$, $g(x, k): R^n \times N \rightarrow R$ 定义如下:

$$h(x) = \frac{1}{2} \|(Ax - b)_+\|_2^2 + \frac{1}{2} \|x\|_2^2$$

$$g(x, k) = \frac{1}{2} \|f(Ax - b, k)\|_2^2 + \frac{1}{2} \|x\|_2^2$$

式中, $f(x, k)$ 为函数(3), 则有如下结论:

- (1) $h(x)$ 和 $g(x, k)$ 都是强凸函数;
- (2) 优化问题 $\min_x h(x)$ 和 $\min_x g(x, k)$ 都存在唯一解, 分别记为 x^* 和 x_k^* ;
- (3) 对任意的 k 都有 $\|x_k^* - x^*\| \leq \frac{m}{22k^2}$ 成立。

以上所提模型(2)、(4)和(16)都源于模型(15), 模型(2)在光滑因子趋于无穷大时的求解性能最好。文献[2]已经证明模型(4)的性能优于模型(2)。本文主要比较模型(16)和模型(4)的求解性能, 对于相同的数据规模 m 和相同的 k 而言, 模型(4)的最优解和原模型(15)的最优解的逼近上限为 $\frac{m}{22k^2}$, 而模型(16)可以达到 $\frac{m}{244k^2}$, 明显要更好。

4 数值实验

实验采用文献[2]的设计, 自动随机生成训练样本和测试样本数据, 比较 SSVM 模型(2)、PSSVM 模型(4)和模型(15)的求解结果。由于使用的光滑函数都一阶光滑的, 采用 BFGS 方法^[7,11]进行训练。比较的数据有训练正确率、测试正确率、算法结束时的目标函数值 opt 、算法结束时下降方向的模 ng 和算法结束时下降方向的步长 $steplength$; 算法的停机条件有两个: 其一是目标函数的梯度模满足 $ng < 10^{-3}$, 另外一个是一维搜索的步长满足 $steplength < 10^{-12}$, 满足停机条件一结束的训练也满足停机条件二, 反过来满足停机条件二结束的训练往往都不满足第一个停机条件, 因此满足停机条件一结束的训练视为完全训练, 而只满足停机条件二结束的训练是不完全的训练; 在完全训练的条件下, 训练正确率和测试正确率表现了训练结果的好坏。训练正确率越高, 说明经验风险越小; 测试正确率越高, 说明泛化能力越好。训练样本 (M, N) 表示 M 个 N 维的样本集, 测试样本个数为 1000, 测试样本属性个数和训练样本相同也为 N , 光滑因子取 5, 算法的初始化点为 $(N+1)$ 维空间的零点。保持样本属性 N 不变而改变训练样本数目 M 的数据结果如表 1 所列, 保持训练样本数目 M 不变而改变样本属性 N 的实验结果如表 2 所列。

表 1 训练样本变化是不同模型的实验结果

(每栏数据为训练正确率、测试正确率、算法结束时的目标函数值 opt 、算法结束时下降方向的模 ng 和算法结束时下降方向的步长 $steplength$)

训练样本规模 (M,N)	(100,10)	(200,10)	(400,10)	(600,10)	(800,10)	(1000,10)
SSVM 模型(2)	93.00%	90.20%	87.75%	84.17%	86.88%	86.10%
	83.40%	81.30%	84.60%	83.20%	84.10%	82.30%
	99.7715	230.0624	383.553261	427.7309	645.2431	900.2076
	2.2930e-004	2.1196e-004	1.2142e-005	9.6984e-005	7.5894e-004	1.6466e-005
	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
PSSVM 模型(4)	93.10%	90.30%	88.00%	84.33%	86.88%	86.15%
	83.50%	81.50%	84.70%	83.30%	84.20%	82.50%
	97.4484	224.000725	297.610713	416.5412	572.6668	858.6351
	1.6178e-004	1.3670e-004	2.0262e-005	7.7129e-005	6.9044e-005	1.6209e-004
	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
样条 PSSVM 模型(16)	93.30%	90.50%	88.10%	84.50%	87.00%	86.30%
	83.60%	81.60%	84.80%	83.40%	84.30%	82.60%
	97.4246	223.3668	229.2175	367.4549	485.6326	681.4930
	1.0198e-004	4.7776e-005	1.1944e-005	2.9465e-005	3.8753e-005	6.0568e-005
	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

表2 样本属性变化时不同模型的实验结果

(每栏数据为训练正确率、测试正确率、算法结束时的目标函数值 opt 、算法结束时下降方向的模 ng 和算法结束时下降方向的步长 $steplength$)

训练样本规模 (M,N)	(500,10)	(500,20)	(500,40)	(500,60)	(500,80)	(500,100)
SSVM 模型(2)	85.70%	91.10%	94.60%	96.00%	97.60%	98.00%
	85.60%	86.90%	90.60%	92.10%	91.40%	91.30%
	452.4911	408.6721	460.2407	541.5291	309.2833	345.9864
	2.7823e-005	2.0021e-004	1.1059e-004	1.3241e-004	2.3678e-004	1.0416e-004
	1.0000	0.1250	0.5000	0.2500	0.2500	0.2500
PSSVM 模型(4)	85.80%	91.20%	94.80%	96.20%	97.80%	98.10%
	85.70%	87.00%	90.90%	92.30%	91.50%	91.40%
	404.5372	379.4460	358.7926	277.3921	270.0451	254.7977
	2.3959e-005	1.8992e-004	6.3441e-005	2.5131e-004	1.6962e-004	1.1578e-004
	1.0000	0.2500	0.5000	0.125000	0.2500	0.2500
样条 PSSVM 模型(16)	85.90%	91.40%	94.90%	96.40%	98.10%	98.30%
	86.00%	87.10%	91.00%	92.50%	91.90%	91.60%
	387.7965	280.3890	355.4398	193.0173	185.9397	196.6830
	1.3163e-005	8.1232e-005	5.8683e-005	1.1065e-004	1.4424e-005	1.8371e-005
	1.0000	0.1250	0.2500	0.2500	0.2500	0.2500

根据以上数值实验分析可以得出:

从每栏数据的 ng 值可以看出以上数据都是在算法完全训练的情况下得到的,对于不同规模的问题使用 Sigmoid 函数的积分函数(1)、多项式函数(3)和样条多项式函数(14)的 SSVM 都可以很好地解决问题。不管是从训练样本 M 还是样本属性 N 的改变来看,模型(2)、模型(4)和模型(16)都有很好的适应能力和扩展性能。

比较表中的 ng 值,可以看出模型(16)的训练结果都明显小于模型(2)和模型(4),而且从目标函数值来看也是样条(16)最小,模型(4)次之,而模型(2)最大,因此样条模型(16)拥有最好的优化效果。

另外,从表1和表2的训练正确率和测试正确率两个方面不难看出模型(2)、模型(4)和模型(16)呈明显的递增趋势,因此模型(16)训练得到的分类器拥有较好的性能。

结束语 为了很好地保持正号函数中原点的信息,本文提出使用样条多项式函数来逼近正号函数并做二次处理的新方法;通过改进逼近正号函数的光滑样条多项式函数(13)给出了一个新的应用于 SSVM 的光滑样条多项式函数(14);对正号函数的逼近程度表明相比而言其有了很大改进。提出了新模型(16),在理论上给出了收敛性证明,并分析了求解性能的表现以及比较,理论分析和实验数据都表明样条模型(16)的求解效果更优越。

SVM 已经在模式识别领域得到了广泛的应用,而 SSVM 和 PSSVM 都是 SVM 的快速求解方案。相比而言,从理论分析和数值实验的角度可知本文提出的新的多项式光滑支持向量机有较好的性能表现,因此在数据挖掘和模式识别的相关领域有着广泛的应用前景。

本研究方向还存在很多有待进一步研究的课题,包括求解更高阶光滑的样条多项式函数和训练算法的改进和创新、将光滑的样条多项式函数应用于光滑支持向量机的回归问题等。

参 考 文 献

- [1] Lee Yuh-Jye, Mangasarian O L. SSVM: A Smooth Support Vector Machine for Classification[J]. Computer Optimization and Applications, 2001, 22(1): 5-21
- [2] 袁玉波, 严杰, 徐成贤. 多项式光滑的支持向量机[J]. 计算机学报, 2005, 28(1): 9-71
- [3] 熊金志, 袁华强, 彭宏. 多项式光滑的支持向量机一般模型研究[J]. 计算机研究与发展, 2008, 45(8): 1346-1353
- [4] 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
- [5] 熊金志, 胡金莲, 袁华强, 等. 支持向量回归机的光滑函数研究[J]. 模式识别与人工智能, 2008, 35(3): 366-370
- [6] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge, UK: Cambridge University Press, 2000
- [7] Yuan Y. A modified BFGS algorithm for unconstrained optimization[J]. IMA Journal of Numerical Analysis, 1991, 11(4): 325-332
- [8] Mangasarian O L, Musicant D R. Successive overrelaxation for support vector machines[J]. IEEE Transactions on Neural Networks, 1999, 10: 1032-1037
- [9] Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines[M]//Scholkopf B, Burges C J C, Smola A J, eds. Advances in Kernel Methods-Support Vector Learning. MIT Press, 1999: 185-208. <http://www.research.microsoft.com/~jplatt/smo.html>
- [10] Joachims T. Making large-scale support vector machine learning practical[M]//Scholkopf B, Burges C J C, Smola A J, eds. Advances in Kernel Methods-Support Vector Learning, MIT Press, 1999: 169-184
- [11] Yuan Y, Byrd R. Non-quasi-Newton updates for unconstrained optimization[J]. Journal of Computing Mathematics, 1995, 13(2): 95-107
- [12] Ahlberg J H, Nilson E N, Walsh J L. The Theory of Splines and Their Applications[M]. New York: Academic Press, 1967
- [13] De Boor C. A Practical Guide to Splines[M]. New York: Springer Verlag, 1978(2001)
- [14] Yuan Yubo, Li Chunzhong. A New Smooth Support Vector Machine[J]. Lecture Notes in Computer Science, 2005: 3802
- [15] Luo L, Lin C, Peng H, et al. A Study on Piecewise Polynomial Smooth Approximation to the Plus Function[C]//Proceeding of ICARCV. 2006: 2177-2182
- [16] Jiang M, Meng Z, Zhou Z. A Smoothing Support Vector Machine Based on Quarter Penalty Function[C]//International Conference on Computational Intelligence and Security, 2007
- [17] Fan Yan-feng, Zhang De-xian. Research on Tangent Circular Arc Smooth Support Vector Machine(TCA-SSVM) Algorithm[C]//Proceedings of the 2008 IEEE International Conference on Information and Automation. Zhangjiajie, China, June 2008: 20-23