

基于 Hermite 插值的 SVM 研究

范艳峰^{1,2}, 张德贤², 何华灿¹

FAN Yan-feng^{1,2}, ZHANG De-xian², HE Hua-can¹

1.西北工业大学 计算机学院, 西安 710072

2.河南工业大学 信息科学与工程学院, 郑州 450001

1.College of Computer Science, Northwest Polytechnical University, Xi'an 710072, China

2.College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450052, China

E-mail: fanyanfeng@haut.edu.cn

FAN Yan-feng, ZHANG De-xian, HE Hua-can. SVM research based on Hermite interpolation. Computer Engineering and Applications, 2007, 43(29): 179-181.

Abstract: In traditional SVM solution algorithms, objective function is a strictly convex unconstrained optimization problem, but is not differentiable due to x_+ , which precludes the use of most used optimization algorithms. Polynomial smooth techniques are applied to SVM model and replace x_+ by a very accurate smooth approximation that is Hermite Interpolation polynomial, thus the undifferentiable model is converted into a differential model. The deduction procedure of Hermite Interpolation polynomial smoothing x_+ is extended. Experiments with UCI datasets show the validity of the model.

Key words: SVM; Hermite interpolation; polynomial smooth

摘要: 在传统 SVM 的分类求解算法中, 由于严格凸的无约束最优化问题中单变量函数 x_+ 是不可微的, 不能使用通常的最优化的算法进行求解。三次 Hermite 插值多项式光滑的支持向量机模型采用的是一种多项式光滑技术, 用三次 Hermite 插值多项式代替单变量函数 x_+ , 将原来不可微的模型变为可微的模型, 并且给出了三次 Hermite 插值多项式光滑化单变量函数 x_+ 的推导过程。使用 UCI 机器学习数据集中的数据, 通过实验验证了该模型的有效性。

关键词: 支持向量机; Hermite 插值; 多项式光滑

文章编号: 1002-8331(2007)29-0179-03 文献标识码: A 中图分类号: TP18

1 支持向量机

1.1 标准 SVM

经验风险是对期望风险的估计, 尽管没有经过充分的理论论证, 在多年的机器学习中经验风险最小化准则占据了主要地位。但是在有限样本下, 经验风险最小化准则得到的结果并不一定能够使真实风险最小, 即训练误差小并不总能导致好的预测效果, 有时训练误差过小反而会致推广能力下降, 也就是导致真实风险的增加, 这就是“过学习”问题产生的原因。支持向量机针对小样本学习和分类问题, 利用了结构风险最小化的原则, 一方面克服了神经网络等方法所固有的过学习问题, 另一方面又有很强的非线性分类能力。

支持向量机用于模式识别的基本思想是将一个超平面作为决策平面, 不但能将分类中的两类样本正确分开, 而且使分类间隔最大, 即将优化问题转化为一个凸型的二次规划问题。

假定大小为 m 的训练样本集 $\{(x_i, y_i), i=1, 2, \dots, m\}$, 由二类组成, 如果 $x_i \in R^n$ 属于第 1 类, 则标记为正 ($y_i=1$), 如果属于第 2 类, 则标记为负 ($y_i=-1$)。学习的目标是构造一个决策函数, 将

测试数据尽可能正确地分类。

$x_i^T \omega - \gamma = \pm 1$ 代表两类数据的分类超平面, 正负类数据分别由以下两式判定

$$x_i^T \omega - \gamma \geq 1 \quad y_i = 1$$

$$x_i^T \omega - \gamma \leq -1 \quad y_i = -1$$

SVM 主要是构造一个分类超平面以分割两类不同的样本, 使得分类间隔 (margin) 最大。由此可以得到下面的二次优化问题

$$\min_{(\omega, \gamma, \nu) \in R^{n+1+m}} \nu e^T y + \frac{1}{2} \|\omega\|_2^2$$

$$\text{s.t. } D(A\omega - e\gamma) + y \geq e$$

$$\nu \geq 0$$

其中, ω 是分类超平面的法向量; γ 决定了边界面到原点的距离; ν 为惩罚参数, ν 越大表示对错误分类的惩罚越大; y 为训练样本线性不可分时引入的非负松弛变量。

1.2 光滑支持向量机 SSVM

给定 n 维实数空间 R^n , 其中有两类数据 A_+ 和 A_- , 分别属于正类和负类, 用 $m \times n$ 维矩阵表示数据集 A , 第 i 行数据 A_i 对

基金项目: 河南省教育厅基础研究 (the Fundamental Research of Department of Education of Henan Province, China under Grant No.2003520261)。

作者简介: 范艳峰 (1973-), 女, 博士生, 副教授, 主研方向: 数据挖掘; 张德贤 (1961-), 男, 教授, 博士生导师, 主研方向: 机器学习、神经网络; 何华灿 (1938-), 男, 教授, 博士生导师, 主要研究方向: 人工智能与逻辑。

应第*i*个数据。正类和负类分别记为+1和-1,可以用 $m \times m$ 维对角矩阵表示,对角元素为+1或-1,也即是 $D_{ii}=\pm 1$,代表第*i*个数据属于正类或负类。

上述最优化问题不是严格凸的二次规划,为了求解方便,可以把他们修正为严格凸二次规划

$$\begin{aligned} \min_{(\omega, \gamma) \in R^{n+1}} & \frac{\nu}{2} \|y\|_2^2 + \frac{1}{2} (\|\omega\|_2^2 + \gamma^2) \\ \text{s.t. } & D(A\omega - e\gamma) + y \geq e \\ & y \geq 0 \end{aligned} \quad (1)$$

该问题的目标函数中包含有 $\|y\|_2^2$ 项,并注意到约束条件,可知问题关于 y 的解应满足

$$y = (e - D(A\omega - e\gamma))_+ \quad (2)$$

其中 $(\cdot)_+$ 是单变量函数

$$x_+ = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$$

将上式代入严格凸的有约束最优化问题,就得到了严格凸的无约束最优化问题

$$\min_{(\omega, \gamma) \in R^{n+1}} \frac{\nu}{2} \|(e - D(A\omega - e\gamma))_+\|_2^2 + \frac{1}{2} (\|\omega\|_2^2 + \gamma^2) \quad (3)$$

它有唯一的最优解,但函数 $(\cdot)_+$ 是不可微的,是非光滑的,所以不能用通常的最优化问题解法求解,因为那里总要求目标函数的梯度、Hesse 矩阵存在。2001年 Yuh-Jye Lee 将式(3)的第一部分作适当的修改,将其光滑化,以建立一个与非光滑无约束问题近似的光滑无约束问题。为此, Lee 引入一个非光滑函数 $(\cdot)_+$ 的近似函数

$$p(x, \alpha) = x + \frac{1}{\alpha} \ln(1 + e^{-\alpha x}) \quad \alpha > 0 \quad (4)$$

其中 α 是参数。可以证明,当 $\alpha \rightarrow \infty$ 时,函数 $p(x, \alpha)$ 收敛于 x_+ , 见图 1。这样无约束最优化问题就近似于最优化问题

$$\min_{(\omega, \gamma) \in R^{n+1}} \frac{\nu}{2} \|p(e - D(A\omega - e\gamma), \alpha)\|_2^2 + \frac{1}{2} (\|\omega\|_2^2 + \gamma^2) \quad (5)$$

可以期望,当 α 充分大时,光滑无约束问题(5)的解会近似于非光滑无约束问题(3)的解。

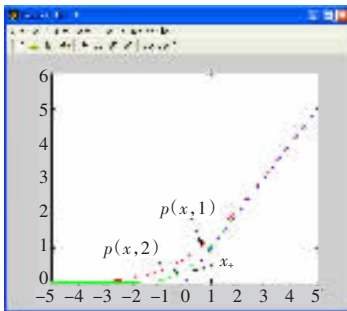


图1 $x_+, p(x, 1)$ 和 $p(x, 2)$ 函数图像

论文给出了一个使用 Hermite 插值多项式来光滑 $(\cdot)_+$ 单变量函数的具体过程,以其替代 $p(x, \alpha) = x + \frac{1}{\alpha} \ln(1 + e^{-\alpha x})$, 实验表明,分类性能较好。

2 Hermite 插值多项式

Hermite 插值多项式在差值节点上函数值相等,而且它的导数值也相等(即要求在节点上具有一阶光滑度),甚至要求高阶导数也相等,满足这种要求的插值多项式就是 Hermite 插值

多项式。

2.1 数学描述

设在节点 $a \leq x_0 < x_1 < \dots < x_n \leq b$ 上, $y_j = f(x_j), m_j = f'(x_j), (j=0, 1, \dots, n)$, 要求插值多项式 $H(x)$, 满足条件 $H(x_j) = y_j, H'(x_j) = m_j, (j=0, 1, \dots, n)$ 。满足 Hermite 插值条件的插值多项式 $H(x) = H_{2n+1}(x)$ 可写成

$$H_{2n+1}(x) = \sum_{j=0}^n [y_j \alpha_j(x) + m_j \beta_j(x)]$$

其中

$$\begin{cases} \alpha_j(x_k) = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}, \alpha_j'(x_k) = 0 \\ \beta_j(x_k) = 0, \beta_j'(x_k) = \delta_{jk} \end{cases} \quad (j, k=0, 1, \dots, n)$$

由上述条件可知, $H_{2n+1}(x_k) = y_k, H'_{2n+1}(x_k) = m_k, (k=0, 1, \dots, n)$ 。

$n=1$ 时称为三次 Hermite 插值, 三次 Hermite 插值多项式 $H_3(x)$ 满足条件 $H_3(x_k) = y_k, H_3(x_{k+1}) = y_{k+1}, H'_3(x_k) = m_k, H'_3(x_{k+1}) = m_{k+1}$ 。

三次 Hermite 插值多项式为:

$$H_3(x) = y_k \alpha_k(x) + y_{k+1} \alpha_{k+1}(x) + m_k \beta_k(x) + m_{k+1} \beta_{k+1}(x)$$

相应的插值基函数为 $\alpha_k(x), \beta_k(x), \alpha_{k+1}(x), \beta_{k+1}(x)$ 它们满足条件:

$$\begin{aligned} \alpha_k(x_k) &= 1, \alpha_k(x_{k+1}) = 0, \alpha'_k(x_k) = \alpha'_k(x_{k+1}) = 0, \\ \beta_k(x_k) &= \beta_k(x_{k+1}) = 0, \beta'_k(x_k) = 1, \beta'_k(x_{k+1}) = 0, \\ \alpha_{k+1}(x_k) &= 0, \alpha_{k+1}(x_{k+1}) = 1, \alpha'_{k+1}(x_k) = \alpha'_{k+1}(x_{k+1}) = 0; \\ \beta_{k+1}(x_k) &= \beta_{k+1}(x_{k+1}) = 0, \beta'_{k+1}(x_k) = 0, \beta'_{k+1}(x_{k+1}) = 1. \end{aligned}$$

根据 Hermite 插值的一般基函数表达式,可得到

$$\begin{cases} \alpha_k(x) = \left(1 + 2 \frac{x-x_k}{x_{k+1}-x_k}\right) \left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2 \\ \alpha_{k+1}(x) = \left(1 + 2 \frac{x-x_{k+1}}{x_k-x_{k+1}}\right) \left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2 \\ \beta_k(x) = (x-x_k) \left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2 \\ \beta_{k+1}(x) = (x-x_{k+1}) \left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2 \end{cases}$$

2.2 Hermite 插值多项式光滑化单变量函数

若对单变量函数 x_+ 进行光滑,可以在 $[-\frac{1}{a}, \frac{1}{a}]$ 区间对单变量函数进行光滑,采用三次 Hermite 插值多项式做光滑函数,并且该函数具有一阶光滑性。在 x_+ 函数 $x < 0$ 和 $x > 0$ 区间上对等距离取两插值点 $x_k = -\frac{1}{a}$ 和 $x_{k+1} = \frac{1}{a} (a > 0)$, 如图 2 所示。

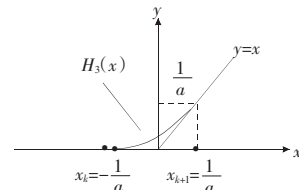


图2 单变量函数 x_+ 、Hermite 插值多项式 $H_3(x)$

单变量函数 x_+ 在 $[-\frac{1}{a}, \frac{1}{a}]$ 区间插值点、插值点对应函数值及其一阶导数值如表 1 所示。

因为 $y_k=0, m_k=0$, 所以无需计算 α_k 和 β_k , 则

$$\alpha_{k+1} = \frac{1}{4} (-a^3 x^3 + 3ax + 2)$$

表1 插值点、插值对应函数值及其一阶导数值

k	0	1
x_k	$-\frac{1}{a}$	$\frac{1}{a}$
y_k	0	$\frac{1}{a}$
m_k	0	1

$$\beta_{k+1} = \frac{1}{4} (a^2 x^3 + ax^2 - x - \frac{1}{a})$$

所以,在 $[-\frac{1}{a}, \frac{1}{a}]$ 区间上对单变量函数 x_+ 进行光滑化的三次

Hermite 插值多项式为

$$H_3(x) = y_{k+1} \alpha_{k+1}(x) + m_{k+1} \beta_{k+1}(x) = \frac{1}{4} ax^2 + \frac{1}{2} x + \frac{1}{4a}$$

根据插值点的位置不同,定义

$$H_3(x, a) = \frac{1}{4} ax^2 + \frac{1}{2} x + \frac{1}{4a}$$

当参数 a 取 1 时, $p(x, a), H_3(x, a)$ 的函数图像如图 3 所示。

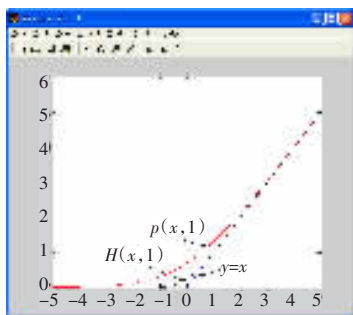


图3 $x_+, p(x, 1)$ 和 $H_3(x, 1)$ 函数图像

由图中可以看出, $H_3(x, a)$ 与 $p(x, a)$ 相比, 在 a 取相同值的情况下对原函数的逼近程度上更高。

文献[5]中提出了使用多项式 $\frac{1}{4}ax^2 + \frac{1}{2}x + \frac{1}{4a}$ 替代 $p(x, a)$, 但是没有给出多项式的推导过程。使用三次 Hermite 函数推导出多项式 $\frac{1}{4}ax^2 + \frac{1}{2}x + \frac{1}{4a}$ 来代替非光滑函数 $(\cdot)_+$, 使无约束最优化问题近似于最优化问题

$$\min_{(\omega, \gamma) \in R^{n+1}} \frac{v}{2} \| H_3(e-D(A\omega - e\gamma), \alpha) \|_2^2 + \frac{1}{2} (\| \omega \|_2^2 + \gamma^2)$$

3 实验结果

为了验证该模型的有效性, 选取 UCI 机器学习数据集中的数据集 Ionosphere 和 checker 进行实验, 每个数据集条件属性和决策属性的个数、测试集和训练集中实例的个数以及训练集和测试集的选择方法见表 2。

表2 Ionosphere 和 Two-Spirals 数据集情况列表

数据集	条件属性/决策属性	训练集、测试集选择方法
Ionosphere	34/1	351 个实例, 随机选取 150 个用于训练, 其余 201 个用于测试, 训练集和测试集中正反例各占 50%; good 为正例, bad 为反例
		1 000 个实例, 随机选取 700 个用于训练, 其余 300 个用于测试, 训练集和测试集中正反例各约占 50%; 0 为正例, 1 为反例
Checker	2/1	

对 Checker 数据集的一次训练实验中, 样本点的随机选取

分布如图 4 所示。

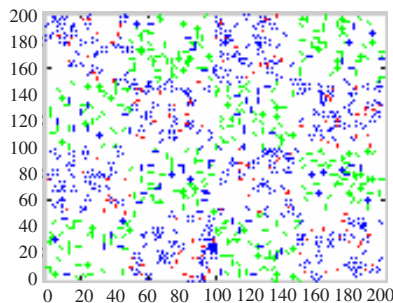


图4 checker 样本点分布

*: 训练集正例 +: 测试集正例 o: 训练集反例 .: 测试集反例

对每个数据集重复进行 30 次实验, 求出错误率的均值和标准差, 标准差反映分类器的稳定性。该算法模型的实验结果如表 3 所示。Lee 提出的 SSVM 模型中对 Ionosphere 数据集的 10-折交叉确认误差为 0.104。

表3 实验结果

	均值	标准偏差
Ionosphere	0.038	0.007
Checker	0.043	0.008

从实验结果看, 使用 Hermite 函数代替非光滑函数, 使无约束最优化问题近似于最优化问题来求解 SVM 分类器的分类效果是令人满意的。

4 结论

在对用于分类问题的 SVM 的求解算法研究中, 传统的带约束的优化模型无法使用通常的最优化的算法进行求解。虽然可以将带约束问题转化为无约束优化模型, 该优化问题的目标具有强凸性, 它具有唯一的最优解, 但是该模型不具有二阶光滑性。基于 Hermite 插值多项式光滑的支撑向量机模型采用的是一种多项式光滑技术, 将原来不可微的模型变为可微的模型, 相应光滑模型可以用最优化的算法进行求解, 是实际应用中的一个很好的选择。

使用三次 Hermite 函数代替非光滑函数 $(\cdot)_+$, 在 a 取相同值的情况下, $H_3(x, a)$ 比 $p(x, a)$ 对原函数的逼近程度更高。采用基于三次样条插值多项式的光滑支持向量机模型能有效地解决二分类问题, 有效地改善分类器的性能, 具有很好的适应性和鲁棒性。(收稿日期: 2007 年 1 月)

参考文献:

- [1] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995.
- [2] Deng N Y, Zhang H B, Zhang C H. Further improvement of the newton-PCG algorithm with automatic differentiation[J]. Optimization Methods and Software, 2001(16): 151-178.
- [3] Lee Yuh -Jye, Mangasarian O L. SSVM: a smooth support vector machine Technical Report 99-03[R]. Data Mining Institute, Computer Science Department, University of Wisconsin, Madison, Wisconsin, 1999-09.
- [4] Dennis J e, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations[M]. Englewood Cliffs, N. J.: Prentice-Hall, 1983.
- [5] 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.