

互补支持向量机

张襄松 刘三阳

(西安电子科技大学理学院 西安 710071)

摘 要 基于支持向量机的修正模型,得到一个互补支持向量机。利用 Fischer-Burmeister 互补函数,提出了一个新的下降算法。该算法不是基于支持向量机最优化问题本身,而是一个与之等价的互补问题。新算法不需要计算任何 Hesse 矩阵或矩阵求逆运算,实现简单,计算量小,克服了 Mangasarian 等人提出的 LSVM 算法需要求逆矩阵而造成不适合求解大规模非线性分类问题的缺陷。在不需要任何假设的情况下,证明了算法的全局收敛性。仿真实验表明算法是可行有效的。

关键词 支持向量机,互补问题,下降算法,全局收敛

中图法分类号 O224,TP18 **文献标识码** A

Complementarity Support Vector Machines

ZHANG Xiang-song LIU San-yang

(Dept. of Mathematical Sciences, Xidian University, Xi'an 710071, China)

Abstract A complementarity support vector machine was obtained which is based on a amended problem of support vector machine. By using Fischer-Burmeister function, a new descent algorithm for support vector machine optimization problem was presented. The proposed algorithm does not base on the primal quadratic programming problem of SVM, but a complementarity problem. It mustn't compute any Hesse or the inverse matrix with simple and small computational work. And the shortcoming of Lagrangian method proposed by Mangasarian et al., which need compute the inverse matrix that is not adapted to handle nonlinear large-scale classification problems, is overcome. Furthermore, without any assumption, the global convergence is proved. Numerical experiments show that the algorithm is feasible and effective.

Keywords Support vector machines, Complementarity problem, Descent method, Global convergence

1 引言

支持向量机(SVM)^[1]是建立在统计学习理论和结构风险最小原理基础上的一种新的机器学习工具。SVM 由于具有简洁的数学形式、直观的几何解释和良好的泛化性能而受到广泛关注,并已成功应用于模式识别、回归分析等众多领域^[2]。标准的支持向量机模型是一个二次规划(QP)问题。求解二次规划问题的算法过程随着数据规模的增大会变得越来越复杂,因此寻找可行有效的算法成为人们研究的热点^[3-5]。常用的算法包括块算法、分解算法和 SMO 算法等,它们的共同思想是用分解的方法,通过求解一系列小规模子 QP 问题来实现对原问题的求解,这些算法在一定程度上节省了计算机内存,提高了计算效率,但算法的设计和实现比较复杂。为了解决大规模数据处理问题,除了基于二次规划问题中 Hesse 矩阵分解这一思路外,另一条思路就是利用标准 SVM 的思想,研究新的 SVM 优化模型与算法。Mangasarian 等人提出了 Lagrangian 支持向量机模型(LSVM),并给出了一个比较简单的 LSVM 算法^[6],但是算法中仍存在矩阵求逆运算,特别是对于非线性分类问题,需要直接求解,

当数据规模很大时,高阶矩阵的求逆运算直接影响了算法的训练速度。为克服这个缺陷,本文基于支持向量机的修正模型,求解一个与原问题等价的互补问题,利用 Fischer-Burmeister 互补函数^[7],从一个新的角度给出了求解该问题的一个简单高效的算法,避免了求解 Hesse 矩阵或是矩阵的求逆运算,减少了运算量,提高了运算效率。在不需要任何假设的情况下,算法具有全局收敛性。为验证算法的可行性和有效性,最后给出了数值实验。

2 标准支持向量机与 LSVM 模型

设输入样本是由分别属于两类的 m 个 n 维的样本向量 $\{(x_i, y_i)\}_{i=1}^m$ 组成,其中 $x_i \in R^n$, $y_i = +1$ 或 -1 表示对应的模式向量的类别。 $A = [x_1, x_2, \dots, x_m]^T$ 是一个 $m \times n$ 矩阵, $D = \text{diag}(y)$ 是 $m \times m$ 的对角阵,其中 $y = [y_1, y_2, \dots, y_m]^T$ 。则线性的标准支持向量机模型就是下面的二次规划问题:

$$\begin{aligned} \min_{(w, \gamma, \xi) \in R^{n+1+m}} & \frac{1}{2} w^T w + C e^T \xi \\ \text{s. t. } & D(Aw - eb) + \xi \geq e, \xi \geq 0 \end{aligned} \quad (1)$$

其中,参数 $C > 0$, $e = (1, 1, \dots, 1)^T$ 。

收稿日期:2009-03-19 返修日期:2009-06-01 本文受国家自然科学基金(项目编号:60674108)资助。

张襄松(1981—),女,博士生,研究方向为最优化理论与算法、数据挖掘,E-mail:xs-zhang@hotmail.com;刘三阳(1959—),男,博士,教授,研究方向为最优化理论与算法、信息网络算法与性能优化。

对二次规划(1),通常转化为它的对偶问题^[8]求解:

$$\begin{aligned} \min_{u \in R^m} & \frac{1}{2} u^T D A A^T D u - e^T u \\ \text{s. t. } & y^T u = 0, 0 \leq u \leq C e \end{aligned} \quad (2)$$

其中, u 是对偶问题(2)的决策变量。通过求解上面的二次规划得到最优分类面。

不妨设 w^* 和 b^* 为优化问题(1)的最优解,则最优分类面为 $w^* x - b^* = 0$ 。而对于非线性分类问题,最优分类面则可通过求解特征空间的对偶规划得到:

$$\begin{aligned} \min_{u \in R^m} & \frac{1}{2} u^T D K(A, A) D u - e^T u \\ \text{s. t. } & y^T u = 0, 0 \leq u \leq C e \end{aligned}$$

其中, $K(A, A)$ 是 $m \times m$ 半正定矩阵,它的第 (i, j) 个元素由核函数 $K(x_i, x_j)$ 确定。

在标准的 SVM 模型(1)中,目标函数对松弛变量取 1-范数,并用 C 进行加权。若对目标函数进行修正,则可得到如下修正模型:

$$\begin{aligned} \min_{(w, \gamma, \xi) \in R^{n+1+m}} & \frac{1}{2} (w^T w + b^2) + \frac{1}{2} C \xi^T \xi \\ \text{s. t. } & D(Aw - eb) + \xi \geq e \end{aligned} \quad (3)$$

式(3)模型的修正能导出目标函数的强凸性等优良性质,且对原问题的影响很小^[8]。式(3)的对偶即得 LSVM 模型:

$$\min_{0 \leq u \in R^m} f(u) := \frac{1}{2} u^T Q u - e^T u \quad (4)$$

其中, $Q = \frac{I}{C} + H H^T$, $H = D[A - e]$ 。利用 KKT 充分必要性条件,式(4)转化为下面问题:

$$\begin{aligned} u \geq 0, Qu - e \geq 0, u^T (Qu - e) &= 0 \\ \text{Mangasarian 等人给出了式(5)的等价形式} \\ Qu - e &= ((Qu^i - e) - \gamma u^i)_+ \end{aligned} \quad (5)$$

并利用 $u^{i+1} = Q^{-1}[e + ((Qu^i - e) - \gamma u^i)_+]$ 提出了一个简单的迭代算法(LSVM 算法),该算法中存在求逆矩阵的运算。对线性分类问题,可以使用等式 Sherman-Morrison-Woodbury(SMW)来计算逆矩阵:

$$Q^{-1} = \left(\frac{I}{C} + H H^T \right)^{-1} = C \left(I - H \left(\frac{I}{C} + H^T H \right)^{-1} H^T \right)$$

从而提高运算效率。但是对于非线性分类问题,却不能直接使用,这就使得 LSVM 算法中需要直接计算 $Q^{-1} = \left(\frac{I}{C} + D(K(A, A) + ee^T)D \right)^{-1}$,因此该方法只适用于求解中小规模的非线性分类问题^[6]。

3 新的下降算法

本节将从不同的角度给出一个求解问题(5)的新算法。由上面的分析可以看出求解支持向量机的二次规划问题等价于求解一个互补问题。对于互补问题,通常是利用互补函数(C-函数)将其转化为一个无约束最优化问题或是一个非线性方程组^[9]来求解。为此,首先介绍下面两个概念^[10]。

定义 1 映射 $F: K \subseteq R^n \rightarrow R^n$ 是强单调的,如果对任意 $(a, b) \in K$,存在一个常数 $\lambda > 0$ 使得:

$$(F(a) - F(b))^T (a - b) \geq \lambda \|a - b\|^2 \text{ 成立。}$$

定义 2 对任意 $(a, b) \in R^2$,映射 $\phi: R^2 \rightarrow R$ 是一个互补函数(C-函数),如果:

$$\phi(a, b) = 0 \Leftrightarrow 0 \leq a \perp b \leq 0 \quad (6)$$

Fischer-Burmeister 函数是求解互补问题常用的一类 C-函数,其定义为 $\phi_{FB}: R^2 \rightarrow R$:

$$\phi_{FB}(a, b) = a + b - \sqrt{a^2 + b^2} \quad (7)$$

令 $q(u) = Qu - e$,则问题(5)可转化为下面的方程组:

$$\Phi(u) := \begin{bmatrix} \phi(u_1, q_1(u)) \\ \vdots \\ \phi(u_m, q_m(u)) \end{bmatrix} = 0$$

它等价于一个价值函数的无约束最小化问题:

$$\min_{u \in R^m} \varphi(u) := \frac{1}{2} \|\Phi(u)\|^2$$

且互补问题与原问题的最优解之间存在如下关系:

$$w = A^T D u, \xi = \frac{u}{C}, b = -y^T u$$

基于上面讨论,给出求解该问题的一个下降算法。

算法 1

Step0 初始化。任意选取初始点 $u^0 \in R^m$,选择参数 $\sigma \in (0, 1)$, $\delta \in (0, 1/2)$ 。置 $k := 0$ 。

Step1 令 $p^k := \varphi(u^k)$,如果 $p^k = 0$,停;否则,转 Step2。

Step2 计算搜索方向: $d^k := -\nabla_q \varphi(u^k, q(u^k))$ 。

Step3 令 l 是满足下面搜索的最小非负整数 l^k , $\varphi(u^k + \delta^k d^k) \leq (1 - \sigma \delta^{2^k}) p^k$ 。

Step4 令 $u^{k+1} := u^k + \delta^k d^k$; $k := k + 1$,转 Step1。

算法复杂度分析:对于非线性分类问题,LSVM 算法^[8]需要记录 2 个 $m \times m$ 的矩阵 Q , Q^{-1} 和 2 个 m 维的向量,而计算 Q^{-1} 需要大量运算,且每次迭代至少需要 2 次矩阵与向量的乘法。从而可知其存储复杂度为 $O(2m^2)$,单次迭代运算复杂度为 $O(4m^2)$,开始迭代前的计算逆矩阵复杂度为 $O(m^3)$;而算法 1 需要记录 2 个 $m \times m$ 的矩阵,其存储复杂度为 $O(2m^2)$,每一次迭代需要计算 10 次矩阵与向量的乘法,因此单次迭代运算复杂度为 $O(10m^2)$,但没有矩阵求逆运算,从而降低了算法的复杂度。下面证明算法是全局收敛的。

引理 1 如果 u^k 不是互补问题(5)的解,则搜索方向 $d^k := -\nabla_q \varphi(u^k, q(u^k))$ 是 $\varphi(u)$ 在点 u^k 处的一个下降方向,即满足 $\nabla \varphi(u^k)^T d^k < 0$ 。

证明:因为 $\nabla \varphi(u^k) = \nabla_u \phi(u^k, q(u^k)) + \nabla_q \varphi(u^k) \nabla_q \phi(u^k, q(u^k)) = \nabla_u \phi(u^k, q(u^k)) + Q \nabla_q \phi(u^k, q(u^k))$ 。

所以 $\nabla \varphi(u^k)^T d^k = -\nabla_u \phi(u^k, q(u^k)) \cdot \nabla_q \varphi(u^k, q(u^k)) - (d^k)^T Q d^k$ 。由 Q 的定义可知, Q 是正定矩阵,故上式中第 2 项 $(d^k)^T Q d^k > 0$,另一方面通过求导运算可得 $\nabla_u \phi(u^k, q(u^k))$ 和 $\nabla_q \varphi(u^k, q(u^k))$ 的表达式,容易验证上面第 1 项中: $\nabla_u \phi(u^k, q(u^k)) \cdot \nabla_q \varphi(u^k, q(u^k)) \geq 0$,故 $\nabla \varphi(u^k)^T d^k < 0$ 成立。

定理 1 设 $\{u^k\}$ 是由算法 1 产生的无穷序列,则至少有一个聚点且任意聚点都是互补问题(5)的解。

证明:因为 Q 是正定矩阵,故 $q(u)$ 是强单调的且互补问题(5)有唯一解。类似文献[11]中的命题 3.1,可以证明定理 1 成立。

4 数值实验

为验证算法 1 的可行性和有效性,本节对给定的样本集进行了数值实验。所有的实验均在 1.86GHz CPU, 1G 的 PC 机上完成,采用 MATLAB 7.0.1 编程。实验中取参数 $C = 1/m$,算法初始点 $u^0 = 0.001(1, 1, \dots, 1)^T$ 。

实验 1 Checkerboard 数据集^[12]由于具有非线性数据的典型特征,因此经常被用于检验非线性分类器学习算法。从 Checkerboard 数据集中读取 50 个样本点作为分类器的训练

(下转第 206 页)

抽取算法[J]. 计算机工程, 2006, 32(2): 203-205

[2] 王雷, 陈治平, 李志成. 基于文本分块的多模板隐马尔可夫模型的文本信息抽取[J]. 山东大学学报: 理学版, 2006, 41(3): 21-24

[3] 杜世平, 李海. 二阶隐马尔可夫模型及其在计算语言学中的应用[J]. 四川大学学报: 自然科学版, 2004, 41(2): 284-289

[4] 林亚平, 刘云中, 陈治平. 基于最大熵的隐马尔可夫模型文本信息抽取[J]. 电子学报, 2005, 33(2): 236-241

[5] 胡宇舟, 王雷, 顾学道. 基于多模板隐马尔可夫模型的文本信息抽取算法[J]. 计算机应用, 2008, 28(3): 609-702

[6] 邢永康, 马少平. 一种基于 Markov 链模型的动态聚类方法[J]. 计算机研究与发展, 2003, 40(2): 129-135

[7] Freitag D, McCallum A. Information extraction with HMMs and shrinkage[A] // Proceedings of the AAAI'99 Workshop on Machine Learning for Information Extraction[C]. Orlando: AAAI Press, 1999: 31-36

[8] Freitag D, McCallum A. Information extraction with HMM structures learned by stochastic optimization[A] // Proceedings of the Eighteenth Conference on Artificial Intelligence[C]. Edmonton: AAAI Press, 2002: 584-589

[9] Ray S, Craven M. Representing sentence structure in hidden Markov models for information extraction[A] // Proceedings of the Seventeenth International Joint Conference On Artificial Intelligence[C]. Washington: Morgan Kaufmann, 2001: 1273-1279

[10] Scheffer T, Deomain C, Wrobel S. Active hidden Markov models for information extraction[A] // Proceedings of the Fourth International Symposium on Intelligent Data Analysis[C]. Lisbon: Springer, 2001: 301-309

[10] 朱征宇, 周智, 罗颖, 等. 基于浏览行为量化分析的兴趣网页提取[J]. 重庆工学院学报: 自然科学版, 2009, 23(7): 79-84

(上接第 166 页)

集, 样本数据分布如图 1 所示。核函数为高斯核函数: 取 $\sigma^2 = 0.02$, $K(x_i, x_j) = \exp(-\sigma^2 \|x_i - x_j\|^2)$ 。

利用算法 1 得到的分类结果如图 2 所示。实验 1 的结果表明本文提出的下降算法是可行有效的。

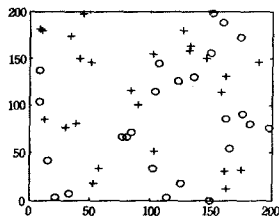


图 1 实验 1 的样本数据

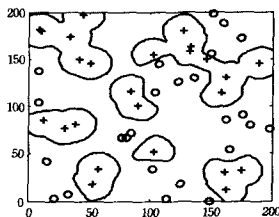


图 2 利用算法 1 的分类结果

实验 2 该实验对来自文献[13, 14]的不同规模和维度的实际数据集进行试验。这些数据都进行了零均值和标准化处理。分别采用 LSVM 迭代算法和本文提出的算法 1 对数据集进行分类。为方便起见, 均采用高斯核函数。结果如表 1 所列。

表 1 算法 1 与 LSVM 对实际数据的性能比较

Input Dataset		Parameters	LSVM		Descent method	
Datasets	Dimension/ train/test	σ^2	Time (s)	Errors (%)	Time (s)	Errors (%)
Titanic	3/150/2051	1	0.023	21.84	0.014	21.84
Diabetis	8/468/300	10	0.291	23.67	0.031	23.67
Breast cancer	9/200/77	0.2	0.047	22.15	0.015	22.08
Heart	13/170/ 100	0.1	0.031	12.35	0.016	11.93
Image	18/1300/ 1010	0.1	7.241	3.61	0.135	3.18
German	20/700/ 300	1	1.328	23.63	0.048	22.86

实验结果表明, 随着数据规模的增大, 算法 1 所需时间并没有明显增加。在错误率与 LSVM 算法相当的情况下, 其耗时较短, 表现出明显的优势。

结束语 基于支持向量机一个修正模型, 提出了互补型支持向量机, 进一步给出了求解该支持向量机的一个新的下降算法。对实际数据集的实验结果表明, 该算法的训练速度要比 LSVM 算法快, 且训练正确率相当甚至更好。同时, 算法由于不需要求解 Hesse 阵以及任何矩阵的求逆运算, 因此适用于求解大规模的非线性分类问题, 为求解 SVM 优化问

题提供了一个新的途径。

参考文献

- [1] Vapnik V N. The Nature of Statistical Learning Theory (second Edition)[M]. SNew York: Springer, 2000
- [2] 王巍, 赵国杰, 陈作斌. 一种新的混合智能分类模型及其应用[J]. 兰州大学学报, 2008, 44(1): 77-81
- [3] 廖东平, 魏玺, 章黎湘, 等. 一种新的支持向量机快速训练算法[J]. 系统工程与电子技术, 2007, 11: 1954-1957
- [4] 王玲, 薄列峰, 刘芳, 等. 稀疏隐空间支持向量机[J]. 西安电子科技大学学报, 2006, 33(6): 896-901
- [5] Zhou S S, Liu H W, Zhou L H, et al. Semismooth Newton support vector machine[J]. Pattern Recognition Letters, 2007, 28: 2054-2062
- [6] Mangasarian O L, Musicant D R. Lagrangian Support Vector Machines[J]. Journal of Machine Learning Research, 2001, 1: 161-177
- [7] Fischer A. A Special Newton-type optimization methods[J]. Optimization, 1992, 24: 269-284
- [8] Mangasarian O L, Musicant D R. Successive over relaxation for support vector machines[J]. IEEE Trans. on Neural Networks, 1999, 10: 1032-1037
- [9] Zhang X Z, Jiang H F, Wang Y J. A smoothing Newton-type method for generalized nonlinear complementarity problem[J]. Journal of Computational and Applied Mathematics, 2008, 212: 75-85
- [10] Facchinei F, Pang J S. Finite-Dimensional Variational Inequalities and Complementarity Problems[M]. New York: Springer, 2003
- [11] Chen J S, Pan S H. A descent method for a reformulation of the second-order cone complementarity problem[J]. Journal of Computational and Applied Mathematics, 2008, 213: 547-558
- [12] Ho T K, Kleinberg E M. Checkerboard dataset[EB/OL]. <http://www.cs.wisc.edu/math-prog/mpml.html>, 1996
- [13] Murphy P M, Aha D W. UCI Repository of machine learning databases[EB/OL]. <http://www.ics.ci.edu/~mlearn/ML-Repository.html>, 1998
- [14] Blake C L, Merz C J. UCI Repository of machine learning databases[EB/OL]. <http://www.ics.uci.edu/mlearn/ML-Repository.html>, 1998