

文章编号: 1001-0920(2008)08-0935-03

直接支持向量机

杜喆, 刘三阳

(西安电子科技大学 数学科学系, 西安 710071)

摘要: 基于最小二乘支持向量机变形, 得到一个极其简单快速的分类器——直接支持向量机. 与最小二乘支持向量机相比, 该分类器只需直接求解一个更小规模矩阵的逆, 大大减小了计算量, 并未降低分类精度. 从理论上证明了该矩阵可逆, 保证了分类面存在的唯一性. 对于线性情形, 采用 Sherman-Morrison-Woodbury 公式降低可逆矩阵的维数, 进一步减少了计算复杂度, 使其可适用于更大规模的样本集. 数值实验表明, 新分类器可行并具有上述优势.

关键词: 支持向量机; 分类; 线性方程组; 最小二乘

中图分类号: TP18 文献标识码: A

Direct support vector machine

DU Zhe, LIU San-yan

(Department of Applied Mathematics, Xidian University, Xi'an 710071, China. Correspondent: DU Zhe, E-mail: duzhe_doog@126.com)

Abstract: Based on the transmutation of least square support vector machine (LSSVM), an extremely fast and simple classifier — direct support vector machine (DSVM) is proposed. Compared with LSSVM, the classifier only needs to calculate the inversion of a smaller scale matrix, which greatly reduces the computing amount and has no decreasing classification accuracy. The inversion of the matrix is guaranteed theoretically, and hence a unique classifying hyperplane is existed. For linear classification, the Sherman-Morrison-Woodbury formula is employed to reduce the dimension of the matrix and the computing complexity is decreased, so that the classifier can be applied to larger scale dataset. Numerical experiments show that the classifier is easily operated and has the above advantages.

Key words: Support vector machine; Classification; Linear equation system; Least squares

1 引言

支持向量机(SVM)^[1,2]因其良好的性能,已成为近年来的研究热点. SVM的研究重点主要集中在精度和速度两大方面,标准 SVM 需求解一个约束二次规划^[1,2]. 为提高训练速度,通常采用 SMO 和 SVM^{light}等技术求解^[3]. 为将原问题简化,很多学者从变形角度作了相关研究. 文献[4]提出基于线性规划的 SVM 变型,计算速度大大加快,泛化能力逼近标准 SVM;文献[5]提出最小二乘支持向量机(LSSVM). 从几何上看,LSSVM 是将两类样本在原空间或特征空间中使用两个平行的超平面最小二乘逼近,从而使这两个超平面尽可能分离. 从表达式上看,LSSVM 将标准 SVM 的损失函数由一次变为二次,将不等式约束改变为等式约束,从而将 KKT 条件转化为一个比核矩阵更大规模的矩阵为

系数矩阵的线性方程组,通过解这个方程组即可求出 Lagrangian 函数的鞍点. 虽然从实验结果^[6]上看,分类精度不及标准 SVM,但避免了求解二次规划,可大大降低计算复杂度. 文献[7]对将最小二乘支持向量机用于回归方面进行了研究. 文献[8]研究了二次损失函数的 SVM,使得对偶规划中没有松弛变量的非负约束,从而简化了问题,并且数值实验表明其分类能力没有下降. Mangasarian 等人^[9,10]提出将阈值的平方加到标准 SVM 的目标函数中使二次规划变得更加简单,并且文献[10]利用 Sherman-Morrison-Woodbury (SMW) 公式^[11]求矩阵逆,给出了线性情形的快速算法.

本文利用文献[9,10]的思想,将阈值的平方加入 LSSVM 的目标函数中,得到一种新的分类器——直接支持向量机. 该分类器在求解分类面时,

收稿日期: 2007-06-06; 修回日期: 2007-08-31.

基金项目: 国家自然科学基金项目(60574075, 60705004).

作者简介: 杜喆(1982—),男,陕西米脂人,博士生,从事机器学习、最优化计算方法的研究; 刘三阳(1959—),男,西安人,教授,博士生导师,从事智能信息处理、最优化方法等研究.

与最小二乘支持向量机相比,其只需对一个关于核矩阵简单变形的矩阵求逆,泛化能力没有降低,且能降低所求方程组的规模,减少计算复杂度.经过理论证明该矩阵可逆,可保证方程组解的唯一性和算法的可行性.对于线性情形,与文献[10]类似,采用SMW公式求矩阵的逆,降低求逆矩阵的维数,进一步减少计算量.最后通过数值实验表明了该方法的有效性.

2 分类器设计

设样本集 $\{x_i, y_i\}_{i=1}^n \subset R^m \times \{-1, +1\}$, 则用矩阵 $A \in R^{n \times m}$ 来表示训练集,每行为一个样本点; D 是一个对角元素为类别标号的对角阵; s 为松弛变量; e 为一个所有元素为 1 的列向量;分类超平面为 $w^T x + b = 0$. 问题可转化为求解规划

$$\begin{aligned} \min \quad & \frac{C}{2} \|s\|^2 + \frac{1}{2} (w^T w + b^2), \\ \text{s.t.} \quad & D(Aw + eb) + s = e, \end{aligned} \quad (1)$$

其中 C 为正实数. 其 Lagrangian 函数为

$$L(u, b, s, v) = \frac{C}{2} \|s\|^2 + \frac{1}{2} (w^T w + b^2) - v^T (D(Aw + eb) + s - e), \quad (2)$$

v 为 Lagrangian 乘子. 根据 KKT 条件, 可得

$$\begin{cases} w - A^T Dv = 0, \\ b - e^T Dv = 0, \\ Cs - v = 0, \\ D(Aw + eb) + s - e = 0. \end{cases} \quad (3)$$

将式(3)中 w, b, s 分别用 v 表示, 有

$$w = A^T Dv, \quad b = e^T Dv, \quad s = C^{-1} v. \quad (4)$$

把式(4)带入(3)中最后一个方程, 有

$$DAA^T Dv + Dee^T Dv + C^{-1} v = e, \quad (5)$$

所以

$$v = (D(AA^T + E)D + C^{-1}I)^{-1} e, \quad (6)$$

E 是一个所有元素为 1 的 $n \times n$ 矩阵. 根据式(4)得到分类器为

$$y(x) = \text{sign}(w^T x + b) = \text{sign}(v^T DAx + e^T Dv).$$

一般情况下, 样本数大于样本的维数, 即 $n > m$. 记 $H = D[A \quad e]$, $H \in R^{n \times (m+1)}$, 则 $v = (HH^T + C^{-1}I)^{-1} e$, 使用 SMW 公式, 可得

$$v = C(I - H(C^{-1}I + H^T H)^{-1} H^T) e. \quad (7)$$

这只需要求解一个 $(m+1) \times (m+1)$ 的矩阵的逆, 大大减少了计算时间和空间复杂度, 进一步提高了分类器的训练速度.

注 1 式(1)可完全解释为线性方程组 $D(Aw + eb) = e$ 关于变量 $\begin{bmatrix} w \\ b \end{bmatrix}$ 的最小二乘解.

对于非线性情形, 通过使用核函数, 将式(6)中矩阵 AA^T 替换为核矩阵 $K = K(A, A^T) = (A, A^T)$, 为隐映射, 则

$$v = (D(K + E)D + C^{-1}I)^{-1} e. \quad (8)$$

这样便得到非线性分类器

$$y(x) = \text{sign}(w^T(x) + b) = \text{sign}(v^T DK(A, x) + e^T Dv).$$

实际上当核函数取为线性核时, 非线性分类器是线性分类. 这样可将线性与非线性分类统一起来使用相同的程序.

在 LSSVM 中, 方程组系数矩阵未必正定, 所以求解算法受到限制, 而 DSVM 的系数矩阵正定可逆, 且维数减少; 若采用 SMW 公式, 则可以降低到更小. 此外若对 DSVM 进行如文献[7]的在线回归方面研究时, 并不需要对阈值调整, 因而更加简单. 下面再从理论上保证 DSVM 的有效性.

3 理论证明

利用对角阵 D 的特性, 对于式(8)有

$$\begin{aligned} v &= (D(K + E)D + C^{-1}I)^{-1} e = \\ &= (D(K + E + C^{-1}I)D)^{-1} e = \\ &= (D(K + E + C^{-1}I)^{-1}D) e, \end{aligned}$$

而 $K + E + C^{-1}I$ 可逆. 事实上, 对 $\forall x \in R^n, x \neq 0$, 有

$$\begin{aligned} x^T (K + E + C^{-1}I) x &= \\ x^T Kx + x^T Ex + x^T (C^{-1}I) x. \end{aligned} \quad (9)$$

由核矩阵性质^[2], 显然 $x^T Kx \geq 0$. 又 $x^T Ex > 0$, $x^T (C^{-1}I) x = C^{-1} \cdot x^T x > 0$, 故式(9)大于 0, 矩阵 $K + E + C^{-1}I$ 正定. 所以, 其逆矩阵是存在的, 因而式(6)和(8)可行, 从而该分类器可操作. 这也保证其解唯一, 分类超平面唯一.

4 数值实验

本文在 CPU 2.4 G, 内存 512 M 的 PC 机上进行数值实验, 编程语言采用 Matlab 7.0.1.

首先在人工数据集上进行线性和非线性分类模拟实验. 图 1 为线性分类情形. 对图 2 中的双螺旋线分类时, 采用径向基核函数 $K(x, y) = \exp(-0.3 \|x - y\|^2)$, $C = 0.1$. 可以看出, 线性与

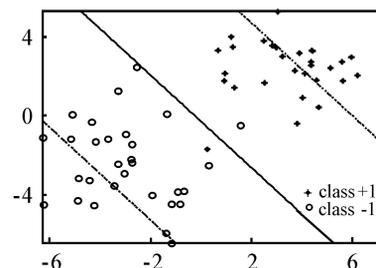


图 1 线性 DSVM 分类示意图

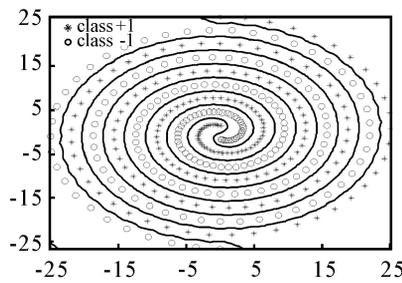


图 2 DSVM 对双螺旋线的分类图

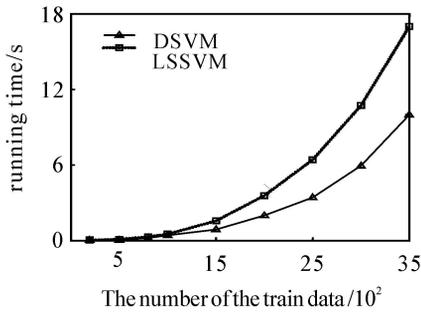


图 3 DSVM 与 LSSVM 运行时间比较图

非线性 DSVM 都有着良好的分类效果. 图 3 是在相同维数、不同数量的样本集下 LSSVM 与 DSVM 的运行时间比较 (核函数为径向基核, 实线为 DSVM). 可以看出, DSVM 具有明显的速度优势, 且随着样本规模增加而更加明显.

再对 UCI 数据库中的 4 个公开测试集进行 10 次交叉验证数值实验. 表 1 为 DSVM 与 LSSVM 在相同参数下的多项式核函数 $K(x, y) = (x^T y + 1)^p$ 分类结果比较; 表 2 为在径向基核函数 $K(x, y) = \exp(-s \|x - y\|^2)$ 下二者的结果. 两表的第 1 列是数据集样本维数、训练和测试样本个数, 以及核参数和惩罚因子 C 的值; 比较指标为对训练和测试集的分类错误率百分比 (均值 \pm 方差), 以及训练时

表 1 多项式核 DSVM 与 LSSVM 对 UCI 数据分类比较

Dataset		DSVM	LSSVM
Breast 9/200/77	Train/ %	26.15 \pm - 1.83	28.70 \pm - 1.73
	Test/ %	30.39 \pm - 4.69	30.64 \pm - 4.50
	Time/ s	0.010	0.016
Diabetis 8/468/300	Train/ %	18.23 \pm - 1.49	18.23 \pm - 1.49
	Test/ %	34.38 \pm - 2.32	34.38 \pm - 2.32
	Time/ s	0.047	0.079
German 20/700/300	Train/ %	29.89 \pm - 0.72	29.97 \pm - 0.68
	Test/ %	30.07 \pm - 1.59	30.07 \pm - 1.59
	Time/ s	0.125	0.188
Image 18/1300/1010	Train/ %	21.30 \pm - 0.21	21.31 \pm - 0.19
	Test/ %	23.18 \pm - 1.01	23.16 \pm - 1.02
	Time/ s	1.051	1.961

表 2 径向基核 DSVM 与 LSSVM 对 UCI 数据分类比较

Dataset		DSVM	LSSVM
Breast 9/200/77	Train/ %	18.52 \pm - 1.36	18.45 \pm - 1.37
	Test/ %	26.60 \pm - 4.33	26.62 \pm - 4.40
	Time/ s	0.008	0.015
Diabetis 8/468/300	Train/ %	17.20 \pm - 1.03	17.20 \pm - 1.03
	Test/ %	23.87 \pm - 1.64	23.87 \pm - 1.64
	Time/ s	0.045	0.079
German 20/700/300	Train/ %	17.36 \pm - 1.06	17.38 \pm - 1.08
	Test/ %	23.03 \pm - 1.97	23.05 \pm - 1.92
	Time/ s	0.126	0.187
Image 18/1300/1010	Train/ %	2.0 \pm - 1.23	2.0 \pm - 1.23
	Test/ %	4.01 \pm - 0.98	4.01 \pm - 0.69
	Time/ s	1.101	1.972

间. 表 1 和表 2 中数据最优值用粗体表示. 可以看出, 在分类错误率上 DSVM 与 LSSVM 相差无几, 有时还更低; 在时间上前者明显要少, 这和理论分析是一样的. 对于数据集 Diabetis 二者在两种核函数下的分类精度完全相同, 此时 DSVM 可代替 LSSVM.

5 结 论

本文通过对 LSSVM 变形, 得到直接支持向量机. 该分类器降低了原来所求解的线性方程组规模, 且在理论上保证系数矩阵非奇异, 也即保证分类面的唯一性, 这样可以直接对此方阵求逆, 因而计算简单快速. 大量数值实验表明, 其分类精度与最小二乘支持向量机基本相当, 但训练速度更快. 对于线性直接支持向量机, 可采用 SMW 公式, 进一步降低训练时间和空间开销, 使其适用于更大规模的样本集. 进一步工作可基于对直接支持向量机和最小二乘支持向量机理论方面的比较研究来展开.

参考文献(References)

- [1] Vapnik V. The nature of statistical learning theory[M], New York: Springer, 1995.
- [2] Cristianini N, Schawe-Taylor J. An introduction to support vector machines [M]. Cambridge: Cambridge University Press, 2000.
- [3] Platt J. Fast training of support vector machines using sequential minimal optimization[C]. Advances in Kernel Methods Support Vector Learning. Cambridge, 1999: 185-208.
- [4] 周伟达, 张莉, 焦李成. 线性规划支撑向量机[J]. 电子学报, 2001, 29(11): 1-5.
(Zhou W D, Zhang L, Jiao L C. Linear programming support vector machines[J]. Acta Electronica Sinica, 2001, 29(11): 1-5.)

(下转第 943 页)

- 682.
- [5] 胡云安, 晋玉强, 崔平远. 带有广义不确定性的导弹非线性控制系统设计[J]. 航空学报, 2004, 25(2): 153-157.
(Hu Y A, Jin Y Q, Cui P Y. Nonlinear controller design for missile system with a general set of uncertainties[J]. Acta Aeronautica and Astronautica Sinica, 2004, 25(2): 153-157.)
- [6] 刘燕斌, 陆宇平. 基于反步法的高超音速飞机纵向逆飞行控制[J]. 控制与决策, 2007, 22(3): 313-317.
(Liu Y B, Lu Y P. Longitudinal inversion flight control based on backstepping for hypersonic vehicle [J]. Control and Decision, 2007, 22(3): 313-317.)
- [7] 胡云安, 晋玉强, 张友安, 等. 基于神经网络的严反馈块非线性系统的鲁棒控制[J]. 控制与决策, 2004, 19(7): 808-812.
(Hu Y A, Jin Y Q, Zhang Y A, et al. NN based robust control for strict-feedback block nonlinear systems [J]. Control and Decision, 2004, 19(7): 808-812.)
- [8] 董文瀚, 孙秀霞, 林岩. 反推自适应控制的发展及应用[J]. 控制与决策, 2006, 21(10): 1081-1087.
(Dong W H, Sun X X, Lin Y. Adaptive backstepping control: Development and applications[J]. Control and Decision, 2006, 21(10): 1081-1087.)
- [9] Swaroop S, Hedrick J K, Yip P P, et al. Dynamic surface control for a class of nonlinear systems [J]. IEEE Trans on Automatic Control, 2000, 45(10): 1893-1899.
- [10] Wang D, Huang J. Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form [J]. IEEE Trans on Neural Networks, 2005, 16(1): 195-202.
- [11] 李红春, 张天平. 基于动态面控制的 MIMO 自适应神经网络控制[J]. 扬州大学学报, 2006, 9(4): 17-24.
(Li H C, Zhang T P. MIMO adaptive neural network control using dynamic surface control [J]. J of Yangzhou University, 2006, 9(4): 17-24.)
- [12] Khalil H K. 非线性系统[M]. 第 3 版. 北京: 电子工业出版社, 2005.
(Khalil H K. Nonlinear systems[M]. 3rd ed. Beijing: Publishing House of Electronics Industry, 2005.)
- [13] 陈谋. 不确定非线性综合火力/飞行/推进系统鲁棒控制方法研究[D]. 南京: 南京航空航天大学, 2004.
(Cheng M. Research on robust control method for uncertain nonlinear integrated fire/flight/propulsion system [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2004.)

(上接第 937 页)

- [5] Suykens J A K, Vandewalle J. Least square support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [6] Gestel T V, Suykens J A K. Benchmarking least squares support vector machine classifiers [J]. Machine Learning, 2004, 54(1): 5-32.
- [7] 王定成, 姜斌. 在线稀疏最小二乘支持向量机回归的研究[J]. 控制与决策, 2007, 22(2): 132-137.
(Wang D C, Jiang B. Online sparse least square support vector machines regression [J]. Control and Decision, 2007, 22(2): 132-137.)
- [8] 朱永生, 王成栋. 二次损失函数支持向量机性能的研究[J]. 计算机学报, 2003, 26(8): 982-989.
(Zhu Y S, Wang C D. Experimental study on the performance of support vector machine with squared cost function[J]. Chinese J of Computers, 2003, 26(8): 982-989.)
- [9] Mangasarian O L, Musicant D R. Lagrangian support vector machines [R]. Madison: University of Wisconsin, 2000.
- [10] Mangasarian O L, Musicant D R. Active support vector machine classification[R]. Madison: University of Wisconsin, 2000.
- [11] Golub G H, Van C F. Matrix computations[M]. 3rd ed. Baltimore: The John Hopkins University Press, 1996.