

最小二乘支持向量机算法研究^{*}

朱家元¹ 陈开陶² 张恒喜¹

(空军工程大学工程学院飞机与发动机工程系 西安710038)¹ (总装备部系统工程研究所 北京100101)²

Study of Least Squares Support Vector Machines

ZHU Jia-Yuan¹ CHEN Kai-Tao² ZHANG Heng-Xi¹

(Department of Aircraft and Engine Engineering, Air Force Engineering University, Xi'an 710038)¹

(General Equipment Department, Institute of System Engineering, Beijing 100101)²

Abstract In this paper, we present a least squares version for support vector machines(SVM) classifiers and function estimation. Due to equality type constraints in the formulation, the solution follows from solving a set of linear equations, instead of quadratic programming for classical SVM. The approach is illustrated on a two-spiral benchmark classification problem. The results show that the LS-SVM is an efficient method for solving pattern recognition.

Keywords Statistical learning theory, Support vector machines, Pattern Recognition, Least squares support vector machines, Neural networks

1 引言

支持向量机(SVM, Support Vector Machines)是基于结构风险最小化的统计学习方法,它具有完备的统计学习理论基础和出色的学习性能,在模式识别和函数估计中得到了有效的应用(Vapnik, 1995, 1998)^[1,2]。支持向量机方法一方面通过把数据映射到高维空间,解决原始空间中数据线性不可分问题;另一方面,通过构造最优分类超平面进行数据分类。神经网络通过基于梯度迭代的方法进行数据学习,容易陷入局部最小值,支持向量机是通过解决一个二次规划问题,来获得全局最优,而且支持向量机的计算复杂性和质量并不直接依赖于输入数据的维数(Vapnik, 1995, 1998),因此在解决模式识别和非线性函数估计中表现出了优秀的性能。近年, Suykens J. A. K^[3]提出一种新型支持向量机方法—最小二乘支持向量机(LS-SVM, Least Squares Support Vector Machines)用于解决分类和函数估计问题。

本文首先阐述支持向量机基本原理,然后给出用于模式识别和非线性函数估计的最小二乘支持向量机详细推理和算法,最后采用典型样本对最小二乘支持向量机学习和模式识别性能进行了研究。

2 支持向量机

Vapnik 提出的支持向量机首先是应用到模式识别当中,包括手写数字识别、人脸检测、文本自动分类和故障诊断等方面。用于分类的支持向量机基本原理和算法如下:

支持向量机是通过构造最优分类超平面(Optimal Hyperplane)实现分类的。设有 N 个训练样本的数据集 $\{y_k, x_k\}_{k=1}^N$, $x_k \in R^n$ 是第 k 个输入数据, $y_k \in R$ 是第 k 个输入数据的所属类别。要实现最优超平面对所有训练样本能够正确分类,必须满足:

$$y_k[(w^T \varphi(x_k) + b) - 1] \geq 0, k=1, \dots, N \quad (1)$$

容易验证,最优分类超平面就是满足(1)式条件,并且使得 $\|w\|^2$ 最小化的超平面。(1)式中的 $\varphi(\cdot)$ 是非线性函数,用于把数据映射到高维空间,但是它并不是显性结构函数。为了使最优分类超平面能够对线性不可分数据进行分类,需要在条件中引入非负松弛变量 $\xi_k \geq 0$, (1)式转化为:

$$y_k[(w^T \varphi(x_k) + b) - 1 + \xi_k] \geq 0, k=1, \dots, N, \quad \xi_k \geq 0, k=1, \dots, N. \quad (2)$$

根据结构风险最小化原则,风险界通过在(2)式的约束条件下优化下式进行最小化:

$$\min_{w, \xi_k} \Phi(w, \xi_k) = \frac{1}{2} w^T w + c \sum_{k=1}^N \xi_k \quad (3)$$

(3)式的解可通过构造下面的拉格朗日方程求出:

$$L(w, b, \xi_k; a_k, v_k) = \Phi(w, \xi_k) - \sum_{k=1}^N a_k \{y_k[(w^T \varphi(x_k) + b) - 1 + \xi_k]\} - \sum_{k=1}^N v_k \xi_k \quad (4)$$

其中, $a_k \geq 0, v_k \geq 0, (k=1, \dots, N)$ 。对 w, b 和 ξ_k 求偏微分,并令等于0,可得:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N a_k y_k \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N a_k y_k = 0 \\ \frac{\partial L}{\partial \xi_k} = 0 \rightarrow 0 \leq a_k \leq c, k=1, \dots, N \end{cases} \quad (5)$$

于是问题转化到求解下面的二次规划问题:

$$\max_{a_k} Q(a_k; \varphi(x_k)) = -\frac{1}{2} \sum_{k,l=1}^N y_k y_l \varphi(x_k)^T \varphi(x_l) a_k a_l + \sum_{k=1}^N a_k \quad (6)$$

$$\text{约束条件: } \begin{cases} \sum_{k=1}^N a_k y_k = 0 \\ 0 \leq a_k \leq c, k=1, \dots, N \end{cases}$$

^{*} 此论文受国防预研资助基金(项目编号: 98J19. 3. 2. JB3201)和 空军重点型号工程课题资助。朱家元 博士生,研究方向为装备发展规划与管理,数字化综合保障工程,支持向量机与智能信息处理,便携式维修计算机研究。陈开陶 助理研究员,研究方向为装备发展规划与管理,智能信息处理。张恒喜 教授,博导,研究方向为飞机系统工程,数字化综合保障工程。

定理 (Mercer 条件) 对于任意的对称函数 $K(x, x')$, 它在某个特征空间中的内积运算的充分必要条件是, 对于任意的 $\varphi(x)$ 不恒等于零, 且 $\int \varphi^2(x) dx > 0$, 有

$$\iint K(x, x') \varphi(x) \varphi(x') dx dx' > 0 \quad (7)$$

根据 Mercer 条件, $\varphi(x)^T \varphi(x_k) = \Psi(x, x_k)$. 因此 (6) 式二次规划问题可转化为求解下式:

$$\max_{a_i} Q(a_i; \Psi(x_k, x_l)) = -\frac{1}{2} \sum_{k,l=1}^N y_k y_l \Psi(x_k, x_l) a_k a_l + \sum_{k=1}^N a_k \quad (8)$$

约束条件不变。

支持向量机的分类决策函数为:

$$y(x) = \text{sgn} \left[\sum_{k=1}^N a_k y_k \Psi(x, x_k) + b \right] \quad (9)$$

其中, b 是实常数。对多数样本 a_k 将为零, 取值不为零 a_k , 且使 (1) 式等号成立的样本为支持向量。 $\Psi(\cdot, \cdot)$ 是核函数, 核函数的形式可以是: $\Psi(x, x_k) = x^T x_k$ (线性 SVM 核); $\Psi(x, x_k) = (x^T x_k + 1)^d$ (d 阶多项式 SVM 核); $\Psi(x, x_k) = \exp\{-\|x - x_k\|^2 / 2\sigma^2\}$ (径向基 SVM 核); $\Psi(x, x_k) = \tan(\kappa x^T x_k + \theta)$ (两层感知神经网络 SVM 核); 其中 σ, κ 和 θ 都是可调常数。

3 用于模式识别的最小二乘支持向量机

最小二乘支持向量机方法是采用最小二乘线性系统作为损失函数, 代替传统的支持向量机采用二次规划方法。设训练样本集 $D = \{(x_k, y_k) | k=1, 2, \dots, N\}$, $x_k \in R^n$, $y_k \in R$, x_k 是输入数据, y_k 是输出类别。在权 w 空间 (原始空间) 中的最小二乘支持向量机分类问题可以描述求解:

$$\min_{w,b,e} \Phi(w, b, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (10)$$

约束条件: $y_k [w^T \varphi(x_k) + b] = 1 - e_k, k=1, \dots, N$ 。

定义拉格朗日函数: $L(w, b, e; a) = \Phi(w, b, e) - \sum_{k=1}^N a_k \{y_k [w^T \varphi(x_k) + b] - 1 + e_k\}$ (11)

其中, 拉格朗日乘子 $a_k \in R$ 。对上式进行优化:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N a_k y_k \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N a_k y_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow e_k = \gamma e_k \\ \frac{\partial L}{\partial a_k} = 0 \rightarrow y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0 \end{cases} \quad (12)$$

上式可化为求解下面的矩阵方程:

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ e \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \rho \end{bmatrix} \quad (13)$$

$$\text{即: } \begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \rho \end{bmatrix}, \quad (14)$$

其中, $Z = [\varphi(x_1)^T y_1; \dots; \varphi(x_N)^T y_N]$, $Y = [y_1; \dots; y_N]$, $\rho = [1; \dots; 1]$, $e = [e_1; \dots; e_N]$, $a = [a_1; \dots; a_N]$ 。同样应用 Mercer 条件到 $\Omega = ZZ^T$ 中, 可得:

$$\Omega_{kl} = y_k y_l \varphi(x_k)^T \varphi(x_l) = y_k y_l \Psi(x_k, x_l) \quad (15)$$

因此, (10) 式的分类问题通过解决 (14) - (15) 式的线性方程获得, 而不用解决二次规划问题。核函数 $\Psi(x_k, x_l)$ 同样可以采用多项式核、多层感知 (MLP) 核、B 样条核、RBF 核等。

4 用于函数估计的最小二乘支持向量机

设训练样本集 $D = \{(x_k, y_k) | k=1, 2, \dots, N\}$, $x_k \in R^n$, $y_k \in R$, x_k 是输入数据, y_k 是输出数据。在权 w 空间 (原始空间) 中的函数估计问题可以描述求解下面问题:

$$\min_{w,b,e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (16)$$

约束条件: $y_k = w^T \varphi(x_k) + b + e_k, k=1, \dots, N$ 。

其中, $\varphi(\cdot): R^n \rightarrow R^h$ 是核空间映射函数, 权矢量 $w \in R^h$ (原始空间), 误差变量 $e_k \in R$, b 是偏差量。损失函数 J 是 SSE 误差和规则化量之和, γ 是可调常数。核空间映射函数的目的是从原始空间中抽取特征, 将原始空间中的样本映射为高维特征空间中的一个向量, 以解决原始空间中线性不可分的问题。根据 (16) 式, 可定义拉格朗日函数:

$$L(w, b, e; a) = J(w, e) - \sum_{k=1}^N a_k \{w^T \varphi(x_k) + b + e_k - y_k\} \quad (17)$$

其中, 拉格朗日乘子 $a_k \in R$ 。对上式进行优化, 即求 L 对 w, b, e, a_k 的偏导数等于 0, 消除变量 w, e , 可得以下矩阵方程:

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (18)$$

其中, $x = [x_1; \dots; x_N]$, $y = [y_1; \dots; y_N]$, $1 = [1; \dots; 1]$, $a = [a_1; \dots; a_N]$, $\Omega_{kl} = \Psi(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$, $k, l=1, \dots, N$ 。根据 Mercer 条件, 存在映射函数 φ 和核函数 $\Psi(\cdot, \cdot)$ 使得:

$$\Psi(x_k, x_l) = \varphi(x_k)^T \varphi(x_l) \quad (19)$$

LS-SVM 最小二乘支持向量机的函数估计为:

$$y(x) = \sum_{k=1}^N a_k \Psi(x, x_k) + b \quad (20)$$

其中, a, b 由 (18) 式求解出。核函数 $\Psi(x_k, x_l)$ 可采用多项式核、多层感知 (MLP) 核、B 样条核、RBF 核等。

5 最小二乘支持向量机分类实例

采用典型的分类测试样本集—螺旋样本对最小二乘支持向量机的学习和模式识别性能进行测试。核函数采用径向基核 $\Psi(x, x_k) = \exp\{-\|x - x_k\|^2 / \sigma^2\}$, σ 取 1, 式 (19) 中的 γ 取 1。计算结果见图 1。从运算结果可知: 采用最小二乘支持向量机进行的分类结果完全正确。

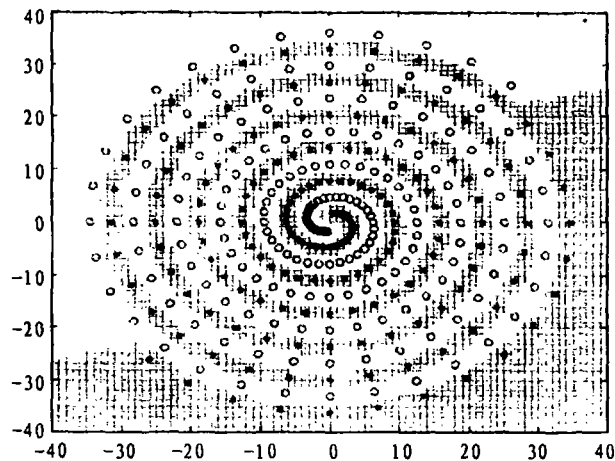


图1 采用 LS-SVM 对螺旋样本进行分类, 其中, 'o' 和 '*' 分别代表两类样本

等价闭包矩阵及其算法

闫德勤¹ 迟忠先²

(辽宁师范大学计算机系 大连116029)¹ (大连理工大学计算机系 大连116026)²

Equivalence Closure Matrix and its Algorithm

YAN De-Qin¹ CHI Zhong-Xian²

(Department of Computer, Liaoning Normal University, Dalian 116029)¹

(Department of computer, Dalian University of Technology, Dalian 116026)²

Abstract Properties of fuzzy equivalence matrix is studied, and the theorem about equivalence matrix is developed, with the theorem the mistake proof in a new algorithm for computing the fuzzy equivalence closure of a fuzzy similarity matrix is corrected.

Keywords Fuzzy clustering, Fuzzy equivalence matrix, Equivalence closure matrix

1 引言

在模糊聚类分析的研究与应用中,基于模糊关系等价闭包的模糊聚类算法,又称等价闭包法是一种重要的方法。等价闭包法即是利用样本间的模糊相似关系矩阵进行模糊矩阵相乘得到模糊等价矩阵进而得到等价闭包矩阵,选取适当的阈值对闭包矩阵截取得到一定的分类。该算法的关键问题就是计算出等价闭包矩阵。设 R 为模糊相似矩阵,其等价闭包矩阵由下式计算:

$$R^* = R \cup R^2 \cup \dots \cup R^{n-1} \quad (1)$$

计算(1)式的复杂度为 $O(n^5)$ ^[6]。当样本数较大时,等价闭包矩阵的计算非常费时。因此(1)式的加速算法受到关注^[2~7]。

最近, Lee^[6]提出了一种计算模糊相似矩阵等价闭包的优化算法。该算法对模糊相似矩阵的最大元素排队,利用求桥元素的方法得出等价闭包矩阵,其复杂度为 $O(n^2)$ 。这一结果是令人鼓舞的,但在文[6]中证明该算法一定产生等价闭包矩阵的过程存在着难以通过的错误环节。本文对等价闭包矩阵的算法以及等价矩阵的性质进行了研究,给出了相关定理,依据本文给出的定理给出了一种文[6]中算法的一种证明。

2 基本概念与定理

定义1 一个 $n \times n$ 的模糊矩阵 $A = (a_{ij})$, 若有 $a_{ii} = 1 (1 \leq i \leq n)$ 则称为自反的, 若 $a_{ij} = a_{ji}$ 则称为对称的。自反对称的模糊矩阵称为模糊相似矩阵。

定义2 若 $a_{ij} \geq \bigvee_k (a_{ik} \wedge a_{kj})$ 则称矩阵 $A = (a_{ij})$ 为传递的。

定义3 一个传递的相似的模糊矩阵称为模糊等价矩阵。

定义4 若对所有的 i, j 都有 $a_{ij} \leq b_{ij}$, 则称矩阵 B 包含矩阵 A , 记作 $A \subseteq B$ 。

定义5 若模糊等价矩阵 B 包含模糊相似矩阵 A , 且被所有包含矩阵 A 的模糊等价矩阵所包含, 则称 B 为 A 的等价闭包。

由定义可知相似矩阵的等价闭包即是等价闭包。一个矩阵的等价闭包即是包含该矩阵的最小模糊等价矩阵。

定理1 设 A 为模糊等价矩阵, P 为初等变换矩阵, 则 $B = PAP^T$ 仍为模糊等价矩阵。更进一步地有若 P 为初等变换矩阵之积, 则 B 仍为模糊等价矩阵。

证明: 首先矩阵 B 是对称的。在初等变换矩阵的作用下

闫德勤 博士, 副教授, 主要研究领域为模式识别、图像处理等。迟忠先 教授, 博士生导师, 主要研究领域为知识发现、数据库、数据挖掘等。

结论 最小二乘支持向量机采用最小二乘线性系统代替传统的支持向量机采用二次规划方法解决模式识别和函数估计, 本文给出了用于模式识别和函数估计的最小二乘支持向量机详细推理和算法。从最小二乘支持向量机对典型样本的学习和模式识别结果看, 采用最小二乘支持向量机进行模式识别, 运算简单, 收敛速度快, 精度高。

参考文献

- 1 Vapnik V N. The nature of statistical learning theory. Springer-Verlag, New-York, 1995
- 2 Vapnik V N. Statistical Learning Theory. John Wiley, New York, 1998
- 3 Suykens J A K, Vandewalle J, De Moor B. Optimal Control by Least Squares Support Vector Machines. Neural Networks,

2001, 14(1): 23~35

- 4 Zhu Jia-Yuan, Ren Bo, Zhang Heng-Xi, Deng Zhen-Ting. Time Series Prediction via New Support Vector Machines. IEEE. In: proc. of ICMLC'2002. China, Beijing. 364~366
- 5 Zhu Jia-Yuan, Zhang Heng-Xi, Guo Ji-Lian, Feng Jing-Lei. Data Distributions Automatic Identification Based on SOM and Support Vector Machines. IEEE. In: proc. of ICMLC'2002, China, Beijing. 340~344
- 6 张恒喜, 郭基联, 朱家元, 虞建飞著. 小样本多元数据分析方法及应用. 西北工业大学出版社, 2002
- 7 朱家元, 张恒喜. 基于支持向量机的 R&D 项目中止决策研究. 计算机科学, 2002, 29(9)
- 8 朱家元, 段宝君, 张恒喜. 新型 SVM 对时间序列预测研究. 计算机科学, 2003, 7