

基于特征可分性的核函数自适应构造

任双桥 魏玺章 黎 湘 庄钊文

(国防科技大学电子科学与工程学院空间电子信息技术研究所 长沙 410073)

摘 要 核函数的选择与构造是支撑向量机研究中的关键问题和难点. 该文针对该问题, 首先讨论了特征空间的线性可分性, 推导了其判别条件. 然后, 根据特征完全可分条件, 基于函数逼近论和核函数的基本性质, 提出了自适应多项式核函数和 B-样条核函数模型, 给出了模型参数的估计算法. 实测数据仿真实验结果表明, 与经典的核函数相比, 该文提出的算法在分类性能上取得了明显改善.

关键词 支撑向量机; 特征可分性; 核函数; 多项式; B-样条

中图法分类号 TP18

Adaptive Construction for Kernel Function Based on the Feature Discriminability

REN Shuang-Qiao WEI Xi-Zhang LI Xiang ZHUANG Zhao-Wen

(Institute of Spatial Electronic Information Technology, School of Electronic Science and Engineering,
National University of Defense Technology, Changsha 410073)

Abstract This paper firstly gives the distinguishable condition for separating the features space by linear classification hyper surface. Then the paper puts forward a novel adaptive construction algorithm for the kernel function, which is based on the feature distinguishable condition and the property of kernel function. This new kernel functions model includes the polynomial model and the B-Spline model. As the experimental results shown, validated with the actually measuring data, the performance of the new adaptive kernel function, such as classification capability and generalized capability are improved obviously in contrast to the classical kernel function.

Keywords SVM; feature discriminability; kernel function; polynomial; B-Spline

1 引 言

支撑向量机(SVM)^[1]是一种新的非常有发展前景的分类识别技术. SVM是建立在统计学习理论中结构风险最小化原理基础上, 根据有限的样本信息, 在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别样本的能力)之间寻求最佳折衷, 以期获得最好的推广能力. SVM克服了神经网络和传统分类器的过学习、局部极值点和维数灾难等诸多缺点, 具备较强的泛化能力, 已成

为备受关注的研究热点.

核函数是基于支撑向量机解决原样本空间中线性不可分问题时引入的一种非线性变换. 其基本思想就是通过核变换, 使得样本在新的高维特征空间中线性可分, 然后在高维特征空间中来实现最优分类面的求解. 核函数及核方法的研究是 SVM 理论与方法研究中极为重要的一个分支.

构造一个具有良好性能的支撑向量机, 模型选择是关键^[2]. 传统的做法包括两个相对独立的步骤, 首先凭经验选定核函数的类型, 其次根据一定的准则确定核函数的参数. 目前, 关于核函数的研究有很

多^[3-9],但是,大多数的方法是在特定的应用领域内,通过实验分析来解决,至今没有一个能够指导 SVM 选择最优核的理论和方法.核函数的选择与构造仍是一个亟待解决的难题.

本文针对该问题,首先讨论了样本空间中特征的线性可分性,推导了其判别条件.然后,根据特征的完全可分条件,基于函数逼近论和核函数的基本性质,提出了自适应构造核函数的两种模型,即自适应多项式核函数和自适应 B-样条核函数,给出了模型参数估计算法.实测数据仿真实验结果表明,与经典核函数相比,本文构造的核函数在分类性能和推广能力上均取得了明显改善.

2 特征空间中目标线性可分条件

本节主要探讨样本空间特征的可分性,首先给出了特征的可分性定义,然后依据 Farkas 引理^[1]推导了其判别条件.这些判别条件为自适应构造可分核函数提供了理论基础.

2.1 线性可分性定义

给定输入空间 l 个训练样本 $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中, $x_i \in R^n, y_i \in \{1, -1\}$ 为类别标示符.如果存在一个线性分类器能将每个样本正确分类,则称样本 D 线性可分.

定义 1. 称样本集 $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中, $x_i \in R^n, y_i \in \{1, -1\}$ 线性可分是指,存在 $w \in R^n, b \in R, \rho > 0$, 使得式(1)成立.

$$y_i(w^T x_i + b) \geq \rho, i = 1, 2, \dots, l \quad (1)$$

若不存在满足不等式(1)的 (w, b, ρ) , 则称样本集 D 线性不可分.

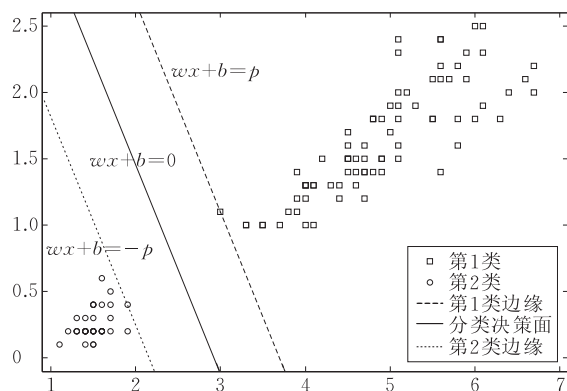


图 1 样本集线性可分示意图

为讨论方便,将式(1)写成矩阵形式,记 $\beta = (w, b, \rho) \in R^{n+2}, B = (0, \dots, 0, 1)^T \in R^{n+2}$, 矩阵

$$A = \begin{bmatrix} -y_1 x_1^T & -y_1 & 1 \\ \vdots & \vdots & \vdots \\ -y_l x_l^T & -y_l & 1 \end{bmatrix}$$

则有

$$A\beta \leq 0 \quad (2)$$

$$B^T \beta > 0 \quad (3)$$

在下述的讨论中, Farkas 引理扮演了一个重要的角色.

引理 1 (Farkas 择一定理)^[10]. 设矩阵 $A \in R^{m \times n}$, 向量 $B \in R^n$, 则不等式组 ① 和组 ② 恰一组有解.

$$\textcircled{1} A\beta \leq 0, B^T \beta > 0;$$

$$\textcircled{2} A^T \alpha = B, \alpha \geq 0.$$

证明. 参见文献[10]. 证略.

2.2 线性可分性条件

根据引理 1, 可从反面考察样本集 $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$ 的线性可分性, 也即, 若 D 线性不可分, 则不等式组 ① 无解, 从而其充要条件是不等式组 ② 有解. 若将式(2)、(3)代入不等式组 ②, 则可得

$$(y_1 x_1 \quad \dots \quad y_l x_l) \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_l \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4)$$

$$\sum_{i=1}^l \alpha_i = 1, \sum_{i=1}^l \alpha_i y_i = 0 \quad (5)$$

$$\alpha_i \geq 0, i = 1, 2, \dots, l \quad (6)$$

从而可得 $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$ 的线性可分性理论形式的判据.

定理 1 (可分性条件). 观测样本集 $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$ 线性可分的充要条件是不等式组(4)、(5)、(6)无解.

为了求解齐次线性方程(4), 需要用到矩阵的广义逆. 下述定理刻画了任意矩阵的全部广义逆.

定理 2^[11]. 设矩阵 $A \in R^{m \times n}$, 秩为 r , 且设

$$A = P \begin{pmatrix} E_{r \times r} & 0 \\ 0 & 0 \end{pmatrix} Q,$$

其中, P, Q 分别为 $m \times m$ 和 $n \times n$ 可逆矩阵, $E_{r \times r}$ 为 r 阶单位阵, 则 A 的全部广义逆为

$$G = Q^{-1} \begin{pmatrix} E_{r \times r} & U \\ F & H \end{pmatrix} P^{-1},$$

其中, U, F, H 分别为任意的 $r \times (m-r), (n-r) \times r$ 和 $(n-r) \times (m-r)$ 矩阵.

证明. 参见文献[11]. 证略.

定理 3^[11]. 设矩阵 $A \in R^{m \times n}, G$ 为任意给定的

广义逆, 则齐次线性方程组

$$A\alpha = 0$$

的全部解为

$$\alpha = (E_{n \times n} - GA)Z,$$

其中, Z 为任意的 n 维向量.

证明. 参见文献[11]. 证略.

根据定理 3, 记矩阵

$$H = (y_1 x_1 \cdots y_l x_l) \quad (7)$$

则不等式组(3)、(4)、(5)的解与下述不等式组同解

$$y^T (E_{l \times l} - GH)Z = 0 \quad (8)$$

$$e^T (E_{l \times l} - GH)Z = 1 \quad (9)$$

$$(E_{l \times l} - GH)Z \geq 0 \quad (10)$$

其中, $y = (y_1, \dots, y_l)^T$, $e = (1, \dots, 1)^T \in R^l$, $Z \in R^l$, G 为任意给定的矩阵 H 的广义逆. 由此可得样本集 $D = \{(x_i, y_i), i=1, 2, \dots, l\}$ 线性可分充要条件是不等式组(8)、(9)、(10)无解.

推论 1. 若矩阵 H 列满秩, 即秩(H)= l , 则样本集 $D = \{(x_i, y_i), i=1, 2, \dots, l\}$ 线性可分.

证明. 由于秩(H)= l , 故存在可逆矩阵 P, Q , 使得

$$H = P \begin{pmatrix} E_{l \times l} \\ 0 \end{pmatrix} Q \quad (11)$$

从而, 其广义逆矩阵

$$G = Q^{-1} (E_{l \times l} \quad 0) P^{-1} \quad (12)$$

齐次方程(4)的解为

$$\alpha = (E_{l \times l} - GH)Z = 0 \quad (13)$$

显然, 式(13)与(9)矛盾, 故不等式组(8)、(9)、(10)无解, 从而样本集 D 线性可分. 证毕.

若记 Gram 矩阵 $K = (k_{ij})_{l \times l}$, $k_{ij} = x_i^T x_j$, 则 $H^T H = Y^T K Y$, 其中

$$Y = \begin{bmatrix} y_1 & & \\ & \ddots & \\ & & y_l \end{bmatrix},$$

由于 $Y^T Y = E$, 因此, 可得一个更有意义的结论.

推论 2. 若 Gram 矩阵 K 正定, 则样本集 $D = \{(x_i, y_i), i=1, 2, \dots, l\}$ 线性可分.

证明. 由于 Gram 矩阵 K 正定, 且 $H^T H = Y^T K Y$, $Y^T Y = E$, 故矩阵 $H^T H$ 也是正定矩阵, 从而秩($H^T H$)= l . 又由于 $l = \text{秩}(H^T H) \leq \text{秩}(H) \leq l$.

故秩(H)= l , 由推论 1 可知, 样本集 $D = \{(x_i, y_i), i=1, 2, \dots, l\}$ 线性可分. 证毕.

根据推论 2 便可以考察经过核映射以后的特征是否线性可分, 这为自适应选择核函数、分析核函数可分性能提供了理论基础.

3 核函数自适应构造

支撑向量机的性能很大程度上取决于核函数的设计. 目前已有的支撑向量核, 由于其通过平移不能生成一组完备的基函数, 因此, 理论上支撑向量机是不能逼近平方可积空间中的任意函数的. 本节在对核函数可分性能的基础上, 依据推论 2 和核函数性质, 采用多项式和 B-Spline 自适应构造线性可分核函数. 首先, 介绍一个重要的定理.

定理 4^[3]. 对任意给定的有限个样本 x_1, x_2, \dots, x_l 以及某个核函数 K , 若其相应的 Gram 矩阵 $K = (k_{ij}) = (k(x_i, x_j))$ 是半正定的, 则可构造映射 ϕ , 使得 $K = \Phi^T(x_i) \Phi(x_j)$.

证明. 参见文献[3].

证略.

定理 4 表明, 对核函数 Mercer 条件^[1]的要求, 可以放宽到对有限个观测样本点值的要求, 也即只要能保证相应的 Gram 矩阵具有半正定性质. 这一重要性质为自适应构造可分核函数提供了一条重要的途径.

3.1 基于多项式表征核函数

不妨假设核函数的一般形式为 $K(u, v) = K(t)$, 其中, $t = u^T v$. 根据 Weierstrass 定理^[5], 闭区间上的连续函数可用多项式函数任意逼近. 因此, 可令

$$K(u, v) = K(t) = \sum_{i=1}^N a_i \frac{t^i}{i!} \quad (14)$$

由于常数对核函数的分类性能没有影响, 所以多项式模型(14)中没有常数项. 又由于 $u^T v$ 为核函数, 根据核函数性质可知, $(u^T v)^i$ 也为核函数. 因此, 为使式(14)为核函数, 只需多项式系数 $a_i \geq 0, i=1, 2, \dots, N$. 进一步, 若要求构造出来的核函数具备线性可分的能力, 由推论 2 和定理 4 可知, 只需 Gram 矩阵 $K = (k_{ij})_{l \times l}$, $k_{ij} = K(x_i, x_j)$ 正定即可. 因此, 可令

$$K(u, v) = K(t) = \sum_{i=1}^N a_i \frac{t^i}{i!} + c \delta(u - v) \quad (15)$$

其中, $a_i \geq 0, c > 0, i=1, 2, \dots, N$.

3.2 基于 B-样条表征核函数

由于多项式的逼近能力有限, 对于函数逼近问题, B-样条函数是一个有力的工具, 可以采用 B-样条函数来表征核函数. 假设 $t = u^T v$ 或者 $t = \|u - v\|$, 且 $a \leq t \leq b$. 给定一组等距划分, $\tau_0 = a < \tau_1 < \dots < \tau_M = b$, 步长 $h = \frac{b-a}{M}$, 节点 $\tau_i = \tau_0 + ih$. 记 $\tau_{-1} = \tau_0 - h$, $\tau_{M+1} = \tau_M + h$, 则有

$$\mathbf{K}(\mathbf{u}, \mathbf{v}) = \mathbf{K}(t) = \sum_{i=-1}^{M+1} b_i B_d\left(\frac{t - \tau_i}{h}\right) + c\delta(\mathbf{u} - \mathbf{v}) \quad (16)$$

其中, $B_d(\cdot)$ 为 d 阶标准 B-样条函数, 一般可取 $d=4$, 则可得

$$B_4(t) = \begin{cases} 0, & |t| \geq 2 \\ \frac{|t|^3}{2} - t^2 + \frac{2}{3}, & |t| \leq 1 \\ -\frac{|t|^3}{6} + t^2 - 2|t| + \frac{4}{3}, & \text{其它} \end{cases}$$

同理, 为保证式(16)为核函数, 由定理 4 可知, 只需其所对应的 Gram 矩阵正定则可. 因此, 可对模型式(16)的参数作如下要求.

$$\sum_{i=-1}^{M+1} b_i = 1, \quad c \geq c_0 > 0 \quad (17)$$

$$b_i \geq 0, \quad i = 1, 2, \dots, N \quad (18)$$

通过选择适当的 c_0 , 由 B-样条函数的归一化性质^[12], 式(17)和(18)总可以保证式(16)所对应的 Gram 矩阵正定. 一般的, c_0 的取值范围可定为 $0.5 \sim 1$.

3.3 核函数模型参数求解

下面考察多项式核函数模型(15)和 B-样条核函数模型(16)的参数求解. 不妨假设核函数 $\mathbf{K}(\mathbf{u}, \mathbf{v}) = \boldsymbol{\Phi}^T(\mathbf{u})\boldsymbol{\Phi}(\mathbf{v})$, 即原空间通过非线性变换 $\boldsymbol{\Phi}(\cdot)$ 映射到高维特征空间. 若记样本集 $D = \{(\mathbf{x}_i, y_i), i = 1, 2, \dots, l\}$ 中第一类样本点数为 l_1 , 第二类样本点数为 l_2 , $l_1 + l_2 = l$, 则在高维特征空间中, 两类样本点的半径分别为

$$r_1^2 = \frac{1}{l_1} \sum_{i=1}^{l_1} \|\boldsymbol{\Phi}(\mathbf{x}_i^1) - \overline{\boldsymbol{\Phi}(\mathbf{x}^1)}\|^2 \quad (19)$$

$$r_2^2 = \frac{1}{l_2} \sum_{j=1}^{l_2} \|\boldsymbol{\Phi}(\mathbf{x}_j^2) - \overline{\boldsymbol{\Phi}(\mathbf{x}^2)}\|^2 \quad (20)$$

其中, $\mathbf{x}_i^1, \mathbf{x}_j^2$ 分别表示样本集 D 中第一类和第二类样本点, $\overline{\boldsymbol{\Phi}(\mathbf{x}^1)}, \overline{\boldsymbol{\Phi}(\mathbf{x}^2)}$ 分别高维特征空间中两类样本点的中心.

$$\overline{\boldsymbol{\Phi}(\mathbf{x}^1)} = \frac{1}{l_1} \sum_{i=1}^{l_1} \boldsymbol{\Phi}(\mathbf{x}_i^1) \quad (21)$$

$$\overline{\boldsymbol{\Phi}(\mathbf{x}^2)} = \frac{1}{l_2} \sum_{j=1}^{l_2} \boldsymbol{\Phi}(\mathbf{x}_j^2) \quad (22)$$

将式(21)、(22)代入式(19)和式(20), 可得

$$r_1^2 = \frac{1}{l_1} \sum_{i=1}^{l_1} \boldsymbol{\Phi}(\mathbf{x}_i^1)^T \boldsymbol{\Phi}(\mathbf{x}_i^1) - \overline{\boldsymbol{\Phi}(\mathbf{x}^1)^T} \overline{\boldsymbol{\Phi}(\mathbf{x}^1)} = \frac{1}{l_1} \sum_{i=1}^{l_1} \mathbf{K}(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{1}{l_1^2} \sum_{i,j=1}^{l_1} \mathbf{K}(\mathbf{x}_i^1, \mathbf{x}_j^1) \quad (23)$$

$$r_2^2 = \frac{1}{l_2} \sum_{j=1}^{l_2} \boldsymbol{\Phi}(\mathbf{x}_j^2)^T \boldsymbol{\Phi}(\mathbf{x}_j^2) - \overline{\boldsymbol{\Phi}(\mathbf{x}^2)^T} \overline{\boldsymbol{\Phi}(\mathbf{x}^2)} = \frac{1}{l_2} \sum_{j=1}^{l_2} \mathbf{K}(\mathbf{x}_j^2, \mathbf{x}_j^2) - \frac{1}{l_2^2} \sum_{i,j=1}^{l_2} \mathbf{K}(\mathbf{x}_i^2, \mathbf{x}_j^2) \quad (24)$$

若记两样本集之间的平均距离为 r_{12} , 则有

$$\begin{aligned} r_{12}^2 &= \frac{1}{l_1 l_2} \sum_{j=1}^{l_2} \sum_{i=1}^{l_1} \|\boldsymbol{\Phi}(\mathbf{x}_i^1) - \boldsymbol{\Phi}(\mathbf{x}_j^2)\|^2 \\ &= \frac{1}{l_1 l_2} \sum_{j=1}^{l_2} \sum_{i=1}^{l_1} \boldsymbol{\Phi}(\mathbf{x}_i^1)^T \boldsymbol{\Phi}(\mathbf{x}_i^1) - 2 \boldsymbol{\Phi}(\mathbf{x}_i^1)^T \boldsymbol{\Phi}(\mathbf{x}_j^2) + \boldsymbol{\Phi}(\mathbf{x}_j^2)^T \boldsymbol{\Phi}(\mathbf{x}_j^2) \\ &= \frac{1}{l_1} \sum_{i=1}^{l_1} \mathbf{K}(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{2}{l_1 l_2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \mathbf{K}(\mathbf{x}_i^1, \mathbf{x}_j^2) + \frac{1}{l_2} \sum_{j=1}^{l_2} \mathbf{K}(\mathbf{x}_j^2, \mathbf{x}_j^2) \end{aligned} \quad (25)$$

又由多项式模型式(15)可知, 若假设向量 $\boldsymbol{\beta} = (a_1, a_2, \dots, a_N, c)^T$, $f(\mathbf{u}, \mathbf{v}) = \left(t, \dots, \frac{t^N}{N!}, \delta(\mathbf{u} - \mathbf{v})\right)^T, t = \mathbf{u}^T \mathbf{v}$, 则有

$$\mathbf{K}(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}, \mathbf{v})^T \boldsymbol{\beta} \quad (26)$$

同理, 对于 B-样条参数模型(15), 只需令

$$\boldsymbol{\beta} = (b_{-1}, b_{-2}, \dots, b_{M+1}, c)^T,$$

$$f(\mathbf{u}, \mathbf{v}) = \left(B_d\left(\frac{t - \tau_{-1}}{h}\right), \dots, B_d\left(\frac{t - \tau_{M+1}}{h}\right), \delta(\mathbf{u} - \mathbf{v})\right)^T, t = \mathbf{u}^T \mathbf{v} \text{ 或 } \|\mathbf{u} - \mathbf{v}\|, \text{ 则同样可得关于向量 } \boldsymbol{\beta} \text{ 的线性模型(26).}$$

将式(26)代入(23)、(24)、(25), 可得

$$r_1^2 = \left(\frac{1}{l_1} \sum_{i=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{1}{l_1^2} \sum_{i,j=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_j^1)\right)^T \boldsymbol{\beta},$$

$$r_2^2 = \left(\frac{1}{l_2} \sum_{j=1}^{l_2} f(\mathbf{x}_j^2, \mathbf{x}_j^2) - \frac{1}{l_2^2} \sum_{i,j=1}^{l_2} f(\mathbf{x}_i^2, \mathbf{x}_j^2)\right)^T \boldsymbol{\beta},$$

$$r_{12}^2 = \left(\frac{1}{l_1} \sum_{i=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{2}{l_1 l_2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} f(\mathbf{x}_i^1, \mathbf{x}_j^2) + \frac{1}{l_2} \sum_{j=1}^{l_2} f(\mathbf{x}_j^2, \mathbf{x}_j^2)\right)^T \boldsymbol{\beta},$$

若记

$$f_1 = \frac{1}{l_1} \sum_{i=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{1}{l_1^2} \sum_{i,j=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_j^1),$$

$$f_2 = \frac{1}{l_2} \sum_{j=1}^{l_2} f(\mathbf{x}_j^2, \mathbf{x}_j^2) - \frac{1}{l_2^2} \sum_{i,j=1}^{l_2} f(\mathbf{x}_i^2, \mathbf{x}_j^2),$$

$$f_{12} = \frac{1}{l_1} \sum_{i=1}^{l_1} f(\mathbf{x}_i^1, \mathbf{x}_i^1) - \frac{2}{l_1 l_2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} f(\mathbf{x}_i^1, \mathbf{x}_j^2) + \frac{1}{l_2} \sum_{j=1}^{l_2} f(\mathbf{x}_j^2, \mathbf{x}_j^2),$$

则有, $r_1^2 = \mathbf{f}_1^T \boldsymbol{\beta}$, $r_2^2 = \mathbf{f}_2^T \boldsymbol{\beta}$, $r_{12}^2 = \mathbf{f}_{12}^T \boldsymbol{\beta}$. 为保证核函数具有较好的推广性能, 一个直观的要求就是使 $r_{12}^2 - r_1^2 - r_2^2 \geq 0$ 的同时尽量大, 即模型参数 $\boldsymbol{\beta}$ 应满足

$$(\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_{12})^T \boldsymbol{\beta} \leq 0 \tag{27}$$

又由于 SVM 模型对偶形式^[4]的目标函数为最小化 $\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{S} \boldsymbol{\alpha}$, 其中, $\mathbf{S} = \mathbf{H}^T \mathbf{H} = \mathbf{Y}^T \mathbf{K} \mathbf{Y}$. 因此, 可通过参数 $\boldsymbol{\beta}$ 依范数达到最小来进一步降低最优化值. 从而, 核函数模型(15)和(16)的参数 $\boldsymbol{\beta}$ 可分别根据下述最优化问题求解.

$$\left\{ \begin{array}{l} \min \frac{1}{2} \sum_{i=1}^{N+1} \beta_i^2 + (\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_{12})^T \boldsymbol{\beta} \\ \text{s. t. } (\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_{12})^T \boldsymbol{\beta} \leq 0 \\ \beta_i \geq 0, i = 1, 2, \dots, N+1 \end{array} \right. \tag{28}$$

$$\left\{ \begin{array}{l} \min \frac{1}{2} \sum_{i=1}^{M+4} \beta_i^2 + (\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_{12})^T \boldsymbol{\beta} \\ \text{s. t. } (\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_{12})^T \boldsymbol{\beta} \leq 0 \\ \sum_{i=1}^{M+3} \beta_i = 1, \beta_{M+4} \geq c_0 \\ \beta_i \geq 0, i = 1, 2, \dots, M+3 \end{array} \right. \tag{29}$$

最优化模型(28)和(29)只是提供了一种较好的求解模型参数 $\boldsymbol{\beta}$ 的方法, 在实际工程应用和后续的研究工作中, 有待进一步考察和挖掘更为有效的参数求解方法以便最大限度提高核函数性能.

4 仿真实验及结果分析

本节将利用示例和外场雷达实测数据分别对特征可分性和核函数自适应构造算法进行有效性验证, 并对实验结果加以分析说明.

例 1. 证明异或问题(XOR)在原空间中线性不可分, 而选用二次多项式核函数 $\mathbf{K}(\mathbf{u}, \mathbf{v}) = (1 + \mathbf{u}^T \mathbf{v})^2$ 后, 则在此核函数导出的高维特征空间中线性可分.

证明. 假设样本集

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix},$$

其中, 矩阵 \mathbf{D} 的第 1 和第 2 行为 4 个二维样本点, 第 3 行为各自的类别标示符. 根据定理 1, 当取 $\boldsymbol{\alpha} = (0.25 \ 0.25 \ 0.25 \ 0.25)^T$ 时, 不等式组(4)、(5)、(6)成立, 因此, 样本集 \mathbf{D} 在原空间中线性不可分.

如果选用二次多项式 $\mathbf{K}(\mathbf{u}, \mathbf{v}) = (1 + \mathbf{u}^T \mathbf{v})^2$ 作为核函数, 则其 Gram 矩阵为

$$\mathbf{K} = \begin{bmatrix} 4 & 1 & 1 & 4 \\ 1 & 4 & 1 & 4 \\ 1 & 1 & 1 & 1 \\ 4 & 4 & 1 & 9 \end{bmatrix},$$

因为 Gram 矩阵 \mathbf{K} 的秩为 4, 故 \mathbf{K} 为正定矩阵. 从而, 由推论 2 可知, 样本集在二次多项式 $\mathbf{K}(\mathbf{u}, \mathbf{v}) = (1 + \mathbf{u}^T \mathbf{v})^2$ 导出的高维特征空间中线性可分的. 图 2 是样本集 \mathbf{D} 在二次多项式核函数下基于 SVM 的实验结果.

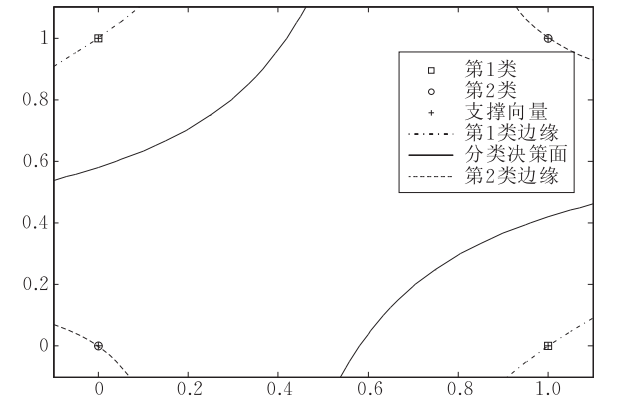


图 2 二阶多项式核函数分类结果

例 2. 本组数据为某机载雷达实测一维距离像数据集, 共有 3939 个样本, 样本维数为 32, 包括了 15 个角度下的测量值, 样本类型为卡车和坦克两种目标. 实验中选取的初始训练样本集为每个角度下 60 个样本, 总共选取了 900 个训练样本, 测试样本为 3039 个. 本次实验将两种自适应核函数模型与经典的多项式核函数和 Gaussian 核函数分别进行了测试率的对比, 实验结果如图 3、图 4 所示.

表 1 一维距离像数据集(900 个训练样本, 3039 个测试样本)

核函数类型	核参数设定	测试率/%
多项式核函数	$p=5$	88.23
	$p=9$	86.65
Gaussian 核函数	$\sigma=1$	81.58
	$\sigma=5$	80.86
自适应多项式核	$N=3$	87.56
	$N=5$	88.19
自适应 B-Spline 核 ($t=\mathbf{u}^T \mathbf{v}$)	$M=8, c_0=0.5$	98.03
	$M=10, c_0=1$	98.97

从表 1 可以看出, 自适应多项式核函数的测试率略高于 Gaussian 核函数而略低于经典的多项式核函数, 而自适应 B-样条核函数性能相比于经典的多项式核函数和 Gaussian 核函数都有了明显的提高. 这一方面是由于 B-样条较强的逼近能力, 同时也是由于模型参数 $\boldsymbol{\beta}$ 的求解算法是使得核函数具有

最大的推广能力.

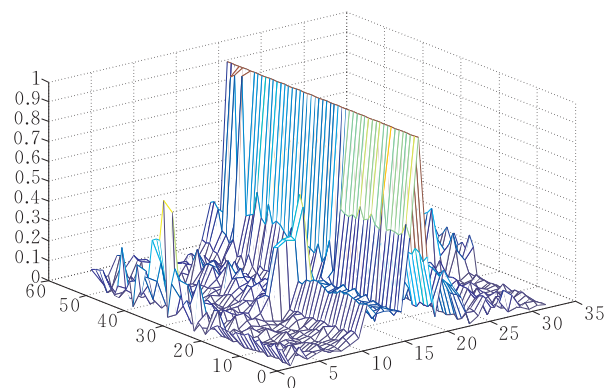


图 3 卡车的一维距离像

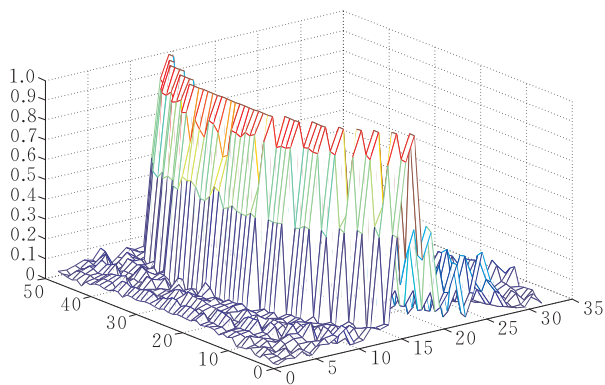


图 4 坦克的一维距离像

5 结束语

分类器的分类能力和推广性能往往是一对相互制衡的量. 好的分类器应能很好地兼顾这两方面的性能要求. 本文首先针对样本空间中的可分性, 给出了其判别条件. 然后, 根据特征的完全可分条件, 基于函数逼近论和核函数的基本性质, 提出了自适应构造核函数的两种模型, 即自适应多项式核函数和自适应 B-样条核函数, 并给出了模型参数估计算法. 总之, 在构造核函数时, 可以选用任何一种基函数, 所要满足的条件仅需其 Gram 矩阵正定, 而且模型的参数估计有待进一步研究以便最大限度提高核函数性能.

参 考 文 献

- [1] Vladimir N Vapnik. Xu Jian-Hua, Zhang Xue-Gong Translating. Statistical Learning Theory. Beijing: Publishing House of Electronics Industry, 2004(in Chinese)
- (Vladimir N Vapnik 著. 许建华, 张学工译. 统计学习理论. 北京: 电子工业出版社, 2004)
- [2] Wu Tao, He Han-Gen, He Ming-Ke. Interpolation based kernel function's construction. Chinese Journal of Computers, 2003, 26(8): 990-996(in Chinese)
- (吴涛, 贺汉根, 贺明科. 基于插值的核函数构造. 计算机学报, 2003, 26(8): 990-996)
- [3] Scholkopf B. Support Vector Learning. Berlin: Berlin University, 1997
- [4] Shun-ichi Amari, Si Wu. Improving support vector machine classifiers by modifying kernel functions. Neural Networks, 1999, 12(2): 783-789
- [5] Vapnik V, Chapelle O. Bounds on error expectation for support vector machines. Neural Computation, 2000, 12(9): 2013-2036
- [6] Müller K-R, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks, 2001, 12(2): 181-201
- [7] Ayat N E, Cheriet M, Remaki L et al. KOMD-a new support vector machine kernel with moderate decreasing for pattern recognition, application to digit image recognition//Proceedings of the 6th International Conference on Document Analysis and Recognition. Seattle, USA, 2001: 1215-1221
- [8] Schölkopf B, Sung K-K, Burges C J C, Girosi F. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing, 1997, 45(11): 2758-2765
- [9] Chapelle O, Vapnik V, Bacsquest et al. Choosing multiple parameters for support vector machines. Machine Learning, 2002, 46(1): 131-159
- [10] Su Ta-Shan et al. The Optimization Theory and Program Design. Changsha: Publishing House of National University of Defence Technology, 2001(in Chinese)
- (粟塔山等. 最优化计算原理与算法程序设计. 长沙: 国防科技大学出版社, 2001)
- [11] The Algebra Groups of Geometry and Algebra Staff Room of the Mathematics School in Peking University. Advanced Algebra. Beijing: Publishing House of Advance Education, 1995(in Chinese)
- (北京大学数学系几何与代数教研室代数小组. 高等代数. 北京: 高等教育出版社, 1995)
- [12] Wang Zheng-Ming et al. The Calibration and Evaluation of Trajectory Tracking Data. Changsha: Publishing House of National University of Defence Technology, 1999 (in Chinese)
- (王正明等. 弹道跟踪数据的校准与评估. 长沙: 国防科技大学出版社, 1999)



REN Shuang-Qiao, born in 1977, Ph. D. , assistant researcher. His research interests include machine learning, pattern recognition and radar signal processing.

WEI Xi-Zhang, born in 1976, Ph. D. , associate profes-

sor. His main research interests include pattern recognition and information fusion.

LI Xiang, born in 1967, Ph. D. , professor. His main research interests include pattern recognition, image processing and information fusion.

ZHUANG Zhao-Wen, born in 1958, Ph. D. , professor. His main research interests include pattern recognition, signal processing and information fusion.

Background

Support vector machines (SVM) is a powerful machines learning method based on VC dimension theory and structural risk minimization principle, which are the important foundation of statistical learning theory (SLT). The kernel function will seriously effect the classification and generalized performance of SVM, so it is a urgent difficult problem in SVM to adaptively construct the kernel function. The general method includes two steps, firstly, selecting the kernel function model by experience, then giving the parameters value of the kernel function by some rules. In the few years, most kernel function construction algorithms are established by analyzing the experimental results in some special engineering fields.

This paper puts forward a novel adaptive construction

algorithm for the kernel function, which is based on the feature distinguishable condition and the property of kernel function. This kernel functions model includes the polynomial and the B-Spline model, and the model parameters can be estimated by an optimization problem. Consequently, the new method can conveniently select the optimum kernel function according to the data.

This work is supported by the National Science Foundation of China, under grant No. 60402032. The aim of this project is to solve the automatic recognition for radar observing objects. This paper will offer the kernel function adaptively construction method, which is an important portion to the automatic objection recognition using SVM.