

文章编号: 1001-0920(2012)01-0077-05

基于最大化密度差的 L_2 核分类器

应文豪, 王士同

(江南大学 信息工程学院, 江苏 无锡 214122)

摘要: 提出一种基于最大化密度差的 L_2 核分类器算法 MDL2KC. 该算法不仅可以保证估计出的两类密度差接近于真实密度差, 而且可以使两类的密度差尽可能大. 利用人工数据集和标准 UCI 数据集进行实验验证, 所得结果表明, MDL2KC 算法较传统的 L_2 核分类器算法具有更好的分类效果和稀疏特性.

关键词: 分类算法; 稀疏; 密度差; 窗宽

中图分类号: TP391.4

文献标识码: A

L_2 kernel classifier based on maximum difference of densities

YING Wen-hao, WANG Shi-tong

(School of Information Engineering, Jiangnan University, Wuxi 214122, China. Correspondent: YING Wen-hao, E-mail: smartywh@163.com)

Abstract: This paper proposes a kernel classification algorithm(MDL2KC) based on the theory of maximum difference of densities. MDL2KC not only ensure the estimate difference of densities fairly close to the true difference of densities, but also maximize the difference of densities between two classes. As demonstrated by extensive experiments in artificial and UCI datasets, the proposed algorithm has better classification effect and sparsity than the traditional L_2 kernel classification algorithm.

Key words: kernel classification algorithm; sparsity; difference of densities; bandwidth

1 引言

在两类分类问题中, 假设已知一系列样本以及样本所属的分类, 并用 $(X_1, Y_1), \dots, (X_n, Y_n)$ 来表示. 其中: $X_i \in R^d$ 为一个 d 维样本, $Y_i \in \{-1, 1\}$ 为该样本的类标签. 分类的目的是在这些先验知识的基础上构造一个分类器, 将未知类别的样本输入该分类器后, 便可得到该样本的类别标签^[1]. 由于可以描绘非线性的判别边界, 核分类器近年来引起了广泛的关注. 一个核分类器可用下式表示^[2]:

$$g(x) = \text{sign} \left\{ \sum_{i=1}^n \alpha_i Y_i k(x, X_i) \right\}.$$

其中: α_i 是参数, k 是一个核函数. 未带补偿项的支持向量机便具有上述的形式^[3].

JooSeuk Kim 等人结合平方积分误差 (ISE)^[4-7] 准则训练核分类器的参数 α_i , 提出了 L_2 核分类器^[8-10]. L_2 核分类器算法关注于最优化样本的两类“密度差”和真实的“密度差”之间的平方积分误差^[11-13], 使样本的密度差接近于真实的密度差, 可得到较好的分类

性能. 但在两类样本概率密度较接近的情况下, 该方法存在一定的局限性.

基于上述观点, 本文提出一种改进的 L_2 核分类器算法——MDL2KC, 即基于最大化密度差的 L_2 核分类器. 该算法可以在两类的“密度差”接近于真实密度差的同时, 使两类的密度差更大. 根据 L_2 核分类器算法对决策函数的定义, 这种改进可以使决策函数的计算结果尽量远离决策分类面.

2 L_2 核分类器

2.1 L_2 核分类器的基本思想

从概率密度的角度, 定义 $f_+(x)$ 和 $f_-(x)$ 分别为正类样本的概率密度和负类样本的概率密度. 根据决策理论, 最优分类器具有如下形式:

$$g^*(x) = \text{sign}\{f_+(x) - \lambda f_-(x)\}. \quad (1)$$

这里定义概率密度差 $d_\lambda(x) = f_+(x) - \lambda f_-(x)$, λ 为人为设定的参数, 用来反映类别的先验概率. 根据前面所述, $Y_i \in \{-1, 1\}$ 是类标签, 这里重定义 $Y_i \in \{\lambda, 1\}$,

收稿日期: 2010-08-24; 修回日期: 2010-11-06.

基金项目: 国家自然科学基金项目(60903100, 60975027).

作者简介: 应文豪(1979—), 男, 讲师, 博士生, 从事模式识别、人工智能等研究; 王士同(1964—), 男, 教授, 博士生导师, 从事模式识别、人工智能等研究.

并定义 $I_+ = \{i|Y_i = +1\}$, $I_- = \{i|Y_i = -1\}$, 则根据核密度估计模型, 正类和负类样本的密度函数可分别表示为

$$\hat{f}_+(x; \alpha) = \sum_{i \in I_+} \alpha_i k_\sigma(x, X_i), \quad (2)$$

$$\hat{f}_-(x; \alpha) = \sum_{i \in I_-} \alpha_i k_\sigma(x, X_i). \quad (3)$$

其中 $\alpha = (\alpha_1, \dots, \alpha_n)$ 为可变参数, 并受如下约束制约 ($\alpha \in A$):

$$A = \left\{ \alpha \mid \sum_{i \in I_+} \alpha_i = \sum_{i \in I_-} \alpha_i = 1, \alpha_i \geq 0, \forall i \right\}. \quad (4)$$

$$k_\sigma(x, X_i) = (2\pi\sigma^2)^{-d/2} \exp\left\{-\frac{\|x - X_i\|^2}{2\sigma^2}\right\} \quad (5)$$

是一个窗宽 $\sigma > 0$ 的高斯核.

L_2 核分类器的核心思想是使用 $\hat{d}_\lambda(x) = \hat{f}_+(x) - \lambda\hat{f}_-(x)$ 来估计真实的 $d_\lambda(x)$, 并提出用最小化 $\hat{d}_\lambda(x)$ 和 $d_\lambda(x)$ 的 L_2 距离 (ISE) 估计出参数 α . ISE 定义如下:

$$\begin{aligned} \text{ISE}(\alpha) &= |\hat{d}_\lambda(x; \alpha) - d_\lambda(x)|_{L_2}^2 = \\ &= \int (\hat{d}_\lambda(x; \alpha) - d_\lambda(x))^2 dx = \\ &= \int \hat{d}_\lambda^2(x; \alpha) dx - 2 \int \hat{d}_\lambda(x; \alpha) d_\lambda(x) dx + \int d_\lambda^2(x) dx. \end{aligned} \quad (6)$$

因不知道真实的 $d_\lambda(x)$, 故需要估计式中的第 2 项. 设

$$H(\alpha) = \int \hat{d}_\lambda(x; \alpha) d_\lambda(x) dx, \quad (7)$$

则有

$$\text{ISE}(\alpha) = \int \hat{d}_\lambda^2(x; \alpha) dx - 2H(\alpha) + \int d_\lambda^2(x) dx. \quad (8)$$

JooSeuk Kimd 等人使用留一误差交叉校验的方法给出了 $H(\alpha)$ 的一个无偏估计, 即

$$H_n(\alpha) = \sum_{i=1}^n \alpha_i Y_i \hat{h}_i. \quad (9)$$

其中

$$\hat{h}_i = \begin{cases} \frac{1}{N_+ - 1} \sum_{j \in I_+, j \neq i} k_\sigma(X_j, X_i) - \\ \frac{\lambda}{N_-} \sum_{j \in I_-} k_\sigma(X_j, X_i), & i \in I_+; \\ \frac{1}{N_+} \sum_{j \in I_+} k_\sigma(X_j, X_i) - \\ \frac{\lambda}{N_- - 1} \sum_{j \in I_-, j \neq i} k_\sigma(X_j, X_i), & i \in I_-; \end{cases}$$

$$N_+ = |I_+|, N_- = |I_-|. \quad (10)$$

因此, 可以定义

$$\hat{\alpha} = \arg \min_{\alpha \in A} \text{ISE}(\alpha). \quad (11)$$

式(11)可使用标准二次规划方法求解.

最终 L_2 核分类器可表示为

$$g(x) = \begin{cases} +1, & \hat{d}_\lambda(x; \hat{\alpha}) \geq 0; \\ -1, & \hat{d}_\lambda(x; \hat{\alpha}) < 0. \end{cases} \quad (12)$$

2.2 L_2 核分类器算法分析

从上文可以看到, L_2 核分类器的核心思想是通过最小化 $\hat{d}_\lambda(x)$ 与 $d_\lambda(x)$ 的 L_2 距离 (ISE) 估计出参数 α , 并利用先验知识找到 ISE 的一个无偏估计; 然后利用标准二次规划方法进行求解. 然而, 当两类样本的概率密度较为接近时, 根据 L_2 核分类器对决策函数的定义 (12), 该类方法会使决策函数的计算结果靠近决策分类面, 从而影响分类的效果. 比较合理的一种设想是使样本密度差逼近于真实密度差的同时, 使两类样本的密度差尽量大, 这样可以使决策函数的计算结果尽量远离决策分类面. 在这种设想下, 本文研究并探讨了基于最大化密度差的 L_2 核分类器.

3 基于最大化密度差的 L_2 核分类器

概率密度差 $\hat{d}_\lambda(x) = \hat{f}_+(x) - \lambda\hat{f}_-(x)$ 用来反映两类概率密度的差异性, 对于分类问题, 希望 $\hat{d}_\lambda(x)$ 越大越好. 假定两类概率密度的最大间隔为 ρ , 则对于给定的样本 $x_i \in R^d$, 应满足

$$\hat{f}_+(x_i) - \lambda\hat{f}_-(x_i) \geq C_i \rho, \quad i \in I_+; \quad (13)$$

$$\hat{f}_+(x_i) - \lambda\hat{f}_-(x_i) \leq C_i \rho, \quad i \in I_-; \quad (14)$$

其中 $\rho \geq 0$. 由于下文中还要沿用 L_2 核分类器中对 Y_i 的定义, 此处定义类标签 $C_i \in \{-1, +1\}$. 基于上述思想, 本文在 ISE(α) 的基础上构造 MDL2KC 的一个目标函数, 最小化该目标函数可以估计出参数 α .

3.1 基于最大化密度差的 L_2 核分类器目标函数

MDL2KC 的目标函数可表示为

$$J_{\text{MDL2KC}} = |\hat{d}_\lambda(x; \alpha) - d_\lambda(x)|_{L_2}^2 - \nu\rho = \int (\hat{d}_\lambda(x; \alpha) - d_\lambda(x))^2 dx - \nu\rho. \quad (15)$$

其中: ρ 用来表示正类和负类的概率密度差, 常量参数 ν 用于平衡两项的贡献. 该目标函数共有 2 项: 第 1 项为 ISE(α), 用来逼近真实的概率密度差; 第 2 项用来控制两类的密度差, 使其足够大. 整个目标函数的含义是在样本密度差逼近于真实密度差的同时, 保证两类概率密度差达到最大.

分类器参数可通过最小化该目标函数求得, 即

$$\hat{\alpha} = \arg \min J_{\text{MDL2KC}}. \quad (16)$$

最终 MDL2KC 可表示为

$$g(x) = \begin{cases} +1, & \hat{d}_\lambda(x; \hat{\alpha}) \geq 0; \\ -1, & \hat{d}_\lambda(x; \hat{\alpha}) < 0. \end{cases} \quad (17)$$

MDL2KC 的核心问题是如何找到对 ρ 的一个估计, 从而通过优化目标函数来估计出参数 α .

3.2 ρ 的估计

定理 1 假设两类的概率密度有一个最大间隔 ρ , 则可以对 ρ 作如下估计:

$$\rho = \frac{1}{2} \sum_{i=1}^n C_i (d_\lambda(x_i) + \hat{d}_\lambda(x_i)) = \frac{1}{2} \left(\sum_{i=1}^{n_+} d_\lambda(x_i) + \hat{d}_\lambda(x_i) - \sum_{i=1}^{n_-} d_\lambda(x_i) + \hat{d}_\lambda(x_i) \right),$$

$$C_i \in \{-1, +1\}. \quad (18)$$

证明 在间隔 ρ 存在的情况下, 对于真实的 $d_\lambda(x_i)$, 若 $i \in I_+$, 则有 $d_\lambda(x_i) - C_i \rho > 0$, 因此对于观测的样本, 同样希望满足 $\hat{d}_\lambda(x_i) - C_i \rho > 0$; 同理, 若 $i \in I_-$, 则 $d_\lambda(x_i) - C_i \rho < 0$, 同样也希望满足 $\hat{d}_\lambda(x_i) - C_i \rho < 0$. 因此可以构建如下的目标函数:

$$J_\rho = \sum_{i=1}^n (d_\lambda(x_i) - C_i \rho)(\hat{d}_\lambda(x_i) - C_i \rho). \quad (19)$$

最大化该目标函数可以保证 $d_\lambda(x_i) - C_i \rho$ 和 $\hat{d}_\lambda(x_i) - C_i \rho$ 同号, 这两项同号可以保证参数的有效性以及分类的正确性.

根据拉氏定理, 要使目标函数(19)最大, 需使下式成立:

$$\frac{\partial J}{\partial \rho} = - \sum_{i=1}^n C_i (d_\lambda(x_i) + \hat{d}_\lambda(x_i)) + 2\rho = 0. \quad (20)$$

因此定理 1 得证. \square

结合式(6), 最终 MDL2KC 的目标函数可表示为

$$J_{\text{MDL2KC}} = |\hat{d}_\lambda(x; \alpha) - d_\lambda(x)|_{L_2}^2 - \nu \rho = \int (\hat{d}_\lambda(x; \alpha) - d_\lambda(x))^2 dx - \frac{\nu}{2} \left(\sum_{i=1}^{n_+} d_\lambda(x_i) + \hat{d}_\lambda(x_i) - \sum_{i=1}^{n_-} d_\lambda(x_i) + \hat{d}_\lambda(x_i) \right) = \int \hat{d}_\lambda^2(x; \alpha) dx - 2 \int \hat{d}_\lambda(x; \alpha) d_\lambda(x) dx + \int d_\lambda^2(x) dx - \frac{\nu}{2} \left(\sum_{i=1}^{n_+} \hat{d}_\lambda(x_i) - \sum_{i=1}^{n_-} \hat{d}_\lambda(x_i) \right) - \frac{\nu}{2} \left(\sum_{i=1}^{n_+} d_\lambda(x_i) - \sum_{i=1}^{n_-} d_\lambda(x_i) \right). \quad (21)$$

3.3 目标函数优化

目标函数(21)可表示为一个二次规划问题. 由于

$$\hat{d}_\lambda(x) = \hat{f}_+(x) - \lambda \hat{f}_-(x) = \sum_{i=1}^n \alpha_i Y_i k_\sigma(x, X_i), \quad (22)$$

根据高斯核的卷积定理^[14], 式(21)中第1项可表示为

$$\int \hat{d}_\lambda^2(x; \alpha) dx = \int \left(\sum_{i=1}^n \alpha_i Y_i k_\sigma(x, X_i) \right)^2 dx = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j Y_i Y_j \int k_\sigma(x, X_i) k_\sigma(x, X_j) dx =$$

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j Y_i Y_j k_{\sqrt{2}\sigma}(X_i, X_j);$$

第2项可表示为 $\sum_{i=1}^n \alpha_i c_i$, 其中 $c_i = Y_i \hat{h}_i$;

第4项可表示为

$$\frac{\nu}{2} \left(\sum_{j=1}^{n_+} \sum_{i=1}^n \alpha_i Y_i k_\sigma(X_j, X_i) - \sum_{j=1}^{n_-} \sum_{i=1}^n \alpha_i Y_i k_\sigma(X_j, X_i) \right);$$

第3项和第5项与 α 无关, 其中 n 为正类和负类样本总和. 因此该优化问题可转化为如下二次规划问题:

$$\hat{\alpha} = \arg \min_{\alpha \in A} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j Y_i Y_j k_{\sqrt{2}\sigma}(X_i, X_j) - \sum_{i=1}^n c_i \alpha_i - \frac{\nu}{2} \left(\sum_{j=1}^{n_+} \sum_{i=1}^n \alpha_i Y_i k_\sigma(X_j, X_i) - \sum_{j=1}^{n_-} \sum_{i=1}^n \alpha_i Y_i k_\sigma(X_j, X_i) \right). \quad (23)$$

3.4 参数讨论

在基于最大化密度差的 L_2 核分类器目标函数中, 有两个参数需要人为设定. 为了最小化概率误差, 根据文献[15]的定理4, 应将 λ 设为 $\lambda = (1-p)/p$, 其中 $0 < p < 1$ 为正类的先验概率. 然而在大多数情况下 p 是未知的, 因此本文采用 $\lambda^* = N_-/N_+$ 来估计 λ , 其中 N_+ 和 N_- 为正类和负类的样本个数. 另外, 常量参数 ν 用于平衡式(15)中两项的贡献, 显然 $\nu \geq 0$. 当 $\nu = 0$ 时, 基于最大间隔的 L_2 核分类器将退化为传统的 L_2 核分类器. 然而, ν 并非越大越好, 因为 ν 越大, 则越侧重于两类样本的密度差, 从而愈加忽略两类样本本身的概率密度分布, 使参数 α 无法正确反映两类样本的概率密度分布, 这将直接影响分类的结果.

4 实验研究

为了评估基于最大化密度差的 L_2 核分类器的性能, 本文将分别在人工数据集和 UCI 数据集上进行实验.

4.1 实验环境

实验中将对基于最大化密度差的 L_2 核分类器 (MDL2KC), 经典的 SVC 算法^[16]以及 L_2 核分类器算法进行比较, 其中 SVC 具有一个窗宽参数 σ , 这里采用 10 倍交叉验证策略在 $(0.01 \times 2^0, 0.01 \times 2^1, \dots, 0.01 \times 2^{10})$ 中找到最优的 σ . L_2 核分类器算法的参数 λ 按照 3.4 节讨论的方法设定, 窗宽参数 σ 采用 10 倍交叉验证策略在 $(0.01 \times 2^0, 0.01 \times 2^1, \dots, 0.01 \times 2^{10})$ 中找到最优. MDL2KC 的参数 λ 和 σ 的设定方法与 L_2 核分类器相同, 参数 ν 采用 10 倍交叉验证策略

在 $(0.01 \times 2^0, 0.01 \times 2^1, \dots, 0.01 \times 2^5)$ 中找到最优.

4.2 人造数据集

为了说明 MDL2KC 算法的相关特性以及分类的有效性, 本文构造了两类人工数据集用于实验.

4.2.1 一维数据集

一维样本点从区间 $[-5, 15]$ 中取得. 首先定义两类样本的概率密度为

$$f_+(x) = 0.2\Phi(x; 10, \sqrt{7}) + 0.8\Phi(x; 5, 1), \quad (24)$$

$$f_-(x) = 0.7\Phi(x; 0, \sqrt{1.5}) + 0.3\Phi(x; 12, \sqrt{6}). \quad (25)$$

其中 $\Phi(x; \mu, \sigma)$ 为均值 μ , 方差 σ^2 的高斯函数. 在区间 $[-5, 15]$ 中取 200 个样本点用来训练, 两类样本点可以有部分重叠. 各参数的设定方法及实验方法如 4.1 节所述, 两类概率密度以及选取的样本点区间如图 1 和图 2 所示.

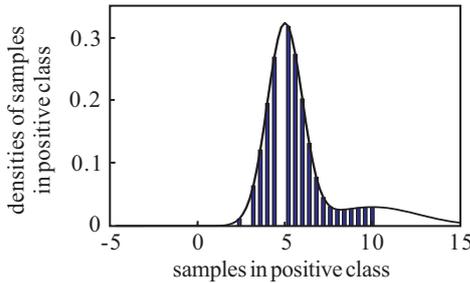


图 1 $f_+(x)$ 以及所取的正类样本区间的柱状图

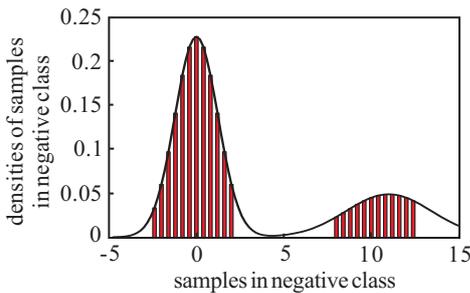


图 2 $f_-(x)$ 以及所取的负类样本区间的柱状图

实验结果如图 3 所示, 可以看到, MDL2KC 算法估计出的 $\hat{d}_\lambda(x; \hat{\alpha})$ 较接近于真实的 $d_\lambda(x)$. 特别需要指出的是, MDL2KC 算法在保证接近于真实密度差的同时, 使两类密度差尽量大 (见图 3), 这与上文的理论分析相一致.

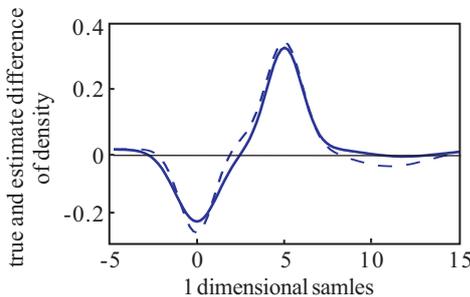


图 3 $d_\lambda(x)$ (实线) 和 $\hat{d}_\lambda(x; \alpha)$ (虚线)

4.2.2 香蕉型数据集

本文构造了一个两类的香蕉形数据集, 如图 4 所示, 标准差为 1, 并且正类和负类各 150 个样本. 本实验从正负类各随机选取 50 个样本作为测试样本, 其余作为训练样本. 将算法运行 20 次得到平均分类精度以及参数 α 的稀疏率, 稀疏率为参数 α 中不为 0 的参数所占参数总数的百分比. 在实验中当参数小于 10^{-10} 时, 它对分类结果的影响非常小, 可认为该参数为 0.

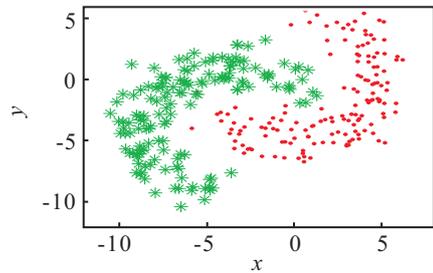


图 4 人造香蕉形数据

图 5 给出了参数 σ 以及 v 对分类精度的影响曲线, 图中 x 轴为 σ 的 11 个值, y 轴为分类精度. 可以看出, 当 $\sigma \in [0.01 \times 2^2 \sim 0.01 \times 2^7]$ 区间时, 可以获得较好的分类精度. 特别地, 如前所述, 当 v 值过大时, 分类精度将有所降低.

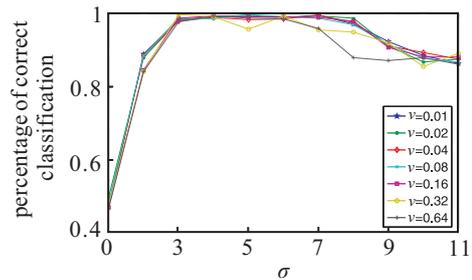


图 5 σ 和 v 对分类精度的影响曲线

图 6 给出了两个参数 σ 以及 v 对稀疏率的影响曲线. 从参数 σ 的角度观察, 随着 σ 的增大, 稀疏率对参数 σ 大致呈单调下降的趋势; 从参数 v 的角度观察, 随着 v 的增大, 稀疏率对参数 v 也大致呈单调下降的趋势. 因此, 通过上述实验可以看到, 算法能够进行有效的分类.

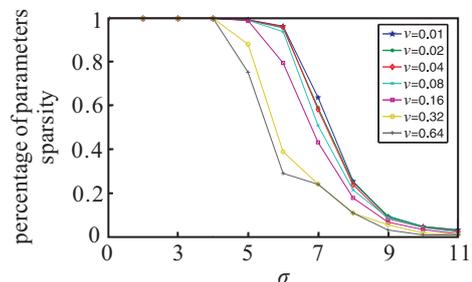


图 6 σ 和 v 对稀疏率的影响曲线

4.3 UCI 数据集

本实验所用的数据集来自于 UCI 数据库, 数据集的描述见表 1, 这些数据集均下载自 <http://homepage.tudelft.nl/n9d04/occ/index.html>. 本实验将在相同数据集下对比三类算法的分类正确率以及稀疏率. 三类算法的实验环境如 4.1 节所述.

表 1 8 个 UCI 数据集

数据集	样本数	特征点数
Breast Wisconsin	699	9
Balance-scale left	625	4
Vehicle bus	846	18
Waveform	900	21
Diabetes	768	8
Ecoli	336	7
Ionosphere	351	34
Delft pump	300	64

表 2 给出了各算法对 UCI 数据集运行 20 次之后得到的平均分类正确率. 由表 2 给出的数据可以看出, MDL2KC 算法较传统的 L_2 核分类器算法在分类正确率上有了全面的提高, 并且在大部分数据集上的正确率优于 SVC.

表 2 不同算法在 UCI 数据集上得到的分类正确率比较

数据集	SVC	L_2 核	MDL2KC
Breast Wisconsin	0.9673	0.9734	0.9795
Balance-scale left	0.9364	0.9250	0.9288
Vehicle bus	0.9373	0.9026	0.9059
Waveform	0.8542	0.8422	0.8567
Diabetes	0.7001	0.6608	0.6660
Ecoli	0.9142	0.9381	0.9512
Ionosphere	0.8811	0.9142	0.9142
Delft pump	0.9449	0.9704	0.9758

表 3 给出了各算法对 UCI 数据集运行 20 次之后得到的平均稀疏率. 稀疏率的大小直接影响到决策函数进行分类计算的复杂度^[17-18], 从而影响分类器的分类效率. 其中 SVC 的稀疏率为支撑向量个数占训练样本总数的百分比. 由表 3 的数据可以看到, 本文的 MDL2KC 算法具有稀疏性, 且稀疏性略优于传统的 L_2 核分类器算法.

表 3 不同算法在 UCI 数据集上得到的稀疏率比较

数据集	SVC	L_2 核	MDL2KC
Breast Wisconsin	0.1705	0.3555	0.3019
Balance-scale left	0.3398	0.8313	0.8223
Vehicle bus	0.3371	1.0000	1.0000
Waveform	0.5124	1.0000	1.0000
Diabetes	0.8134	1.0000	0.9980
Ecoli	0.4392	0.9800	0.0202
Ionosphere	0.3544	0.6641	0.7341
Delft pump	0.3853	1.0000	0.3695

根据上述分析, 本文提出的 MDL2KC 算法较传统的 L_2 核分类器算法具有更好的分类效果.

5 结 论

本文通过引入最大化概率密度差的思想, 提出了基于最大化密度差的 L_2 核分类器 (MDL2KC). 所提出的新算法的有效性在人工数据集和 UCI 数据集上均得到了验证. 关于 MDL2KC 参数选取的理论依据以及 ρ 是否有更合适的估计方式是进一步需探讨的课题.

参考文献(References)

- [1] Kim J, Scott C. Kernel classification via integrated squared error[C]. IEEE Workshop on Statistical Signal Processing. 2007: 783-787.
- [2] Schölkopf B, Smola A J. Learning with kernels[M]. Cambridge: MIT Press, 2002.
- [3] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [4] Mark Girolami, Chao He. Probability density estimation from optimally condensed data samples[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1253-1264.
- [5] Turlach B A. Bandwidth selection in kernel density estimation: A review[C]. CORE and Institut de Statistique, Universit'e Catholique de Louvain, 1993.
- [6] David W Scott. Parametric statistical modeling by minimum integrated square error[J]. Technometrics, 2001, 43: 274-285.
- [7] Bunea F, Tsybakov A B, Wegkamp M H. Sparse density estimation with l_1 penalties[C]. Proc of the 20th Annual Conf on Learning Theory. 2007, 4539: 530-543.
- [8] JooSeuk Kim, Clayton D Scott. Kernel classification via integrated squared error[C]. IEEE Workshop on Statistical Signal Processing. 2007: 783-787.
- [9] JooSeuk Kim, Clayton D Scot. Performance analysis for L_2 kernel classification[C]. Proc of Advances in Neural Information Processing Systems. 2008, 21: .
- [10] JooSeuk Kim, Clayton D Scott. L_2 kernel classification[J]. IEEE Trans on Pattern Analysis Mach Intell, 2010, 32(10): 1822-1831.
- [11] Peter Hall, Matthew P Wand. On nonparametric discrimination using density differences[J]. Biometrika, 1988, 75(3): 541-547.
- [12] Di Marzio M, Taylor C C. Kernel density classification and boosting: An L_2 analysis[J]. Statistics and Computing, 2005, 15: 113-123.
- [13] Meinicke P, Twellmann T, Ritter H. Discriminative densities from maximum contrast estimation[C]. Neural Information Proceeding Systems. Vancouver, 2002: 985-992.