

基于 Fisher 准则和最大熵原理的 SVM 核参数选择方法

周绍磊¹, 廖剑^{1,2}, 史贤俊¹

(1. 海军航空工程学院 控制工程系, 山东 烟台 264001; 2. 中国人民解放军 91550 部队, 辽宁 大连 116000)

摘要: 针对支持向量机(SVM)核参数选择困难的问题, 提出一种基于 Fisher 准则和最大熵原理的 SVM 核参数优选方法. 首先, 从 SVM 分类器原理出发, 提出 SVM 核参数优劣的衡量标准; 然后, 根据此标准利用 Fisher 准则来优选 SVM 核参数, 并引入最大熵原理进一步调整算法的优选性能. 整个模型采用粒子群优化算法(PSO)进行参数寻优. UCI 标准数据集实验表明了所提方法具有良好的参数选择效果, 优选出的核参数能够使 SVM 具有较高的泛化性能.

关键词: 支持向量机; 核函数; 参数选择; Fisher 准则; 最大熵原理; 粒子群优化算法

中图分类号: TP391

文献标志码: A

SVM parameters selection method based on Fisher criterion and maximum entropy principle

ZHOU Shao-lei¹, LIAO Jian^{1,2}, SHI Xian-jun¹

(1. Department of Control Engineering, Naval Aeronautical and Astronautical University, Yantai 264001, China; 2. Unit 91550 of PLA, Dalian 116000, China. Correspondent: LIAO Jian, E-mail: 251250544@qq.com)

Abstract: Aiming at the support vector machine(SVM) kernel parameters selection problem, an optimal selection method is proposed based on Fisher criterion and maximum entropy principle. A criterion for measuring the performance of the SVM kernel parameters is proposed by analyzing the classification principle of SVM. According to this criterion, the Fisher criterion is used to optimizing the SVM kernel parameters. Then, the maximum entropy principle is introduced to adjust the optimization performance of the algorithm. Finally, the whole model uses the particle swarm optimization(PSO) algorithm for parameters optimization. The experiment results on datasets from UCI show that the proposed method is more apt on parameter selection and can improve the generalization performance of SVM.

Key words: SVM; kernel function; parameter selection; Fisher criterion; MEP; PSO

0 引言

支持向量机(SVM)^[1]作为一种基于统计学习理论的机器学习方法, 克服了神经网络方法容易陷入局部最优、结构难于确定等问题, 在解决小样本、非线性及高维模式识别等问题中表现出结构简单、全局最优和泛化能力强等特点, 已成为计算智能领域的研究热点^[2-4]. SVM 的核心思想是通过引入核函数将低维空间中线性不可分的问题映射到高维空间中, 然后在空间构造最优分类超平面. 在不同核函数的作用下, 数据的分布结构不同, 从而直接影响 SVM 学习算法的性能. 尽管实验结果表明 SVM 核函数的具体形式对 SVM 分类效果的影响不大, 其核函数参数才是影响 SVM 性能的关键因素^[5], 但是目前就 SVM 核函数参数的选择也没有固定的方法, 这已成为制

约 SVM 应用于实际问题的“瓶颈”. 因此, SVM 核函数参数的选择一直是一个热点研究问题.

目前, 对于 SVM 核参数的选择主要有以下 3 种思路: 1) 采用梯度下降法或其他搜索算法在整个参数空间中进行搜索, 利用交叉验证法或留一法验证 SVM 泛化误差, 误差最小的参数即是最佳参数. 文献 [6] 对上述典型方法进行了概括, 并指出这类方法稳定性较高、误差估计准确, 其缺点是计算复杂程度高、计算量过大. 2) 简化 SVM 泛化误差估计上界, 训练时直接替代交叉验证法或留一法估计误差. 如文献 [7] 中的支持向量数上界; 文献 [8] 中介绍的几种估计上界方法 (Oppen-Winther 上界、Jaakkola-Haussler 上界、跨距上界等). 但这些方法不可避免地同样需要重复训练 SVM, 没有完全克服计算效率和计算精

收稿日期: 2013-08-03; 修回日期: 2013-11-08.

基金项目: 国家青年科学基金项目(61203168).

作者简介: 周绍磊(1963-), 男, 教授, 博士生导师, 从事模式识别与机器学习等研究; 廖剑(1985-), 男, 博士生, 从事模式识别与数据挖掘的研究.

度这一矛盾. 3) 寻找影响 SVM 分类器性能的各种因素, 构造目标函数, 然后利用各种搜索算法优化目标间接得到参数. 这种方法可大大减少 SVM 训练次数, 速度快, 因此得到广泛的研究^[5,9-12]. 文献 [9] 提出一种基于核相似性差异最大化的参数选择算法, 优化目标是使样本数据在核空间中方差最大; 文献 [10] 提出一种基于最佳相似核矩阵的参数选择算法, 优化目标是求一个与理想核矩阵最相近的矩阵. 文献 [5]、文献 [11-12] 都是基于最大间隔准则的核参数优化方法, 利用 Fisher 准则来衡量样本在特征空间中的类别可分性, 不同之处在于参数的寻优方法. 这类方法都是通过间接定义反映 SVM 泛化性能指标来求取最优核参数, 最大化减少 SVM 训练次数, 可以有效地减少计算量.

本文针对 SVM 核参数的选择问题, 提出一种基于 Fisher 准则和最大熵原理的核参数优选方法. 该算法利用 Fisher 准则衡量样本在核空间中的类别间隔, 同时引入信息熵进一步描述数据样本的类内分布性状, 并采用粒子群优化算法实现对 Fisher 准则函数和熵函数的同时最大化, 最后采用加权因子的方式来平衡这两类算法的分类效果. 与传统的基于 Fisher 准则的核参数选择方法相比, 本文方法不但考虑数据的类间绝对间隔, 而且注重类内数据的分布性状. 通过实验验证, 该方法具有良好的参数选择效果, 优选出的核参数能够使 SVM 具有更高的泛化性能.

1 支持向量机

SVM 是一种基于统计学习理论的机器学习算法, 建立在结构风险最小化原则之上. 用于分类的 SVM 原理如图 1 所示, 图中白点和黑点分别代表两类样本; H 为最优分类超平面; H_1 、 H_2 分别为过两类中离分类面最近的样本且平行于最优分类面的超平面, 它们之间的距离即为分类间隔.

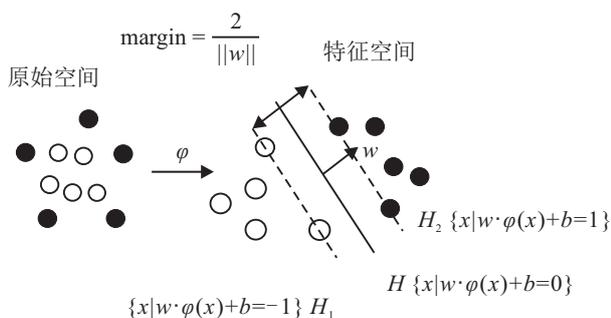


图 1 SVM 分类原理

SVM 的核心思想是通过引入核函数(隐性非线性映射), 将低维空间(图中原始空间)中线性不可分的问题映射到高维空间(图中特征空间)中, 然后在特征空间中建立最优分类超平面, 使两类数据点到超平面的距离最大.

1.1 支持向量机分类算法

设样本集为 $\{(x_i, y_i)\}_{i=1}^l$, $x_i \in R^d$ 为特征样本, $y_i \in \{1, -1\}$ 为输入样本 x_i 的类别标号, l 为样本个数. 首先通过核函数的非线性映射将输入数据映射到高维特征空间; 然后在特征空间中构造最优分类超平面 $w \cdot \varphi(x) + b = 0$, $w = \sum_{i=1}^l \alpha_i y_i \varphi(x_i)$ 为法向量, $b = y_j - \sum_{i=1}^l \alpha_i y_i K(x_i, x_j)$ 为偏置, α_i 通过求解最小化 $\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$ 的约束规划的对偶形式得到^[1], 如下式所示:

$$\begin{aligned} \max W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j), \\ \text{s.t. } \sum_{i=1}^l \alpha_i y_i &= 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l. \end{aligned} \quad (1)$$

其中: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)$ 为 Lagrange 乘子, α 中不为零的系数称为支持向量; $K(x_i, x_j)$ 为核函数; C 为惩罚系数. 则求解式 (1) 得到最优分类函数为

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right\}. \quad (2)$$

常用的核函数有多项式核函数 $K(x_i, x_j) = ((x_i \cdot x_j) + 1)^q$, 高斯径向基核函数 (RBF) $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/h^2)$, Sigmoid 核函数 $K(x_i, x_j) = \tanh(a(x_i \cdot x_j) + b)$. 在这 3 种核函数中, 以 RBF 核函数在 SVM 中的应用最广, 是一个普适的核函数^[13].

2 基于 Fisher 准则和最大熵原理的 SVM 核参数选择算法

2.1 核空间中的 Fisher 准则

Fisher 准则的基本原理是要找到一个最佳的投影方向, 使两类样本在该方向轴上投影的交迭部分最少, 从而使分类效果达到最佳. 即保证投影之后样本在新的空间中异类数据之间的距离尽可能大, 而同类数据的距离尽可能小, 其基本思想与 SVM 方法“不谋而合”, 因此可以借鉴 Fisher 准则来寻找核参数.

将原始数据经过非线性映射到特征空间之后, 数据的变换和计算都必须在特征空间中进行, 因此需要推导 Fisher 准则在核空间中的具体形式, 如下所示.

设 $X_1(x_{11}, x_{12}, \dots, x_{1i}), X_2(x_{21}, x_{22}, \dots, x_{2j})$ ($i = 1, 2, \dots, l_1, j = 1, 2, \dots, l_2$) 是原始数据空间中的两类样本, 两类样本在特征空间中的均值向量分别为

$$\mu_1 = \frac{1}{l_1} \sum_{i=1}^{l_1} \varphi(x_{1i}), \quad (3)$$

$$\mu_2 = \frac{1}{l_2} \sum_{i=1}^{l_2} \varphi(x_{2i}). \quad (4)$$

其中 $\varphi(\cdot)$ 为核函数 $K(\cdot, \cdot)$ 对应的非线性映射.

样本类间散布矩阵为

$$S_b = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \frac{1}{l_1^2} K_{11} + \frac{1}{l_2^2} K_{22} - \frac{2}{l_1 l_2} K_{12}; \quad (5)$$

样本类内散布矩阵为

$$S_{w1} = \sum_{i=1}^{l_1} \|\varphi(x_{1i}) - \boldsymbol{\mu}_1\|^2 = K_1 - \frac{1}{l_1} K_{11}, \quad (6)$$

$$S_{w2} = \sum_{i=1}^{l_2} \|\varphi(x_{2i}) - \boldsymbol{\mu}_2\|^2 = K_2 - \frac{1}{l_2} K_{22}. \quad (7)$$

其中

$$\begin{aligned} K_1 &= \sum_{i=1}^{l_1} K(x_{1i}, x_{1i}), \\ K_2 &= \sum_{i=1}^{l_2} K(x_{2i}, x_{2i}), \\ K_{11} &= \sum_{i=1}^{l_1} \sum_{j=1}^{l_1} K(x_{1i}, x_{1j}), \\ K_{22} &= \sum_{i=1}^{l_2} \sum_{j=1}^{l_2} K(x_{2i}, x_{2j}), \\ K_{12} &= \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} K(x_{1i}, x_{2j}). \end{aligned}$$

由 Fisher 准则, 最优核参数 κ 为

$$\kappa_1^* = \arg \max \left(\frac{\text{tr}(S_b)}{\text{tr}(S_{w1} + S_{w2})} \right). \quad (8)$$

2.2 核空间中的最大熵原理

在信息论中, 熵^[4]用于描述不确定性, 熵越大不确定性越大. 由熵的性质可知, 熵表示任何一种能量在空间中分布的均匀程度, 能量分布得越均匀, 熵就越大. 当某一系统的能量完全均匀地分布时, 系统的熵达到最大值. 因此, 本文利用熵反映类内样本的分布性状, 熵越大表示类内样本分布越均匀. 同 Fisher 准则, 熵的计算也必须在特征空间中进行, 因此, 根据信息论中有关熵的定义, 定义两类样本的分布熵分别为

$$E_1 = - \sum_{i=1}^{l_1} \frac{\|\boldsymbol{\mu}_1\|}{\|\varphi(x_{1i})\|} \log \frac{\|\boldsymbol{\mu}_1\|}{\|\varphi(x_{1i})\|} = - \sum_{i=1}^{l_1} \frac{K_{1i}}{l_1 K} \log \frac{K_{1i}}{l_1 K}, \quad (9)$$

$$E_2 = - \sum_{i=1}^{l_2} \frac{\|\boldsymbol{\mu}_2\|}{\|\varphi(x_{2i})\|} \log \frac{\|\boldsymbol{\mu}_2\|}{\|\varphi(x_{2i})\|} = - \sum_{i=1}^{l_2} \frac{K_{2i}}{l_2 K} \log \frac{K_{2i}}{l_2 K}. \quad (10)$$

其中

$$K = K(x_{*i}, x_{*i}), \quad * = 1, 2,$$

$$K_{1i} = \sum_{j=1}^{l_1} K(x_{1i}, x_{1j}),$$

$$K_{2i} = \sum_{j=1}^{l_2} K(x_{2i}, x_{2j}).$$

两类样本的分布熵之和为

$$E = E_1 + E_2. \quad (11)$$

由最大熵原理, 最优核参数 κ 为

$$\begin{aligned} \kappa_2^* &= \\ \arg \max(E) &= \arg \max(E_1 + E_2). \end{aligned} \quad (12)$$

2.3 本文方法的数学模型

Fisher 准则强调的是类间样本间隔的最大和类内样本的紧凑程度, 并不考察映射之后样本的类内分布状况, 而熵定理则能表示类内样本的均匀分布程度. 如果能同时考虑数据在非线性映射下的类间间隔和类内样本分布性状, 则有可能取得更好的核参数优化效果, 所以本文提出基于 Fisher 准则和最大熵原理的 SVM 核参数选择方法. 该方法的数学模型为

$$\begin{aligned} \kappa^* &= \lambda \kappa_1^* + (1 - \lambda) \kappa_2^* = \\ \lambda \times \arg \max \left(\frac{\text{tr}(S_b)}{\text{tr}(S_{w1} + S_{w2})} \right) &+ \\ (1 - \lambda) \times \arg \max(E_1 + E_2), \end{aligned} \quad (13)$$

其中 λ 为权重系数.

权重由类别样本内的相似度一致性确定. 如果类别样本的类内方差越大, 则样本相似度越小, 类别差异性越好, 也越有利于分类, 所以权值越大.

设使用 Fisher 准则计算的最优核参数为 κ_1^* , 将两类样本映射到高维空间之后, 两类样本的方差分布为 σ_1^2 和 σ_2^2 , 类内方差计算方法分别为

$$\sigma_1^2 = \frac{1}{l_1} S_{w1}, \quad (14)$$

$$\sigma_2^2 = \frac{1}{l_2} S_{w2}. \quad (15)$$

同理, 可以计算出使用最大熵原理计算的最优核参数为 κ_2^* , 将两类样本映射到高维空间之后, 两类样本的方差分别为 σ_3^2 和 σ_4^2 , 计算公式同样如式 (14) 和 (15) 所示. 则使用基于 Fisher 准则方法计算的最优核参数对应的权重为

$$\lambda = \frac{\sigma_1^2 + \sigma_2^2}{(\sigma_1^2 + \sigma_2^2) + (\sigma_3^2 + \sigma_4^2)}. \quad (16)$$

使用最大熵原理计算的最优核参数对应的权重为 $1 - \lambda$. λ 的取值介于 $0 \sim 1$ 之间, 当 $\lambda = 1$ 时为基于 Fisher 准则方法. 由于该算法引入了权重因子 λ 使得算法变得可调, 增强了算法的灵活性, 使算法具有更

广的适应性,同时也增强了算法的优选精度和稳定性.

2.4 使用 PSO 求解最优核参数

要求得最优核参数就必须求式(8)和(12)的最大值.文献[5]中采用梯度下降法求解,该方法需要计算类内离散度矩阵的逆矩阵,求解过程复杂,容易陷入局部最优,且当其逆矩阵不存在时将无法直接求解.因此本文考虑采用粒子群算法优化式(8)和(12),求得最优核参数.

粒子群优化算法(PSO)最早是由 Kennedy 等^[15]于 1995 年提出的. PSO 作为一种基于群体智能原理的全局并行优化方法,算法操作简便,容易实现,特别适合工程应用.由于篇幅所限,详细介绍请参见文献[15].

在 PSO 中,适应度函数决定了粒子的搜索方式和最终解的形式.本文方法为求最大类间间隔和类内最大熵,因此,适应度函数采用 Fisher 准则和最大熵原理,即式(8)和(12)的形式. κ^* 为粒子种群所要搜索的最优解. PSO 搜索 Fisher 准则和最大熵原理的最优核参数实现步骤如下.

Step 1: 初始粒子群. 随机初始化粒子群中各个粒子的起始位置和起始速度,构成初始候选解集.

Step 2: PSO 迭代开始. 计算各粒子式(8)或(12)的适应度函数值.

Step 3: 更新粒子个体极值. 比较每个粒子的当前适应度 $fitness(t)$ 和迭代前的个体极值 $fitnessBest(t-1)$, 若粒子的当前适应度优于迭代前的个体极值,则更新个体极值,否则保留原个体极值.

Step 4: 更新全局极值. 当前所有个体极值中适应度最优的个体极值存储为全局极值.

Step 5: 更新粒子下一时刻粒子速度和位置. 通过下式更新粒子的速度和位置:

$$\begin{aligned} v_{ij}(t+1) &= \\ v_{ij}(t) + c_1 r_1 [p_{ij} - x_{ij}(t)] + c_2 r_2 [p_{gj} - x_{ij}(t)], \\ x_{ij}(t+1) &= x_{ij}(t) + v_{ij}(t+1), \end{aligned} \quad j = 1, 2, \dots, d. \quad (17)$$

其中: v 和 x 分别为粒子的速度和位置; w 为惯性权因子; c_1, c_2 为正的学习因子; r_1, r_2 为 0 到 1 之间均匀分布的随机数; d 为粒子维数; t 为当前迭代步数.

Step 6: 更新惯性权重系数. 为了增强算法的全局搜索能力,算法在开始时采用较大的惯性权重系数,而在后期为了增强算法的区域搜索能力,采用较小的惯性权重系数. 本文采用如下式所示的线性递减权重变化公式:

$$w = \frac{w_{\max} - (w_{\max} - w_{\min})t}{t_{\max}}. \quad (18)$$

其中: w_{\max} 和 w_{\min} 分别表示 w 的最大值和最小值, t 表示当前迭代步数, t_{\max} 表示最大迭代步数.

Step 7: 判断是否满足迭代停止条件(为预设的运算精度或迭代次数). “是”转到 Step 8, 否则返回 Step 2 继续搜索.

Step 8: 全局极值对应的粒子为种群的最优解,即 Fisher 准则和最大熵原理得到的最优核参数 κ_1^* 和 κ_2^* .

Step 9: 根据式(14)和(15)分别计算两个方法下的类内样本方差,利用式(16)得到两个方法下的权重 λ 和 $1 - \lambda$, 最后代入式(13)得到 SVM 最优核参数.

3 实验分析

为了验证基于 Fisher 准则和最大熵原理的 SVM 核参数优化方法及其有效性,使用几个 UCI 机器标准学习数据集进行实验分析. 实验中采用 RBF 核函数作为基本核函数,具体形式如 1.1 节所示,并与交叉验证法和传统的基于 Fisher 准则的核参数选择方法进行对比. 为使实验具有代表性,在相同的软硬件平台下进行实验.

3.1 实验数据

实验所用样本采用 UCI 机器学习数据库中的 8 个两类分类数据集,实验中所采用的所有数据均可从网址: <http://www.ics.uci.edu/~mlearn/MLRepository.html> 中下载. 实验过程中,分别随机选取数据集的 40% 作为训练数据集、60% 为测试数据集. 表 1 为所用数据集的基本情况.

表 1 数据集基本情况

UCI 数据集	总样本数	特征维数
Breast-Cancer	277	9
Ionosphere	351	34
Sonar	208	60
Heart-c	296	13
Promoters	106	57
Diabetes	768	8
Clean 1	476	166
Heart-statlog	270	13

3.2 实验结果与分析

利用 SVM 进行模式分类时需要确定两个参数: 惩罚因子 C 和 RBF 核参数 h . 本文采用“试凑法”确定惩罚因子 C , 即从指数序列 $\{2^{-1}, 2^0, \dots, 2^{10}\}$ 中进行选择,核参数 h 采用本文提出的基于 Fisher 准则和最大熵原理的方法来确定. 首先利用 PSO, 根据 UCI 的训练数据集分别得到 RBF 核函数在 Fisher 准则和最大熵原理下的最优核参数 h_1^* 和 h_2^* , 并在相应的最优核参数下求得两类数据的类内方差分别为 σ_1^2, σ_2^2 和 σ_3^2, σ_4^2 , 然后根据式(13)和(16)得到最优核参数 h_1^* , 实验结果如表 2 所示.

表 2 两种方法下的最优核参数和相应的类内方差

UCI 数据集	Fisher 准则方法			最大熵原理方法			本文方法	
	h_1^*	σ_1^2	σ_2^2	h_2^*	σ_3^2	σ_4^2	h^*	λ
Breast-Cancer	0.98	0.52	0.4	1.15	0.13	0.15	1.02	0.77
Ionosphere	1.71	0.58	0.68	2.14	0.19	0.21	1.82	0.76
Sonar	8.43	0.21	0.15	10.55	0.11	0.09	9.24	0.64
Heart-c	11.18	0.37	0.56	8.18	0.26	0.22	10.26	0.66
Promoters	18.26	0.8	0.64	22.36	0.45	0.51	20.1	0.6
Diabetes	4.88	0.26	0.32	5.21	0.18	0.15	5	0.64
Clean 1	2.58	0.67	0.42	3.16	0.57	0.56	2.89	0.49
Heart-statlog	10.1	0.45	0.28	11.19	0.26	0.21	10.54	0.61

表 3 本文方法与其他模型评估准则的分类结果对比

UCI 数据集	交叉验证法			Fisher 准则方法			本文方法		
	C	h	正确率 %	C	h	正确率 %	C	h	正确率 %
Breast-Cancer	8	1.1	84.5	8	0.98	80.1	16	1.02	83.3
Ionosphere	128	1.9	96	16	1.71	95.2	32	1.82	96.9
Sonar	64	9.2	94.4	32	8.43	92.8	32	9.24	93.6
Heart-c	64	10.5	86.8	512	11.18	83.1	1024	10.26	84.5
Promoters	32	21.2	92.5	64	18.26	91.4	32	20.1	93.4
Diabetes	64	4.6	77.6	128	4.88	77.5	32	5	77.5
Clean1	1024	2.7	94.1	128	2.58	93.2	128	2.89	94.8
Heart-statlog	64	11.3	83.3	64	10.1	82.1	32	10.54	84.4

表 2 列出了每一种方法单独作用时取得的最优核参数和相应的类内方差, 最后两列则是采用式 (13) 求得的加权平均值及其相应权重. 从表 2 可以看出: 基于 Fisher 准则的方法求得的两类类内方差较大, 说明经过核映射之后虽然取得了最大的类间间隔, 但样本“参差不齐”的情况较严重; 采用基于最大熵原理的方法求得的两类类内方差明显小于基于 Fisher 准则的方法, 说明基于最大熵原理的方法使得映射之后样本的一致性较好. 如果能兼顾二者, 极有可能取得更好的优化效果. 最后采用表 2 中的最优核参数建立 SVM 对测试样本进行测试, 相应的实验结果如表 3 所示.

为了对比本文算法的效果, 表 3 还列出了采用网格搜索的交叉验证法、基于传统的 Fisher 准则方法^[5]和本文方法取得的核参数以及相应的核参数训练的 SVM 对测试数据集测试得到的正确率数据. 值得一提的是, 表 3 中交叉验证法的结果是采用 10 折交叉验证法获得的, 优化的结果包括惩罚因子 C 和核参数 h . 惩罚因子 C 的取值同样采用“试凑法”从指数序列 $\{2^{-1}, 2^0, \dots, 2^{10}\}$ 中进行选择, 核参数 h 在 $0 \sim 25$ 之间按照间隔 0.1 等差增长的方式进行扫描选取.

从表 3 可以看出: 本文方法取得了与交叉验证法相当的效果, 甚至在某些数据集上的正确率高于交叉验证法, 但本文方法在优选参数时只需重复训练 SVM 11 次, 而交叉验证法为了获得最佳的参数组合

需要重复训练 SVM $250 \times 11 = 2750$ 次, 可见本文算法效率明显高于交叉验证法; 基于 Fisher 准则的方法取得了最差的分分类效果, 在所有数据集上取得的分类正确率都不如本文方法, 这验证了 2.3 节中所分析的结论. 因此, 本文方法兼顾类别间隔最大原则和类内样本分布性状, 能取得令人满意的结果.

4 结 论

与其他机器学习算法一样, SVM 在用于模式识别时, 首先面临的问题是如何选择核函数参数以得到较好的分类性能. 针对这一问题, 本文提出了一种基于 Fisher 准则和最大熵原理的 SVM 核参数选择方法. 本文从分析 SVM 分类器的原理出发, 指出基于 SVM 分类的映射思想与 Fisher 准则的投影思想是一致的, 都是期望样本在特征空间中的类别间隔最大, 而类内紧凑程度最好, 因此可以借鉴 Fisher 准则来优化得到核参数. 但 Fisher 准则关注的只是投影之后类别的线性可分性, 并不考察映射之后样本的类内分布性状. 基于此, 本文进一步引入信息熵的概念来描述数据样本的类内分布性状, 采用加权因子的方式调整算法的优选性能, 且整个模型采用粒子群算法进行寻优, 避免了复杂的矩阵求逆过程, 算法简单, 容易实现. 针对 8 个 UCI 标准数据集的实验结果验证了本文参数选择方法的有效性, 其参数选择效果优于传统的基于 Fisher 准则方法, 优选出的核参数能够使 SVM 具有较高的泛化性能. 需要指出的是, 本文方法并不局限

于 RBF 核函数中的参数选择, 同样可用于其他核函数类型的参数优化中.

参考文献(References)

- [1] Vapnik V. The nature of statistical learning theory[M]. New York: Springer Verlag, 1995: 101-122.
- [2] Bradford J R, West Head D R. Improved prediction of protein binding sites using a support vector machines approach[J]. *Bioinformatics*, 2005, 21(8): 1487-1494.
- [3] Tao Q, Chu D J, Wang J. Recursive support vector machines for dimensionality reduction[J]. *IEEE Trans on Neural Networks*, 2008, 19(1): 189-193.
- [4] 孙健, 王成华. 基于 mRMR 原则和优化 SVM 的模拟电路故障诊断[J]. *仪器仪表学报*, 2013, 34(1): 221-226. (Sun J, Wang C H. Analog circuit fault diagnosis based on mRMR and optimized SVM[J]. *Chinese J of Scientific Instrument*, 2013, 34(1): 221-226.)
- [5] 刘飏, 陈春萍, 封化民, 等. 基于 Fisher 准则的 SVM 参数选择算法[J]. *山东大学学报: 理学版*, 2012, 47(7): 50-55. (Liu B, Chen C P, Feng H M, et al. A SVM parameters selection algorithm based on Fisher criterion[J]. *J of Shandong University: Natural Science*, 2012, 47(7): 50-55.)
- [6] 奉国和. SVM 分类核函数及参数选择比较[J]. *计算机工程与应用*, 2011, 47(3): 123-125. (Feng G H. Parameter optimizing for support vector machines classification[J]. *Computer Engineering and Applications*, 2011, 47(3): 123-125.)
- [7] 林升梁, 刘志. 基于 RBF 核函数的支持向量机参数选择[J]. *浙江工业大学学报*, 2007, 35(2): 163-167. (Lin S L, Liu Z. Parameter selection in SVM with RBF kernel function[J]. *J of Zhejiang University of Technology*, 2007, 35(2): 163-167.)
- [8] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines[J]. *Machine Learning*, 2002, 46(1): 131-159.
- [9] 唐耀华, 郭为民, 高静怀. 基于核相似性差异最大化的支持向量机参数选择算法[J]. *模式识别与人工智能*, 2010, 23(2): 210-215. (Tang Y H, Guo W M, Gao J H. SVM parameter selection algorithm based on maximum kernel similarity diversity[J]. *Pattern Recognition and Artificial Intelligence*, 2010, 23(2): 210-215.)
- [10] 李晓宇, 张新峰, 沈兰荪. 一种确定径向基核函数参数的方法[J]. *电子学报*, 2005, 33(12A): 2459-2463. (Li X Y, Zhang X F, Shen L S. A selection means on the parameter of radius basis function[J]. *Acta Electronica Sinica*, 2005, 33(12A): 2459-2463.)
- [11] Xiong H L, Swamy M N S, Ahmad M O. Optimizing the kernel in the empirical feature space[J]. *IEEE Trans on Neural Networks*, 2005, 16(2): 460-474.
- [12] 魏秀业, 潘宏侠. 粒子群优化及智能故障诊断[M]. 北京: 国防工业出版社, 2010: 69-73. (Wei X Y, Pan H X. Particle swarm optimization and intelligent fault diagnosis[M]. Beijing: National Defense Industry Press, 2010: 69-73.)
- [13] Smola A J, Schelkopf B. A tutorial on support vector regression[J]. *Statistics and Computing*, 2004, 14(3): 199-222.
- [14] Penalver B A, Escolano R F, Saez J M. Learning gaussian mixture models with entropy-based criterion[J]. *IEEE Trans on Neural Networks*, 2009, 20(11): 1756-1771.
- [15] Kennedy J K, Eberhart R C. Particle swarm optimization[C]. *Proc of the IEEE Int Conf on Neural Networks*. Perth Western, 1995: 1942-1946.

(责任编辑: 齐 霖)