

文章编号: 1001-0920(2006)01-0007-06

正则化最小二乘分类的AlignLoo模型选择方法

杨辉华^{1,2}, 王行愚¹, 王勇^{1,3}, 高海华¹

(1. 华东理工大学 信息科学与工程学院, 上海 200237; 2. 桂林电子工业学院 计算机系, 广西 桂林 541004; 3. 桂林电子工业学院 网络信息中心, 广西 桂林 541004)

摘要: 正则化最小二乘分类(RLSC)是一种基于二次损失函数的正则化网络,其推广能力受模型参数影响,传统的模型选择方法是耗时的参数网格搜索.为此,提出一种新颖的AlignLoo模型选择方法,其关键在于将核参数与超参数分开优化,即最大化核-目标配准以选择最优核参数,最小化RLSC的留一法误差的界以选择最优超参数.该方法效率高且不需验证样本,并在DA数据集上进行了测试,结果表明方法有效.

关键词: 核函数; 正则化; 最小二乘法; 分类; 模型选择; 支持向量机

中图分类号: TP18; TP393 **文献标识码:** A

AlignLoo Model Selection Approach for Regularized Least-squares Classification

YANG Hui-hua^{1,2}, WANG Xing-yu¹, WANG Yong^{1,3}, GAO Hai-hua¹

(1. College of Information Science and Engineering, East China University of Science & Technology, Shanghai 200237, China; 2. Department of Computer, Guilin University of Electronic Technology, Guilin 541004, China; 3. Network Information Center, Guilin University of Electronic Technology, Guilin 541004, China. Correspondent: WANG Xing-yu, E-mail: xywang@ecust.edu.cn)

Abstract: Regularized Least-Squares Classification (RLSC) is a kind of kernel-based square loss regularization network, and its generalization performance is severely influenced by the setting of its kernel and hyper parameters. A novel approach named AlignLoo is presented for optimal parameters selection, and the key of this approach lies in tuning the kernel parameter and hyper parameter sequentially. First, the optimal kernel parameter is selected by maximizing kernel target alignment, and then the optimal hyper-parameter is determined via minimizing RLSC's leave-one-out error bound. The experimental results on DA benchmark datasets demonstrate that the proposed approach is superior to traditional grid search method in that it is much more time efficient and it needs no independent validation set.

Key words: Kernel function; Regularization; Least-squares methods; Classification; Model selection; Support vector machines

1 引言

Rifkin 和 Poggio 称一种基于二次损失函数的正则化网络为“正则化最小二乘分类(RLSC)”^[1].最近, Poggio 和 Smale 明确指出“经典的二次损失函数正则化网络在分类中能同样工作得很好”^[2],称这一算法是学习理论的一个关键算法(KA)^[2],并将它建

立在“学习的数学基础”^[3]理论之上. KA 用二次损失函数处理回归和分类,为区别起见,本文称 KA 用于分类时为 RLSC. KA 与支持向量机(SVM)有密切联系,特别是 KA 与 LS-SVM^[4]和 PSVM^[5]都利用二次损失函数,故不需解二次规划,只需求解线性方程组.尽管源于不同的思想,KA 与核岭回归和高斯

收稿日期: 2005-01-04; 修回日期: 2005-03-15

基金项目: 国家重点基础研究发展规划项目(2002CB312200); 高等学校博士点基金项目(20040251010).

作者简介: 杨辉华(1972—),男,湖南澧县人,副教授,博士生,从事智能信息处理、模式识别等研究; 王行愚(1944—),男,上海人,教授,博士生导师,从事智能控制、网络信息安全等研究.

过程有相同的求解公式

RLSC 的推广能力与其核参数和超参数的选择有关 传统的模型选择方法是利用交叉验证技术作参数网格搜索, 特别耗时, 并需要单独的验证集 因此, 探讨高效的、不需要验证集的模型选择方法十分必要

本文的核心思想是如果能将核参数与超参数分开、依次优化, 则可为RLSC 提供一种简单高效的模型选择方法 为此, 需要有两个独立判据 一方面, 若能建立起目标(样本的类别标号)与核函数之间匹配程度的一种度量, 则可为确定最优核参数提供依据 Cristianini 等^[6]提出的“核-目标配准”正是这样的一种度量 由于核-目标配准是关于核的Gram 矩阵的函数, 从而是关于核参数的函数, 并且这一度量与超参数无关 因此, 最大化“核-目标配准”可为RLSC 选择最优核参数 另一方面, 由于RLSC 与SVM 存在密切联系, 故可借鉴最近提出的SVM 模型选择方法 基于最小化可导的SVM 的RM 界或Span 界, Chapelle 等^[7]提出了一种基于梯度下降的SVM 参数选择方法, 它可同时优化多个模型参数 受此启发, 本文利用RLSC 推广能力的某种界来选择参数 R ifkin^[1]给出了RLSC 的一个留一法误差的界(lo 界), 而这一界是关于超参数的函数 由于最小的 lo 界对应最小的期望预测误差, 因此通过最小化这一 lo 界可以确定最优超参数

2 正则化最小二乘分类

给定训练数据集 $S = (x_i, y_i)_{i=1}^m$, 其中 $x_i \in X$, $y_i \in Y$, X 是 R^n 的闭子集, $Y \subset R$, 文献[2] 描述的学习理论的一个关键算法(KA) 是采用如下的步骤建立预测模型 $f: X \rightarrow Y$.

1) 选择一个正定对称核 $K(x, x) = K(x, x)$, 如 Gaussian 核

$$K(x, x) = \exp(-\|x - x\|^2 / (2\sigma^2)).$$

2) 预测模型 $f: X \rightarrow Y$ 为

$$f(x) = \sum_{i=1}^m c_i K(x, x_i), \tag{1}$$

式中 $c = (c_1, \dots, c_m)^T$, 且

$$(mI + K)c = y. \tag{2}$$

其中: I 是一单位矩阵, K 是一 $m \times m$ 正定矩阵, 其元素 $K_{ij} = K(x_i, x_j)$, y 是向量 $(y_1, \dots, y_m)^T$, 超参数 γ 是一正实数

KA 的推导是基于最小化正则化风险泛函

$$\min_{f \in H_K} R_{\text{reg}}(f) = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 + \gamma \int f^2. \tag{3}$$

其中 $f \in H_K$ 是正定核 K 诱导的再生核希尔伯特空

间 H_K 中函数 f 的范数

在分类情况下(也即RLSC), 取 $y_i \in \{1, -1\}$, 并对式(1) 取符号函数从而得到分类决策函数 KA 及RLSC 的详细理论及推导见文献[1~ 3]

3 核 - 目标配准及核参数选择

基于核的学习算法将数据嵌入在核的 Gram 矩阵中 如果能建立起目标与核函数之间匹配程度的一种度量, 则可为确定最优核参数提供依据 Cristianini 等^[6]提出的“核 - 目标配准”正是这样的一种度量

定义 1(配准) 两个核 k_1 和 k_2 的配准定义为

$$A(k_1, k_2) = \frac{k_{1, k_2, P}}{\sqrt{k_{1, k_1, P} k_{2, k_2, P}}}, \tag{4}$$

其中 P 是产生数据的分布函数

尽管现实问题的配准无法准确计算, 但可通过它的样本 $S = (x_i, y_i)_{i=1}^m$ 作经验估计 首先定义两个 Gram 矩阵的内积为其 Frobenius 内积, 即

$$K_1, K_2_F = \sum_{i, j=1}^m K_1(x_i, x_j) K_2(x_i, x_j). \tag{5}$$

定义 2(经验配准) 两个核的经验配准定义为

$$\hat{A}(k_1, k_2) = \frac{K_1, K_2_F}{\sqrt{K_1, K_1_F K_2, K_2_F}}. \tag{6}$$

核 - 目标配准是经验配准的一个特例, 取 $K_2 = yy^T$, y 是样本的标号列向量, 于是

$$\hat{A}(K, yy^T) = \frac{K, yy^T}{\sqrt{K, K yy^T, yy^T}}. \tag{7}$$

配准的一个关键特性是它的实际值可以通过经验值有效估计, 相关定理见文献[6] 配准已被应用于基于转导的半监督学习、直接构建 Gram 矩阵及核函数的组合与选择^[6].

借鉴配准的思想, 由式(7) 可见核 - 目标配准(此处及以后皆指其经验值) 是关于核的 Gram 矩阵 K 的函数, 从而是关于核参数的隐函数, 故可通过最大化核 - 目标配准来为RLSC 选择最优的核参数

4 RLSC 留一法误差的界及超参数选择

学习算法的 lo 界使我们能有效地估计算法的推广能力, 特别是在没有独立的验证集或样本数很少的情况下. Jaakkola 和 Haussler^[8] 提出了基于核的分类器的一种简单的 lo 界 R ifkin^[1] 证明了这一界对RLSC 也成立, 其 lo 界为 $(|x_i: c|$ 表示满足条件 c 的样本点 x_i 的个数)

$$|x_i: y_i \sum_{j=1, j \neq i}^m c_j K_{ij} = 0| \tag{8}$$

这一界在给定 c_i 后可直接计算 然而, 一个简单的几何条件使得RLSC 能具有更好的界 以式(8) 为基

础, 给出一种较文献[1]更为简洁的 loo 界的推导:

结合式(1)和(2), 可得 $c_i = (y_i - f(x_i))/m\lambda$
 又由式(1)有

$$f(x_i) = \sum_{j=1}^m c_j K_{ij},$$

可得

$$\sum_{j=1, j \neq i}^m c_j K_{ij} = f(x_i) - c_i K_{ii}$$

现将 c_i 从式(8)中消去, 并考虑到 $y_i y_i = 1, m\lambda > 0$, 于是 RLSC 的 loo 误差以下式为界:

$$L = \left| x_i: y_i \left(f(x_i) - \left(\frac{y_i - f(x_i)}{m\lambda} \right) K_{ii} \right) \quad 0 \right| =$$

$$\left| x_i: y_i \left(\left(1 + \frac{K_{ii}}{m\lambda} \right) f(x_i) - \frac{y_i}{m\lambda} K_{ii} \right) \quad 0 \right| =$$

$$\left| x_i: y_i \left((m\lambda + K_{ii}) f(x_i) - y_i K_{ii} \right) \quad 0 \right| =$$

$$\left| x_i: y_i \left(m\lambda + K_{ii} \right) f(x_i) - K_{ii} \quad 0 \right| =$$

$$\left| x_i: y_i f(x_i) \quad K_{ii} / (m\lambda + K_{ii}) \right| \quad (9)$$

式(9)是关于超参数的函数, 而最小的 loo 界对应模型最强的推广能力, 因此只要通过第 3 部分的方法获得了最优核参数, 即可通过最小化(9)求最优超参数 λ

5 AlignLoo 模型选择方法

将核-目标配准的基本思想与 RLSC 的 loo 界相结合, 本文提出如下的 AlignLoo 模型选择方法

5.1 AlignLoo 模型选择方法

- 1) 预处理输入数据 $S = \{x_i, y_i\}_{i=1}^m$, 使 $y_i \in \{-1, 1\}$, $X = \{x_i\}$ 均值为零, 标准差为 1.
- 2) 最大化核-目标配准从而选择最优核参数
 以 Gaussian 核为例, 它有唯一核参数 σ^2 . 首先为每一个 σ^2 计算 Gram 矩阵 K , 然后根据式(7)计算配准 $A(K, yy^T)$, 使配准 A 最大的 σ^2 即为最优核参数
- 3) 最小化 loo 界从而选择最优超参数 λ
 首先根据步骤 2) 所得的最优核参数求出 Gram 矩阵 K , 然后对每一个 λ , 根据式(2)训练 RLSC 得到 c , 再根据式(1)得到 $f(x)$, 从而可根据式(9)得这一 λ 对应的 loo 界. 最优的 λ 值最小化 loo 界.
- 4) 利用步骤 2) 和步骤 3) 所得最优 σ^2 和 λ 再次训练模型, 得到 c .
- 5) 利用 $f(x) = \text{sign}\left(\sum_{i=1}^m c_i K(x, x_i)\right)$ 测试新的经预处理的数据 x .

5.2 计算技巧

首先, 由于 K 是对称阵, $y_i \in \{-1, 1\}$, 核-目标配准可根据下式高效计算:

$$K, yy^T = \text{trace}(K^T yy^T) =$$

$$\text{trace}((Ky)y^T) = y, Ky = y^T Ky =$$

$$\sum_{i,j=1}^m y_i y_j K_{ij} = \sum_{y_i=y_j} K_{ij} - \sum_{y_i \neq y_j} K_{ij}, \quad (10)$$

$$K, K = \text{trace}(KK) =$$

$$\sum_{i,j=1}^m K_{ij}^2, \quad (11)$$

$$yy^T, yy^T = \text{trace}((yy^T)^T yy^T) =$$

$$\text{trace}(y(y^T y)y^T) = m \cdot \text{trace}(yy^T) = m^2. \quad (12)$$

式(10)和(11)可根据不同的软件环境, 相应地选择矩阵或是下标循环的计算方式. 式(11)中 $\text{vec}K$ 表示矩阵按列拉直. 式(10)中最后一个等号的前一项为同类元素和, 后一项为异类元素和, 这揭示了 Align 判据与 Fisher 判据相关.

其次, 由于 Gaussian 核 K_{ii} 总为 1, 此时 RLSC 的 loo 界简化为

$$L = \left| x_i: (1 + m\lambda) y_i f(x_i) - 1 \quad 0 \right| =$$

$$\sum_{i=1}^m \Psi(1 - y_i f(x_i) (1 + m\lambda)), \quad (13)$$

其中 $\Psi(x)$ 是一阶跃函数

5.3 核-目标配准 A 与 loo 界 L 的导数及其优化

AlignLoo 方法的第 2) 和第 3) 步, 均涉及单参数优化问题. 如果 A 与 L 可导且存在极值点, 则可通过令 $\frac{\partial A}{\partial \sigma^2} = 0$ 及 $\frac{\partial L}{\partial \lambda} = 0$, 从而求得最优参数. 下面仍以 Gaussian 核为例, 对有关问题分析如下:

设

$$K_{ij} = \exp\left[-\frac{(x_i - x_j)^2}{2\sigma^2}\right] = \exp\left[-\frac{d_{ij}^2}{2\sigma^2}\right],$$

则有 $\frac{\partial K_{ij}}{\partial \sigma^2} = d_{ij}^2 K_{ij} / 2\sigma^4,$

A 对 σ^2 求偏导, 可得

$$\frac{\partial A}{\partial \sigma^2} = \frac{1}{2m\sigma^4 B \sqrt{B}} \left[\left[\sum_{i,j=1}^m K_{ij}^2 \right] \left[\sum_{i,j=1}^m y_i y_j K_{ij} d_{ij}^2 \right] - \left[\sum_{i,j=1}^m K_{ij}^2 d_{ij}^2 \right] \left[\sum_{i,j=1}^m y_i y_j K_{ij} \right] \right], \quad (14)$$

其中 $B = K, K = \sum_{i,j=1}^m K_{ij}^2$. 由于 K_{ij} 是关于 σ^2 的隐函数, 故令 $\partial A / \partial \sigma^2 = 0$ 的极值点 σ^2 只能通过数值方法求解, 并且这一极值点完全由核函数及数据 (d_{ij}^2, y_i) 的分布确定.

由式(13)可知, 因 L 不连续, 故不能直接对 L 求导.

综合上述分析, 不能利用 A 和 L 的导数进行优化, 故对 σ^2 和 λ 分别使用一维线搜索. 这种简单方法能有效处理目标函数存在多个局部极值的情况, 实际应用表明该方法快速、有效.

6 实验测试

本文提出的方法在 DA 标准数据集^[9]上进行了测试. DA 数据集包括 13 个两类分类问题, 并且为每一分类问题划分好 100 或 20 个训练与测试样本子集, 其目的是消除数据划分的不一致, 从而使该数据

集能更好地为评估各种算法的性能服务

对于RLSC算法, 本文将AlignLoo模型选择方法与传统的参数网格搜索模型选择方法进行比较。对于参数网格搜索模型选择方法(简记为Grid), 使用参数范围 $\ln\sigma^2 = \{-10, -1, 10\}$ 及 $\ln\gamma = \{-15, -1, 5\}$, 因此共有 21×21 个参数组合。对每一数据集, 本文采用与文献[9]相同的方法, 即对参数网格搜索, 使用最前面的5个训练/测试数据划分进行训练与测试, 取5个测试错误率的平均值最小的参数格点为最优参数对。因此, 为完成最优参数选择, Grid需要进行 $21 \times 21 \times 5 = 2205$ 次训练与测试运算。在得到最优参数后, 对该数据集所有的100或20个数据划分进行测试, 并对测试误差率求平均从而得到该算法在该数据集上的性能估计。

对于AlignLoo模型选择方法, 为便于比较, 本文使用相同的参数网格, 对应着 $21 + 21 = 42$ 次搜索运算, 但前21次运算只需要计算核-目标配准, 只有后21次搜索才需要进行RLSC训练与测试。同时, 考虑到AlignLoo模型选择方法可以为每一训练数据子集(划分)作最优参数估计, 不需要进行交叉验证, 因此对每一个训练/测试数据划分均作最优参数选择。

Grid和AlignLoo两种模型选择方法的性能分析见表1和表2。表1显示两种不同方法的时间消耗, 其中: Grid的时间对应着在前5个划分上进行网格搜索, 而AlignLoo的时间对应着在一个数据划分上得到最优参数。可见AlignLoo的时间代价远远小于Grid。

表2显示的是两种不同方法在DA数据集上的

测试误差率及其标准差, 表中的粗体表示对应该算法在该数据集上具有最优性能。前5种方法的结果来自文献[9], 第6列的结果来自基于BFGS优化的SVM算法(Gaussian核)^[10]。可见, RLSC具有较好的性能, AlignLoo的性能略低于Grid。性能略低的原因一方面在于Grid利用了多个数据划分作交叉验证, 而AlignLoo为每个划分求最优参数; 另一方面, 它们的工作原理不同, Grid直接利用真实数据验证, 而AlignLoo利用核-目标配准和looo界两种判据。进一步的研究显示, 通过在AlignLoo第一轮参数搜索所得最优参数附近进行细化搜索, AlignLoo一般能够提高性能, 但个别情况导致性能下降。例如, 对Titanic数据集, 本文用此方法得到100个划分的平均测试错误率为 22.38 ± 1.00 , 这是所有结果中最好的。当然, 这一方法对Grid也有效, 但其计算代价巨大而难以承受。

图1显示了Banana数据集上Gaussian核的核-目标配准随核参数变化的情况。图2显示了Banana数据集上RLSC的looo界随超参数变化的情况。本文发现RLSC的looo界是紧的。如图2所示, 第一个数据划分的looo界为37, 由100个划分得到的looo界的均值为42.57, 对应looo误差率的均值10.64%, 而在100个测试划分上, Banana数据集的真正平均测试错误率为11.01%。

本文利用AlignLoo为每一数据集的各个划分选择最优参数, 那么这些最优参数是否稳定, 与Grid所得的最优参数关系怎样就成为本文关心的问题。本文发现AlignLoo方法得到的最优核参数与最优超参数均相当稳定。以数据集Thyroid为例, 如图

表1 DA数据集的简要描述及RLSC-Grid和RLSC-AlignLoo算法的时间代价

No.	数据集	样本维数	训练集	测试集	划分数	RLSC-Grid /min	RLSC-AlignLoo /s
1	Banana	2	400	4 900	100	79.6	1.88
2	Breast Cancer	9	200	77	100	8.7	0.24
3	Diabetes	8	468	300	100	81.6	2.55
4	Flare Sonar	9	666	400	100	149.8	6.68
5	German	20	700	300	100	258.6	7.19
6	Heart	13	170	100	100	6.4	0.14
7	Image	18	1 300	1 010	20	1 410.2	40.30
8	Ringnom	20	400	7 000	100	162.8	1.57
9	Splice	60	1 000	2 175	20	1 009.7	20.21
10	Thyroid	5	140	75	100	3.2	0.11
11	Titanic	3	150	2 051	100	10.7	0.19
12	Twonom	20	400	7 000	100	179.1	1.69
13	Waveform	21	400	4 600	100	147.3	1.75

表2 RLSC-Grid 和 RLSC-AlignLoo 模型选择算法的平均测试错误率及标准差

No.	RBF- Network	AdaBoost RBF	LP- Reg- AdaBoost	AdaBoost - Reg	SVM RBF	SVM BFGS	RLSC-Grid	RLSC -AlignLoo
1	10.76 ± 0.42	12.26 ± 0.67	10.73 ± 0.43	10.85 ± 0.42	11.53 ± 0.66	11.45	10.41 ± 0.44	11.01 ± 0.62
2	27.64 ± 4.71	30.36 ± 4.73	26.79 ± 6.08	26.51 ± 4.47	26.04 ± 4.74	28.57	25.43 ± 4.03	25.60 ± 4.26
3	24.29 ± 1.88	26.47 ± 2.29	24.11 ± 1.90	23.79 ± 1.80	23.53 ± 1.73	24.33	22.99 ± 1.69	23.70 ± 1.93
4	34.37 ± 1.95	35.70 ± 1.79	34.74 ± 2.00	34.20 ± 2.18	32.43 ± 1.82	35.00	33.48 ± 1.79	35.86 ± 1.82
5	24.71 ± 2.38	27.45 ± 2.50	24.79 ± 2.22	24.34 ± 2.08	23.61 ± 2.07	19.67	24.61 ± 2.29	26.81 ± 2.91
6	17.55 ± 3.25	20.29 ± 3.44	17.49 ± 3.53	16.47 ± 3.51	15.95 ± 3.26	19.00	16.61 ± 3.95	17.03 ± 3.45
7	3.32 ± 0.65	2.73 ± 0.66	2.76 ± 0.61	2.67 ± 0.61	2.96 ± 0.60	2.97	2.84 ± 0.53	3.15 ± 0.95
8	1.70 ± 0.21	1.93 ± 0.24	2.24 ± 0.46	1.58 ± 0.12	1.66 ± 0.12	—	2.44 ± 0.16	2.58 ± 0.24
9	9.95 ± 0.78	10.14 ± 0.51	10.22 ± 1.59	9.50 ± 0.65	10.88 ± 0.66	10.16	10.91 ± 0.81	12.95 ± 0.90
10	4.52 ± 2.12	4.40 ± 2.18	4.59 ± 2.22	4.55 ± 2.19	4.80 ± 2.19	2.67	4.21 ± 2.12	5.24 ± 2.48
11	23.26 ± 1.34	22.58 ± 1.18	23.98 ± 4.38	22.64 ± 1.20	22.42 ± 1.02	22.91	22.55 ± 1.13	23.12 ± 1.46
12	2.85 ± 0.28	3.03 ± 0.28	3.17 ± 0.43	2.70 ± 0.24	2.96 ± 0.23	—	2.39 ± 0.12	2.47 ± 0.20
13	10.66 ± 1.08	10.84 ± 0.58	10.53 ± 1.02	9.79 ± 0.81	9.88 ± 0.43	—	9.54 ± 0.46	10.43 ± 0.92

3(a) 所示, AlignLoo 得到的 100 个最优参数对 (σ^2, γ) 在图上用星号表示, 很明显, 它们重叠为 3 个点, 并且位于 Grid 所得的最优参数点 (用菱形点表示) 附近. 图 3(b) 为 Grid 方法在 Thyroid 数据集上参数网格的测试精度的 mesh 可视化表示, 其中单独标出的点对应 Grid 所得最优参数. AlignLoo 算法所得最优核参数的稳定性说明了经验配准是真实配准的一个良好逼近. 值得指出的是, AlignLoo 所得最优参数与网格搜索所得最优参数并不一致, 类似现象在

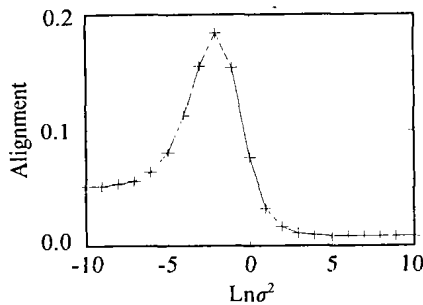


图1 Banana 数据集 Gaussian 核的核-目标配准

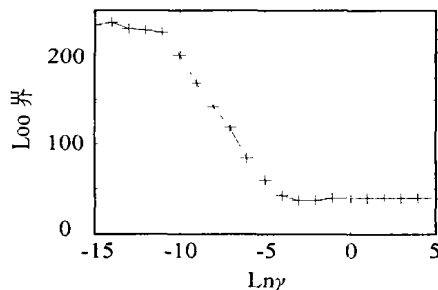
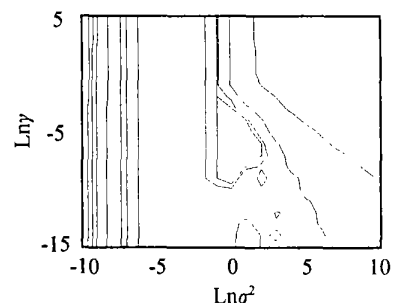
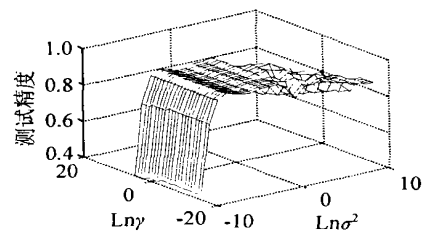


图2 Banana 数据集上 RLSC 的 loo 界



(a) contour 图



(b) mesh 图

图3 Thyroid 数据集上参数网格的测试精度
基于梯度的 SVM 最优参数搜索中也同样存在^[11].

7 结 语

为确定 RLSC 的最优模型参数, 本文提出了一种 AlignLoo 算法, 它首先最大化核-目标配准以选择最优核参数, 再最小化 loo 界以选择最优超参数. 由于将核参数和超参数分开优化, 故计算量大大低于传统的网格搜索, 且不需要单独的验证集. 对于基于核的其他分类器, 若知道某种推广能力的界, 则本文提出的模型选择方法也可适用.

参考文献(References)

- [1] Rifkin R M, Yeo G, Poggio T. Regularized Least Squares Classification [A]. *Advances in Learning Theory: Methods, Model and Applications*[C]. NATO Science Series III: Computer and Systems Sciences, Amsterdam: DS Press, 2003: 131-153
- [2] Poggio T, Smale S. The Mathematics of Learning: Dealing With Data [J]. *Notice of American Mathematical Society*, 2003, 50(5): 537-544
- [3] Cucker F, Smale S. On the Mathematical Foundations of Learning [J]. *Bulletin of American Mathematical Society*, 2001, 39(1): 1-49
- [4] Suykens J A K, Gestel V, De Brabanter J, et al. *Least Squares Support Vector Machine* [M]. Singapore: World Scientific Press, 2002
- [5] Fung G, Mangasarian O L. Proximal Support Vector Machine Classifiers [A]. *Seventh ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining* [C]. San Francisco: Association for Computing Machinery Press, 2001: 77-86
- [6] Cristianini N, Kandola J, Elisseeff A, et al. *On Kernel Target Alignment* [EB/OL]. <http://www.support-vector.net/papers/alignment>. JMLR. ps 2005-01-01.
- [7] Chapelle O, Vapnik V, Mukherjee S. Choosing Multiple Parameters for Support Vector Machines [J]. *Machine Learning*, 2002, 46(1): 131-159
- [8] Jaakkola T, Haussler D. Probabilistic Kernel Regression Models [A]. *Proc of the 1999 Conf on AI and Statistics* [C]. San Mateo: Morgan Kaufmann, 1999
- [9] Ratsch G, Onoda T, Müller K-R. Soft Margins for AdaBoost [J]. *Machine Learning*, 2001, 42(3): 287-320
- [10] Keerthi S S. Efficient Tuning of SVM Hyperparameters Using Radius/margin Bound and Iterative Algorithms [J]. *IEEE Trans on Neural Networks*, 2002, 13(5): 1225-1229
- [11] Staelin C. *Parameter Selection for Support Vector Machines* [R]. HP Laboratories, 2003
- [12] Guyon I, Stork D G. Linear Discriminant and Support Vector Classifiers [A]. *Advances in Large Margin Classifiers* [C]. Cambridge MA: MIT Press, 2000: 147-170
- [13] 许建华, 张学工, 李衍达. 一种基于核函数的非线性感知器算法 [J]. *计算机学报*, 2002, 25(7): 689-695 (Xu J H, Zhang X G, Li Y D. A Nonlinear Perceptron Algorithm Based on Kernel Functions [J]. *Chinese J of Computers*, 2002, 25(7): 689-695)
- [14] 许建华, 张学工, 李衍达. 基于核函数的非线性口袋算法 [J]. *电子学报*, 2003, 31(4): 612-615 (Xu J H, Zhang X G, Li Y D. Nonlinear Pocket Algorithm with Kernels [J]. *Acta Electronica Sinica*, 2003, 31(4): 612-615)
- [15] Gallant S I. Optimal Linear Discriminant [A]. *Proc of the Eighth Int Conf on Pattern Recognition* [C]. New York: IEEE Press, 1986: 849-853
- [16] Muselli M. On Convergence Properties of Pocket Algorithm [J]. *IEEE Trans on Neural Networks*, 1997, 8(3): 623-629
- [17] Xu J, Zhang X, Li Y. Large Margin Kernel Pocket Algorithm [A]. *Proc of 2001 Int Joint Conf on Neural Networks* [C]. New York: IEEE Press, 2001: 1480-1485
- [18] Xu J, Zhang X, Li Y. Kernel Neuron and its Training Algorithm [A]. *Proc of the Eighth Int Conf on Neural Information Processing* [C]. Shanghai: Fudan University Press, 2001, 2: 861-866
- [19] Ishikawa M. Structural Learning with Forgetting [J]. *Neural Networks*, 1996, 9(3): 509-521
- [20] Xu J, Zhang X, Li Y. Sparse Training Procedure for Kernel Neuron [A]. *Proc of Int Conf on Neural Networks and Signal Processing* [C]. New York: IEEE Press, 2003, 1: 49-53
- [21] Xu J, Zhang X. A Learning Algorithm with Gaussian Regularizer for Kernel Neuron [A]. *2004 Int Symposium on Neural Networks* [C]. Berlin: Springer, 2004: 252-257.
- [22] Yu K, Ji L, Zhang X. Kernel Nearest Neighbor Algorithm [J]. *Neural Processing Letters*, 2002, 15(2): 147-156
- [23] Joachims T. Making Large-scale SVM Learning Practical [A]. *Advances in Kernel Methods - Support Vector Learning* [C]. Cambridge MA: MIT Press, 1999: 169-184
- [24] Baldi P, Brunak S. *Bioinformatics: The Machine Learning Approach* [M]. 2nd ed. Cambridge MA: MIT Press, 2001.
- [25] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. *IEEE Trans on Neural Networks*, 2002, 13(2): 415-425

(上接第6页)

- [26] Guyon I, Stork D G. Linear Discriminant and Support Vector Classifiers [A]. *Advances in Large Margin Classifiers* [C]. Cambridge MA: MIT Press, 2000: 147-170
- [27] 许建华, 张学工, 李衍达. 一种基于核函数的非线性感知器算法 [J]. *计算机学报*, 2002, 25(7): 689-695 (Xu J H, Zhang X G, Li Y D. A Nonlinear Perceptron Algorithm Based on Kernel Functions [J]. *Chinese J of Computers*, 2002, 25(7): 689-695)
- [28] 许建华, 张学工, 李衍达. 基于核函数的非线性口袋算法 [J]. *电子学报*, 2003, 31(4): 612-615 (Xu J H, Zhang X G, Li Y D. Nonlinear Pocket Algorithm with Kernels [J]. *Acta Electronica Sinica*, 2003, 31(4): 612-615)
- [29] Gallant S I. Optimal Linear Discriminant [A]. *Proc of the Eighth Int Conf on Pattern Recognition* [C]. New York: IEEE Press, 1986: 849-853
- [30] Muselli M. On Convergence Properties of Pocket Algorithm [J]. *IEEE Trans on Neural Networks*, 1997, 8(3): 623-629
- [31] Xu J, Zhang X, Li Y. Large Margin Kernel Pocket Algorithm [A]. *Proc of 2001 Int Joint Conf on Neural Networks* [C]. New York: IEEE Press, 2001: 1480-1485
- [32] Xu J, Zhang X, Li Y. Kernel Neuron and its Training Algorithm [A]. *Proc of the Eighth Int Conf on Neural Information Processing* [C]. Shanghai: Fudan University Press, 2001, 2: 861-866
- [33] Ishikawa M. Structural Learning with Forgetting [J]. *Neural Networks*, 1996, 9(3): 509-521
- [34] Xu J, Zhang X, Li Y. Sparse Training Procedure for Kernel Neuron [A]. *Proc of Int Conf on Neural Networks and Signal Processing* [C]. New York: IEEE Press, 2003, 1: 49-53
- [35] Xu J, Zhang X. A Learning Algorithm with Gaussian Regularizer for Kernel Neuron [A]. *2004 Int Symposium on Neural Networks* [C]. Berlin: Springer, 2004: 252-257.
- [36] Yu K, Ji L, Zhang X. Kernel Nearest Neighbor Algorithm [J]. *Neural Processing Letters*, 2002, 15(2): 147-156
- [37] Joachims T. Making Large-scale SVM Learning Practical [A]. *Advances in Kernel Methods - Support Vector Learning* [C]. Cambridge MA: MIT Press, 1999: 169-184
- [38] Baldi P, Brunak S. *Bioinformatics: The Machine Learning Approach* [M]. 2nd ed. Cambridge MA: MIT Press, 2001.
- [39] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. *IEEE Trans on Neural Networks*, 2002, 13(2): 415-425