

文章编号: 1001-0920(2012)06-0866-05

无偏置 ν -支持向量回归优化问题研究

丁晓剑, 赵银亮

(西安交通大学 电子与信息工程学院, 西安 710049)

摘要: 在高维特征空间中, 具有支持向量机形式的学习机的决策超平面倾向于通过原点, 并不需要偏置. 但在 ν -支持向量回归机 (ν -SVR) 中存在偏置, 为了研究偏置在 ν -SVR 中的作用, 提出了无偏置的 ν -SVR 优化问题并给出其求解方法. 在标准数据集上的实验表明, 无偏置 ν -SVR 的泛化性能好于 ν -SVR. 根据对偶优化问题的解空间分析, 偏置 b 不应包含在 ν -SVR 优化问题中, ν -SVR 的决策超平面在高维特征空间中应通过原点.

关键词: 偏置; ν -支持向量回归机; 优化问题

中图分类号: TP273

文献标识码: A

Study on ν -support vector regression optimization problem without bias

DING Xiao-jian, ZHAO Yin-liang

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China. Correspondent: ZHAO Yin-liang, E-mail: zhaoy@mail.xjtu.edu.cn)

Abstract: In the high-dimensional feature space, the decision hyperplane of learning machine with support vector machine style tends to pass through the origin and the bias b is not needed. However, bias exists in ν -support vector regression (ν -SVR). To study the role of bias in ν -SVR, optimization formulation of ν -SVR without bias is proposed and the corresponding method of solving the dual optimization formulation is presented. The experimental results on benchmark data sets show that the generalization ability of ν -SVR without bias is better than ν -SVR. Based on the analysis of solution space on dual optimization formulation, the bias b should not be contained in the optimization formulation of ν -SVR, and the hyperplane of ν -SVR should pass through the origin.

Key words: bias; ν -support vector regression; optimization formulation

1 引言

Huang 和 Ding^[1]指出, 分类问题在高维空间中, 分类超平面倾向于通过原点, 即不需要偏置 b . 对于回归问题, 因为回归超平面在高维空间中的位置与偏置 b 的值有关, 而不同位置的超平面会影响 ν -支持向量回归机 (ν -SVR)^[2-4] 的泛化性能, 所以在理论上 ν -SVR 的泛化性能应与 b 有关. Poggio^[5] 从核函数的正定理论分析得出, 如果核函数是正定的, 则在支持向量机 (SVM) 的优化问题中不需要 b ; 文献 [6-7] 也指出, 使用高斯核等常用核函数的 SVM 优化问题不需要偏置 b . 因此无偏置的 ν -SVR 的泛化性能是一个值得研究的问题.

鉴于 Chang 和 Lin^[8] 从理论上已研究了 ν -SVR 与 Vapnik 提出的 ε -支持向量回归机之间的关系, 探讨了参数 ν 的一些性质; Chalimourda 等人^[9] 从不同的噪声

模型和参数设置下研究了 ν 的最优值, 本文仅研究不同参数 ν 和核函数参数的值对下无偏置的 ν -SVR 和 ν -SVR. 首先提出了无偏置 ν -SVR 的优化问题; 然后给出了优化问题的求解方法, 包括迭代停止条件和子优化问题的求解方法.

2 无偏置 ν -SVR 的优化问题

给定一个含有 l 个样本的训练样本集 $\{\mathbf{x}_i, y_i\}_{i=1}^l$. 其中: 输入为 d 维向量 \mathbf{x}_i , 输出为 y_i . 通常训练样本集在输入空间中是线性不可分的, 需要引入映射函数 $\phi(\mathbf{x})$ 将 \mathbf{x}_i 映射到高维空间 $\phi(\mathbf{x}_i)$ 中. 由于无偏置 ν -SVR 优化问题的约束条件中没有 b 项, 其优化问题可表示为

$$\min \tau(\mathbf{w}, \xi^{(*)}, \varepsilon) =$$

收稿日期: 2010-12-01; 修回日期: 2011-05-02.

基金项目: 国家自然科学基金项目(61173040).

作者简介: 丁晓剑(1982-), 男, 博士生, 从事神经网络和机器学习的研究; 赵银亮(1960-), 男, 教授, 博士生导师, 从事并行计算、数据挖掘等研究.

$$\begin{aligned} & \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \right); \\ \text{s.t. } & \mathbf{w}\phi(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i, \\ & y_i - \mathbf{w}\phi(\mathbf{x}_i) \leq \varepsilon + \xi_i^*, \\ & \xi_i^{(*)} \geq 0, \varepsilon \geq 0. \end{aligned} \quad (1)$$

对式(1)引入拉格朗日函数

$$\begin{aligned} L(\mathbf{w}, \alpha^{(*)}, \beta, \xi^{(*)}, \varepsilon, \eta^{(*)}) = & \\ & \frac{1}{2} \|\mathbf{w}\|^2 + C\nu\varepsilon + \frac{C}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) - \beta\varepsilon - \\ & \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^l \alpha_i (\xi_i + y_i - \mathbf{w}\phi(\mathbf{x}_i) + \varepsilon) - \\ & \sum_{i=1}^l \alpha_i^* (\xi_i^* + \mathbf{w}\phi(\mathbf{x}_i) - y_i + \varepsilon). \end{aligned} \quad (2)$$

其中: $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ 为非负拉格朗日乘子, 即

$$\alpha_i \geq 0, \alpha_i^* \geq 0, \eta_i \geq 0, \eta_i^* \geq 0. \quad (3)$$

为了最小化式(3)的目标函数, 对 L 关于向量 \mathbf{w} 和变量 $\varepsilon, \xi_i, \xi_i^*$ 求偏导数, 当下列条件都满足时便可得到式(3)的最优解:

$$\partial L / \partial \mathbf{w} = 0 \Rightarrow \mathbf{w} = \sum_i (\alpha_i^* - \alpha_i) \mathbf{x}_i, \quad (4)$$

$$\partial L / \partial \varepsilon = 0 \Rightarrow C\nu - \sum_i (\alpha_i + \alpha_i^*) - \beta = 0, \quad (5)$$

$$\partial L / \partial \xi_i^{(*)} = 0 \Rightarrow C/l - \alpha_i^{(*)} - \eta_i^{(*)} = 0. \quad (6)$$

将式(4)~(6)代入(1), 可得对偶优化问题

$$\begin{aligned} \min & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) - \\ & \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i; \\ \text{s.t. } & 0 \leq \alpha_i^{(*)} \leq C/l, \\ & \sum_{i=1}^l (\alpha_i + \alpha_i^*) \leq C\nu. \end{aligned} \quad (7)$$

其中 $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ 为满足 Mercer 定理的核函数. 根据 Chang 和 Lin^[8] 的证明, 对于任意给定的 ν , 式(7)至少有一个最优解满足 $\sum_{i=1}^l (\alpha_i + \alpha_i^*) = C\nu$,

即约束 $\sum_{i=1}^l (\alpha_i + \alpha_i^*) \leq C\nu$ 可以转化为 $\sum_{i=1}^l (\alpha_i + \alpha_i^*) = C\nu$ 进行求解.

3 无偏置 ν -SVR 求解方法

文献[10]提出的有效集算法能够高效求解 SVM 优化问题. 该有效集算法将 SVM 对偶优化问题分解为一系列子优化问题来求解, 其主要计算代价是每次迭代过程中均需求解子优化问题. 无偏置 ν -SVR 对

偶优化问题的形式与 SVM 优化问题较为相似, 但无法直接利用[10]的有效集算法求解. 本节将对无偏置 ν -SVR 对偶优化问题进行一系列推导和转换, 包括式(7)的迭代停止条件推导和将式(7)分解为一系列子优化问题并给出求解方法; 然后利用[10]的循环过程寻找最优解.

3.1 Karush-Kuhn-Tucker 条件

为了得到式(7)的最优解, 需要推导其完整的 Karush-Kuhn-Tucker (KKT) 条件, 即:

初始可行性

$$\begin{cases} \xi_i^{(*)} \geq 0, \varepsilon \geq 0, \\ \mathbf{w}\phi(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i, \quad \forall i; \\ y_i - \mathbf{w}\phi(\mathbf{x}_i) \leq \varepsilon + \xi_i^*, \end{cases} \quad (8)$$

对偶可行性

$$\alpha_i \geq 0, \alpha_i^* \geq 0, \eta_i \geq 0, \eta_i^* \geq 0, \quad \forall i; \quad (9)$$

对偶与原问题的补条件

$$\begin{cases} \alpha_i (\xi_i + y_i - \mathbf{w}\phi(\mathbf{x}_i) + \varepsilon) = 0, \\ \alpha_i^* (\xi_i^* + \mathbf{w}\phi(\mathbf{x}_i) - y_i + \varepsilon) = 0, \quad \forall i. \\ \beta \varepsilon = 0, \eta_i \xi_i = 0, \eta_i^* \xi_i^* = 0, \end{cases} \quad (10)$$

由式(6)和补条件(10)可推得

$$(C/l - \alpha_i) \xi_i = 0, (C/l - \alpha_i^*) \xi_i^* = 0. \quad (11)$$

令 f_i 为无偏置 ν -SVR 对样本 (\mathbf{x}_i, y_i) 预测的值, 有 $f_i = \mathbf{w}\phi(\mathbf{x}_i)$. 当 $\alpha_i = 0$ 时, 由式(11)可知 $\xi_i = 0$, 再由式(8)可知 $f_i - y_i \leq \varepsilon$; 当 $\alpha_i = C/l$ 时, 由式(10)可知 $\xi_i + y_i - f_i + \varepsilon = 0$, 再由式(8)可知 $f_i - y_i \geq \varepsilon$; 当 $0 \leq \alpha_i \leq C/l$ 时, 由式(11)可知 $\xi_i = 0$, 再由式(10)可知 $\xi_i + y_i - f_i + \varepsilon = 0$, 即有 $f_i - y_i = \varepsilon$. 同理可推得 α_i^* 的 KKT 条件. 综上, 有下面条件成立:

$$\begin{aligned} \alpha_i^{(*)} = 0 & \Leftrightarrow f_i - y_i \leq \varepsilon, \\ 0 < \alpha_i^{(*)} < C/l & \Leftrightarrow f_i - y_i = \varepsilon, \\ \alpha_i^{(*)} = C/l & \Leftrightarrow f_i - y_i \geq \varepsilon. \end{aligned} \quad (12)$$

只有当式(12)满足时, 优化问题(7)才能得到最优解.

3.2 求解过程

令 $W(\alpha^{(*)})$ 为式(7)的目标优化函数, 则 $W(\alpha^{(*)})$ 可表示为如下形式:

$$\begin{aligned} W(\alpha^{(*)}) = & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) - \\ & \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i = \\ & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i \alpha_j - \alpha_i \alpha_j^* - \alpha_i^* \alpha_j + \alpha_i^* \alpha_j^*) \times \\ & K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i. \end{aligned} \quad (13)$$

令 $\bar{\alpha} = [\alpha; \alpha^*]$, $\bar{K} = \begin{bmatrix} K & -K^T \\ -K & K \end{bmatrix}$, $\bar{p} = [y; -y]$, $\bar{e} = [e; e]$, $\mathbf{y} = [y_1, \dots, y_l]^T$, $\mathbf{e} = [1, \dots, 1]^T$, 则式(7)可表示为如下形式:

$$\begin{aligned} \min W(\bar{\alpha}) &= \frac{1}{2} \bar{\alpha}^T \bar{K} \bar{\alpha} - \bar{p}^T \bar{\alpha}, \\ \text{s.t. } & 0 \leq \bar{\alpha}_i \leq C/l, \bar{e}^T \bar{\alpha} = C\nu. \end{aligned} \quad (14)$$

若将 C/l 视为 C , 则相应的约束 $\bar{e}^T \bar{\alpha} = C\nu$ 将变为 $\bar{e}^T \bar{\alpha} = C\nu$. 为了更好地描述有效集方法, 先定义一些集合和向量. 定义 3 个集合 $S_0 := \{i | \bar{\alpha}_i = 0\}$, $S_C := \{i | \bar{\alpha}_i = C\}$ 和 $S_{\text{work}} := \{i | \bar{\alpha}_i \in (0, C)\}$, 根据这些集合可定义向量 $\bar{\alpha}_0$, $\bar{\alpha}_C$ 和 $\bar{\alpha}_{\text{work}}$.

由于向量 $\bar{\alpha}_0$ 和 $\bar{\alpha}_C$ 中元素的值是不变的, 求解式(14)等价于求解下式:

$$\begin{aligned} \min W(\bar{\alpha}_{\text{work}}) &= \\ & \frac{1}{2} \bar{\alpha}_{\text{work}}^T \bar{K}_{ww} \bar{\alpha}_{\text{work}} + \bar{\alpha}_C^T \bar{K}_{wC} \bar{\alpha}_{\text{work}}; \\ \text{s.t. } & 0 \leq \bar{\alpha}_i \leq C, \\ & \sum \bar{\alpha}_{\text{work}} = C\nu - \sum \bar{\alpha}_C. \end{aligned} \quad (15)$$

其中: \bar{K}_{ww} 和 \bar{K}_{wC} 为 \bar{K} 的子矩阵, 行和列由相应的集合 S_C 和 S_{work} 标识. 同理, 由于式(15)是对 $\bar{\alpha}_{\text{work}}$ 求最小值, 由 $\bar{\alpha}_{\text{work}}$ 的定义可知, $\bar{\alpha}_{\text{work}}$ 中分量的值都在区间 $(0, C)$ 中, 即约束 $0 \leq \bar{\alpha}_i \leq C$ 必然满足. 于是, 优化问题(15)变为如下等式约束优化问题:

$$\begin{aligned} \min_{\delta} & \frac{1}{2} \delta^T Q \delta + \mathbf{c}^T \delta, \\ \text{s.t. } & \mathbf{A} \delta = t. \end{aligned} \quad (16)$$

其中: $\delta = \bar{\alpha}_{\text{work}} \in \mathbf{R}^n$, $Q = \bar{K}_{ww}$, $\mathbf{c} = \bar{K}_{wC} \bar{\alpha}_C$, $\mathbf{A} = \bar{e}^T$, $t = C\nu - \sum \bar{\alpha}_C$, $\bar{e} = [1, \dots, 1]^T \in \mathbf{R}^n$, n 为 $\bar{\alpha}_{\text{work}}$ 中变量的个数.

式(16)可以通过解条件问题的乘子法来求解. 构造拉格朗日函数

$$\bar{L}(\delta, \lambda) = \frac{1}{2} \delta^T Q \delta + \mathbf{c}^T \delta + \lambda^T (\mathbf{A} \delta - t), \lambda \in \mathbf{R}^n. \quad (17)$$

令 $\bar{L}(\delta, \lambda)$ 对 δ 和 λ 的导数为零, 得到线性方程组

$$\begin{bmatrix} Q & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \delta \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ t \end{bmatrix}. \quad (18)$$

其中: $\mathbf{0} = [0, \dots, 0] \in \mathbf{R}^n$, δ 和 λ 可通过求解式(18)得到. 至此, 无偏置 ν -SVR 对偶优化问题(7)便转化为子优化问题(16)来求解, 而(16)可通过解线性方程组(18)得出.

根据求解 SVM 优化问题的有效集算法^[10], 式(7)可以通过 2 层循环来求解. 即先判断外循环 $\bar{\alpha}$ 中边界约束元素是否都满足 KKT 条件(12). 如果满足, 则算法终止; 如果不满足, 则选取违反 KKT 条件的变量进入内循环求解. 内循环目标是求解子优化问题(16), 使得最优解 $\bar{\alpha}_{\text{work}}$ 满足上下界约束.

4 实验分析

为了公平地比较 ν -SVR 和无偏置 ν -SVR 的泛化性能, 分别记作 NUSVR 算法和 NBNUSVR 算法. 核函数均选择高斯核函数 $K(\mathbf{u}, \mathbf{v}) = \exp(-\|\mathbf{u} - \mathbf{v}\|^2 / 2\gamma^2)$. 所有算法都是基于 Pentium 4, 2.83 GHz CPU, Matlab 2007 环境下实现. 实验中使用人工数据集和标准数据集进行测试, 前者的生成方法将在实验中给出, 后者来自 UCI 数据库^[11]和 StatLib 数据库^[12]. 数据的输入和输出都归一化到 $[0, 1]$ 之间, 数据的具体描述见表 1. 其中 Abalone 数据集有一个属性“sex”是由 M, F 和 I 组成, 分别用 1, 2, 3 表示.

表 1 标准数据集

数据集	训练样本数	测试样本数	属性数
Babyfat	130	122	14
Basketball	50	46	4
Bolts	20	20	7
Elusage	25	20	2
fruitfly	60	65	4
Mbgrade	30	31	2
Pollution	30	30	15
Quake	1000	1178	3
Sensory	300	276	11
Stock	200	750	9
Strike	300	325	6
Space-ga	1500	1607	6
Abalone	2000	2177	8
Housing	250	256	13

4.1 参数设置

NUSVR 和 NBNUSVR 算法均有 3 个参数需要调节: ν , C 和 γ . 根据文献[8-9], 可以在固定参数 C 的情况下调节参数 ν 和 γ 以找到最佳的泛化性能. 经过反复测试, 参数 C 设为 1000 时两种算法的泛化性能较好. 参数 ν 按照文献[2]的选择方法, 分别取 0.1, 0.2, \dots , 1; 参数 γ 按照文献[13]的选择方法, 分别取 0.001, 0.01, 0.1, 0.2, 0.4, 0.8, 1, 2, 5, 10, 20, 50, 100, 1000, 10000. 组合 (ν, γ) 共有 150 种参数对. 对于每个数据集采用 50 次测试, 每次测试随机产生表 1 描述的训练样本数目和测试样本数目. 实验评价指标包括平均训练时间和平均测试均方根误差 (RMSE).

4.2 人工数据集测试

人工数据集训练样本的输入 \mathbf{x} 和输出 \mathbf{y} 由下面方法产生:

$$\mathbf{x}_1 = \mathbf{x}_2 = 2 \text{randn}(100, 1), \mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2],$$

$$\mathbf{y} = \exp(-2(\mathbf{x}_1^2 + \mathbf{x}_2^2)).$$

测试样本由下面方法产生:

$$\mathbf{x}_{\text{test}} = \text{meshgrid}([-3 : 0.1 : 3], [-3 : 0.1 : 3]).$$

NUSVR 算法和 NBNUSVR 算法的参数对 (ν, γ) 均选为 (0.1, 0.4), 图 1 和图 2 分别为两种算法的测试结果.

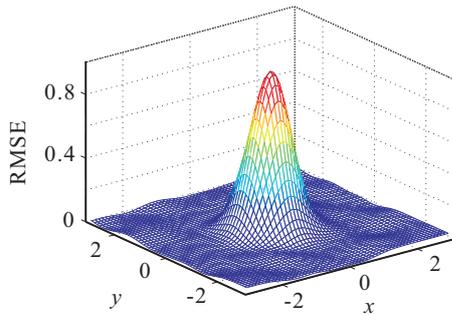


图 1 NUSVR 的测试性能

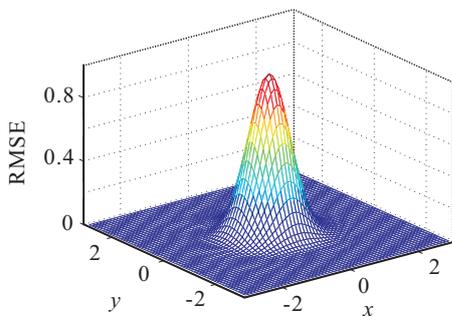


图 2 NBNUSVR 的测试性能

由图可以看出, NBNUSVR 算法在圆锥形外的测试样本上的 RMSE 曲面非常光滑, 且 RMSE 值基本趋于一致; 而 NUSVR 的 RMSE 曲面波动较大, 且 RMSE 值没有规律。

4.3 标准数据集测试

表 2 为 NUSVR 与 NBNUSVR 在标准数据集上的测试结果比较。在 50 次随机测试中, 除了 Strike 数据集, NBNUSVR 在其他 13 个数据集上的平均 RMSE 值都较低。两个算法在所有数据集上的训练时间相差不大。可以看出 NUSVR 与 NBNUSVR 的最优参数对并不总是一致, 这说明它们在解空间中搜索最优

表 2 NUSVR 与 NBNUSVR 在标准数据集上的泛化性能比较

数据集	NUSVR			NUSVR		
	(ν, γ)	Time/s	RMSE	(ν, γ)	Time/s	RMSE
Babyfat	(0.8, 5)	0.244 8	0.025 3	(0.8, 5)	0.194 7	0.023 7
Basketball	(0.5, 10)	0.030 3	0.131 5	(0.5, 10)	0.031 7	0.129 4
Bolts	(0.4, 5)	0.025 0	0.144 9	(0.4, 5)	0.021 7	0.140 9
Elusage	(0.7, 1)	0.029 9	0.137 9	(0.8, 1)	0.028 0	0.133 7
fruitfly	(0.7, 1)	0.045 0	0.208 5	(0.7, 5)	0.056 6	0.200 5
Mbgrade	(0.8, 2)	0.027 5	0.227 9	(0.8, 20)	0.028 0	0.215 9
Pollution	(0.3, 10)	0.035 0	0.139 1	(0.3, 10)	0.019 2	0.135 2
Quake	$(0.3, 10^3)$	4.900 8	0.174 1	$(0.3, 10^3)$	4.530 9	0.173 4
Sensory	$(0.3, 10^{-3})$	1.168 1	0.165 0	$(0.2, 10^4)$	0.153 2	0.164 4
Stock	(0.2, 0.8)	0.173 6	0.039 7	(0.6, 0.8)	0.373 5	0.038 7
Strike	(0.3, 1)	0.561 7	0.077 5	(0.2, 10)	0.292 8	0.079 3
Space-ga	(0.5, 1)	39.70	0.038 6	(0.5, 1)	39.65	0.038 0
Abalone	(0.6, 0.4)	92.22	0.077 7	(0.3, 0.4)	100.15	0.076 7
Housing	(0.7, 5)	0.256 4	0.084 9	(0.2, 1)	0.251 6	0.079 2

解的方向不一样。这个结论也可以由前面 ν -SVR 与无偏置的 ν -SVR 对偶优化问题得出, 即两个对偶优化问题的约束条件不同, 其寻找最优解的方向也不同。

4.4 ν -SVR 与无偏置 ν -SVR 对偶优化问题最优解的比较

ν -SVR 对偶优化问题比无偏置 ν -SVR 对偶优化问题多一个约束条件 $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$ 。为了分析该约束条件对求解最小化目标函数值的影响, 本实验将进行两个优化问题目标函数最优值的比较。为了公平, 两个目标函数中的核矩阵必须一致, 在此将两个算法中的参数对一致设为 NUSVR 中的最优参数对 (见表 2)。在实验中, 每次顺序选取各个数据集中如表 1 描述的训练样本数目作为训练样本集。

表 3 从对偶优化问题求解的角度给出了有约束条件 $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$ 的最优值和无该约束条件的最优值比较。除了 Strike 数据集, 无偏置 ν -SVR 总能得到更优的目标函数值。在 Strike 数据集上, ν -SVR 与无偏置 ν -SVR 的对偶优化问题得到了相同的最优函数值, 而由表 2 可见, 前者的 RMSE 值更低, 这是由于对偶间隙的存在导致原始优化问题的最优解搜索方向不同 (最优参数对不一致)。总体上可以看出, 约束条件 $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$ 限制目标函数得到了更优的解。从前文的推导中可知, 该约束条件恰好是由偏置 b 引出的, 所以偏置 b 并不应包含在 ν -SVR 优化问题中, ν -SVR 的回归超平面在高维特征空间中应通过原点。

表 3 ν -SVR 与无偏置 ν -SVR 最优目标函数值的比较

数据集	NUSVR	NBNUSVR
Babyfat	-1.741 0e-005	-1.786 4e-005
Basketball	-7.440 1e-005	-7.449 4e-005
Bolts	-6.842 6e-005	-6.848 6e-005
Elusage	-6.582 0e-005	-6.726 6e-005
fruitfly	-1.167 3e-004	-1.172 8e-004
Mbgrade	-1.454 3e-004	-1.895 1e-004
Pollution	-4.991 8e-005	-5.001 5e-005
Quake	-7.447 8e-005	-7.454 5e-005
Sensory	-4.517 9e-005	-4.827 5e-005
Stock	-8.427 3e-006	-8.521 1e-006
Strike	-2.323 5e-005	-2.323 5e-005
Space-ga	-2.336 8e-005	-2.446 8e-005
Abalone	-4.762 9e-005	-4.814 8e-005
Housing	-5.943 1e-005	-5.986 4e-005

5 结 论

本文从优化对偶问题的角度分析了偏置 b 对 ν -SVR 泛化性能的影响, 提出了无偏置的 ν -SVR 优化问题, 并将其与 ν -SVR 的优化问题进行了对比分析。实验结果表明, 没有偏置 b 的 ν -SVR 对偶目标函数值

能得到更小的值,且无偏置 ν -SVR 的泛化性能好于传统的 ν -SVR. 本文工作可推广到 ν -SVR 的一些改进算法中. 进一步的研究工作是将无偏置的 ν -SVR 应用于日常生活中,如图像处理 and 模型检测等.

参考文献(References)

- [1] Huang G B, Ding X J, Zhou H M. Optimization method based extreme learning machine for classification[J]. Neurocomputing, 2010, 74(1-3): 155-163.
- [2] Schölkopf B, Smola A J, Williamson R, et al. New support vector algorithms[J]. Neural Computation, 2000, 12(5): 1207-1245.
- [3] Zhao Y P, Sun J G. Rough ν -support vector regression[J]. Expert Systems with Applications, 2009, 36(6): 9793-9798.
- [4] Zhao Y P, Sun J G, Zou X Q. Reducing samples for accelerating multikernel semiparametric support vector regression[J]. Expert Systems with Applications, 2010, 37(6): 4519-4525.
- [5] Poggio T, Mukherjee S, Rifkin R, et al. *b*[R]. USA: Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2001.
- [6] Evgeniou T, Pontil M, Poggio T. Regularization networks and support vector machines[J]. Advances in Computational Mathematics, 2000, 13(1): 1-50.
- [7] Kecman V, Vogt M, Huang T M. On the equality of kernel adatron and sequential minimal optimization in classification and regression tasks and alike algorithms for kernel machines[C]. Proc of the 11th European Symposium on Artificial Neural Networks. Bruges: Springer, 2003: 215-222.
- [8] Chang C C, Lin C J. Training ν -support vector regression: Theory and algorithms[J]. Neural Computation, 2002, 14(8): 1959-1977.
- [9] Chalimourda A, Schölkopf B, Smola A J. Experimentally optimal ν in support vector regression for different noise models and parameter settings[J]. Neural Networks, 2004, 17(1): 127-141.
- [10] Scheinberg K. An efficient implementation of an active set method for SVMs[J]. J of Machine Learning Research, 2006, 7(10): 2237-2257.
- [11] Blake C K, Merz C J. UCI machine learning repository[EB/OL]. (1998-04-02)[2009-02-12]. <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- [12] Michie D, Spiegelhalter D J, Taylor C C. Machine learning, neural and statistical classification[EB/OL]. (2001-02-19)[2009-02-12]. <ftp://ftp.ncc.up.pt/pub/stalog/>.
- [13] Ghanty P, Paul S, Pal N R. NEUROSVN: An architecture to reduce the effect of the choice of kernel on the performance of SVM[J]. J of Machine Learning Research, 2009, 10(3): 591-622.

(上接第865页)

- [5] Tiryaki F. Interactive compensatory fuzzy programming for decentralized multi-level linear programming(DMLLP) problems[J]. Fuzzy Sets and Systems, 2006, 157(23): 3072-3090.
- [6] Ahlatcioglu M, Tiryaki F. Interactive fuzzy programming for decentralized two-level linear fractional programming (DTLLFP) problems[J]. Omega, 2007, 35(4): 432-450.
- [7] 李春好, 杜元伟. 不确定环境下的两层交互式有限理性决策方法[J]. 系统工程理论与实践, 2010, 30(11): 2003-2012.
(Li C H, Du Y W. Interactive bounded rationally approach to two level decision making under uncertainty[J]. System Engineering-Theory & Practice, 2010, 30(11): 2003-2012.)
- [8] 胡晓翔, 达庆利, 徐泽水. 知识管理的风险分析、评估与控制[J]. 管理科学学报, 2003, 6(5): 17-23.
(Hu X X, Da Q L, Xu Z S. Risk analysis, appraisal and control on knowledge management[J]. J of Management Sciences in China, 2003, 6(5): 17-23.)
- [9] 吕卫文. 隐性知识和编码知识[J]. 科研管理, 2007, 28(6): 31-35.
(Lv W W. The tacit knowledge and the codified knowledge[J]. Science Research Management, 2007, 28(6): 31-35.)
- [10] 李春好, 杜元伟, 刘成明, 等. 基于基元前景交叉判断的前景价值模型[J]. 管理科学学报, 2010, 13(2): 11-23.
(Li C H, Du Y W, Liu C M, et al. Prospect value model via cross judgments of basic prospects[J]. J of Management Sciences in China, 2010, 13(2): 11-23.)
- [11] Lai Y J, Hwang C L. Fuzzy multiple objective decision making: Methods and applications[M]. Berlin: Springer, 1996: 30-35.