

具有多分段损失函数的多输出支持向量机回归

胡根生¹, 邓飞其²

(1. 安徽大学 电子科学与技术学院, 安徽 合肥 230039; 2. 华南理工大学 自动化科学与工程学院, 广东 广州 510640)

摘要: 对多维输入、多维输出数据的回归, 可以采用多输出支持向量机回归算法. 本文介绍具有多分段损失函数的多输出支持向量机回归, 其损失函数对落在不同区间的误差值采用不同的惩罚函数形式, 并利用变权迭代算法, 给出回归函数权系数和偏置的迭代公式. 仿真实验表明, 该算法的精确性和计算工作量都优于使用多个单输出的支持向量机回归算法.

关键词: 支持向量机; 损失函数; 多输出回归

中图分类号: TP18 **文献标识码:** A

Multi-output support vector regression with piecewise loss function

HU Gen-sheng¹, DENG Fei-qi²

(1. School of Electronic Science and Technology, Anhui University, Hefei Anhui 230039, China;

2. College of Automation Science and Technology, South China University of Technology, Guangzhou Guangdong 510640, China)

Abstract: Multi-output support vector regression (MSVR) algorithm can be used for the regression of multi-input multi-output data. An MSVR with piecewise loss function is proposed in this paper. For error values on different interval, the loss function adopts different forms of penalty functions. By using re-weight iterative algorithm, the iterative formulas of weight coefficients and bias of regression function are then given. Experiments also show that the accuracy and workload of this algorithm are superior to that using several single-output SVR algorithms.

Key words: support vector machine; loss function; multi-output regression

1 引言(Introduction)

支持向量机(SVM)是由Vapnik等人提出的一种基于统计学习理论和结构风险最小化原理的新型学习机^[1]. 目前的支持向量机回归算法主要是多输入、单输出的, 即给定一组数据 $\{(x_i, y_i), i = 1, \dots, n\}$, 这里 $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$, 构造线性回归函数 $f \in F, F = \{f|f: \mathbb{R}^d \rightarrow \mathbb{R}\}$ 是一个函数类, 使得结构风险

$$R_{\text{reg}}(f) = R_{\text{emp}}(f) + C\Omega(f) \quad (1)$$

最小, 其中: $R_{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i))$ 是经验风险函数, $\Omega(f)$ 是提高函数泛化能力的正则项.

实际生活中经常会出现多输出函数的回归问题, 这些多输出之间有时并不是相互独立的, 如果用多个单输出支持向量机回归算法来估计多输出函数, 则不能平等地对待每个样本点, 因而精确性较差. 本文利用F.Perez-Cruz等人提出的变权迭代算法^[2], 求解多输出支持向量机回归问题(MSVR), 提高估计的精确性, 降低多维回归问题的计算工作量.

利用变权迭代算法求解MSVR时, 必须选择目标函数中恰当的损失函数. 标准的支持向量机回归模型采用Vapnik定义的 ε -不敏感损失函数, 这是一种线性惩罚函数, 对于误差小于 ε 的样本点不予惩罚. Suykens等人提出的最小方差支持向量机(LS-SVM)^[3]用平方损失函数代替 ε -不敏感损失函数, 其约束条件就由等式约束代替了不等式约束, 求解这类问题就变成了求解对偶空间中的一组线性方程, 降低了计算复杂性, 但同时也降低了它的抗噪声能力, 如果噪声分布具有长尾, 则解的结果会由于少量噪声的影响而与实际值有很大的偏离. 除此之外, Laplace损失函数用误差项的绝对值作为惩罚函数, Huber损失函数把Laplace损失函数和二次损失函数结合了起来^[4], 这几种损失函数不再具有 ε -不敏感损失函数解的稀疏性. 本文结合 ε -不敏感损失函数和Huber损失函数的优点, 采用多分段损失函数的形式, 在不同的区间段内, 损失函数具有不同的形式, 这样既保持了解的稀疏性, 又增强了模型的抗噪声能力, 提高了估计的精确度.

2 具有多分段损失函数的多输出支持向量机回归算法(MSVR algorithm with piecewise loss function)

给定数据集 $\{(x_i, y_i), i=1, \dots, n\}$, 这里 $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}^m$. 多输出情况下, 要构造函数 $f^j(x) = \Phi^T(x)w^j + b^j, j=1, \dots, m$, 使式(1)的结构风险最小化.

根据支持向量机理论, 多输出情况下, 最小化式(1)等价于下面的约束优化问题^[5]:

$$\begin{aligned} \min \quad & \sum_{j=1}^m \|w^j\|^2 + C \sum_{i=1}^n L(\xi_i), \\ \text{s.t.} \quad & \|y_i - \Phi^T(x_i)w - b\| \leq \varepsilon + \xi_i, \\ & \xi_i \geq 0, i=1, \dots, n. \end{aligned} \quad (2)$$

这里 $L(\xi_i)$ 是损失函数, $y_i = (y_i^1, y_i^2, \dots, y_i^m)$, $w = (w^1, w^2, \dots, w^m)$, $b = (b^1, b^2, \dots, b^m)$.

式(2)定义了一个 ε -超球不敏感域, 当实际值与估计值之差的欧几里得范数小于 ε 时, 即样本点位于多维输出估计的 ε -超球不敏感域内, 损失函数的惩罚值为0, 如果样本点位于多维输出估计的 ε -超球不敏感域之外时, 损失函数按同样的因子系数 C 进行惩罚.

如果采用 m 个单输出的支持向量回归来代替多输出支持向量回归, 则目标函数里的损失函数不能平等地对待多输出回归中的每一个样本点.

式(2)需要事先定义一个损失函数, 本文结合 ε -不敏感损失函数和Huber损失函数的优点, 采用多分段损失函数的形式, 其损失函数定义为:

$$L(\delta_i) = \begin{cases} -\delta_i - \varepsilon, & \delta_i \in (-\infty, -\beta_p \varepsilon), \\ \vdots & \vdots \\ (-\delta_i - \varepsilon) \prod_{j=l}^p \frac{(-\delta_i - \varepsilon)}{(\beta_j - 1)\varepsilon}, & \delta_i \in [-\beta_l \varepsilon, -\beta_{l-1} \varepsilon], \\ \vdots & \vdots \\ 0, & \delta_i \in [-\beta_0 \varepsilon, \beta_0 \varepsilon], \\ \vdots & \vdots \\ (\delta_i - \varepsilon) \prod_{j=l}^p \frac{(\delta_i - \varepsilon)}{(\beta_j - 1)\varepsilon}, & \delta_i \in (\beta_{l-1} \varepsilon, \beta_l \varepsilon], \\ \vdots & \vdots \\ \delta_i - \varepsilon, & \delta_i \in (\beta_p \varepsilon, +\infty). \end{cases} \quad (3)$$

这里 $\delta_i = y_i - f(x_i)$, $1 = \beta_0 \leq \beta_1 \leq \dots \leq \beta_p < +\infty$.

损失函数式(3)对落在越小区间值的误差项, 其惩罚值也越小, 且小于相应的线性惩罚函数. 实际应用表明^[6], 三阶及以上的损失函数区域适合于次高斯噪声, 二阶损失函数区域适合于高斯噪声, 线性损失函数区域适合于超高斯噪声.

当 $\beta_p = 1$ 时, 上述损失函数等价于 ε -不敏感损失函数, 当 $\beta_{p-1} = 1$ 时, 损失函数式(3)等价于下述函数形式

$$L(\delta) = \begin{cases} -\delta - \varepsilon, & \delta \in (-\infty, -\beta\varepsilon), \\ \frac{(-\delta - \varepsilon)^2}{(\beta - 1)\varepsilon}, & \delta \in [-\beta\varepsilon, -\varepsilon], \\ 0, & \delta \in [-\varepsilon, \varepsilon], \\ \frac{(\delta - \varepsilon)^2}{(\beta - 1)\varepsilon}, & \delta \in (\varepsilon, \beta\varepsilon], \\ \delta - \varepsilon, & \delta \in (\beta\varepsilon, +\infty), \end{cases} \quad \beta > 1. \quad (4)$$

式(4)中, ε -不敏感区域 $\delta \in [-\varepsilon, \varepsilon]$ 适用于低频变化, 例如游荡或基线偏离, 二次损失函数域 $|\delta| \in (\varepsilon, \beta\varepsilon]$ 适用于处理高斯噪声, 线性惩罚函数区域限制异常点或颤动噪声对模型的影响. 在SVR的贝叶斯模型选择方法中^[7,8], 噪声分布常常被假定为具有下述指数分布形式: $P(\delta_i) \propto e^{[-CL(\xi_i)]}$, 这里 $\xi_i = \max(0, |y_i - f(x_i)| - \varepsilon)$. 如果使用式(4)的损失函数, 对于落在二次损失函数域 $|\delta| \in (\varepsilon, \beta\varepsilon]$ 的误差项, 则 $P(\delta_i) \propto e^{(-C \frac{\delta_i^2}{(\beta-1)\varepsilon})}$. 如果噪声分布是均值为0、方差为 r^2 的高斯分布, 则 $\beta \sim \frac{2Cr^2}{\varepsilon} + 1$, 由此可以看出, 恰当的 β 值是与噪声分布的方差有密切的关系.

将式(2)的损失函数采用式(3)的形式, 则式(2)中的损失函数

$$L(\xi_i) = \begin{cases} 0, & \xi_i = 0, \\ \vdots & \vdots \\ \xi_i \prod_{j=l}^p \frac{\xi_i}{(\beta_j - 1)\varepsilon}, & \xi_i \in ((\beta_{l-1} - 1)\varepsilon, (\beta_l - 1)\varepsilon], \\ \vdots & \vdots \\ \xi_i, & \xi_i \in ((\beta_p - 1)\varepsilon, +\infty). \end{cases} \quad (5)$$

损失函数式(5)具有多分段的形式, 该模型可以利用F.Perez-Cruz等人提出的变权迭代算法来求解, 算法的收敛性已经由F.Perez-Cruz等人给出^[9].

定义Lagrange乘子

$$L = \sum_{j=1}^m \|w^j\|^2 + C \sum_{i=1}^n L(\xi_i) - \sum_{i=1}^n \alpha_i [(\varepsilon + \xi_i)^2 - \|y_i - \Phi^T(x_i)w - b\|^2] - \sum_{i=1}^n \mu_i \xi_i. \quad (6)$$

由KKT条件得到

$$\frac{\partial L}{\partial w^j} = 2w^j - 2\Phi^T D_\alpha [y^j - \Phi w^j - \mathbf{1}b^j] = 0. \quad (7)$$

这里: $\Phi^T = (\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n))$, $y^j = (y_1^j, y_2^j, \dots, y_n^j)^T$, $D_\alpha = \text{diag}\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, $\mathbf{1} = (1, 1, \dots, 1)^T$.

$$\frac{\partial L}{\partial b^j} = \alpha^T [y^j - \Phi w^j - \mathbf{1}b^j] = 0 \quad (8)$$

这里 $\alpha^T = (\alpha_1, \alpha_2, \dots, \alpha_n)$.

$$\frac{\partial L}{\partial \xi_i} = Ca_i - \mu_i - 2\alpha_i(\varepsilon + \xi_i) = 0, \quad (9)$$

这里:

$$a_i = \begin{cases} (p-l+2) \prod_{j=l}^p \frac{\xi_i}{(\beta_j-1)\varepsilon}, & \xi_i \in ((\beta_{l-1}-1)\varepsilon, (\beta_l-1)\varepsilon], \\ \vdots & \vdots \\ 1, & \xi_i \in ((\beta_p-1)\varepsilon, +\infty). \end{cases} \quad (10)$$

$$\mu_i, \alpha_i \geq 0, \quad (10)$$

$$\alpha_i[(\varepsilon + \xi_i)^2 - \|y_i - \Phi^T(x_i)w - b\|^2] = 0, \quad (11)$$

$$\mu_i \xi_i = 0, \quad (12)$$

$$\|y_i - \Phi^T(x_i)w - b\| - \varepsilon - \xi_i \leq 0. \quad (13)$$

由式(7)(8)我们得到

$$\begin{bmatrix} \Phi^T D_\alpha \Phi + I & \Phi^T \alpha \\ \alpha^T \Phi & \alpha^T \mathbf{1} \end{bmatrix} \begin{bmatrix} w^j \\ b^j \end{bmatrix} = \begin{bmatrix} \Phi^T D_\alpha y^j \\ \alpha^T y^j \end{bmatrix}. \quad (14)$$

根据表示定理^[10]

$$w^j = \sum_{i=1}^n \gamma_i^j \Phi(x_i) = \Phi^T \gamma^j, \quad (15)$$

这里 $\gamma^j = (\gamma_1^j, \gamma_2^j, \dots, \gamma_n^j)^T$.

将式(15)代入式(7), 并在方程的两边同时乘以 $\Phi^T D_\alpha$ 的广义逆 $(\Phi^T D_\alpha)^+$, 得到 $[\Phi \Phi^T + D_\alpha^+] \gamma^j = [y^j - \mathbf{1}b^j]$, 如果某个正定核函数 K 满足 Mercer 定理的条件(比如高斯核函数), 就能定义 $K = \Phi \Phi^T$, 其中 $K_{ij} = K(x_i, x_j) = \Phi^T(x_i) \Phi(x_j)$, 因而式(14)变为

$$\begin{bmatrix} K + D_\alpha^+ & \mathbf{1} \\ \alpha^T K & \alpha^T \mathbf{1} \end{bmatrix} \begin{bmatrix} \gamma^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ \alpha^T y^j \end{bmatrix}, \quad (16)$$

所以

$$\begin{bmatrix} \gamma^j \\ b^j \end{bmatrix} = \begin{bmatrix} K + D_\alpha^+ & \mathbf{1} \\ \alpha^T K & \alpha^T \mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} y^j \\ \alpha^T y^j \end{bmatrix}. \quad (17)$$

记 $e_i = \|y_i - \Phi^T(x_i)w - b\|$, 根据式(9)~(13)可得

$$\alpha_i = \begin{cases} 0, & e_i \leq \varepsilon, \\ \frac{Ca_i}{2e_i}, & e_i > \varepsilon. \end{cases} \quad (18)$$

这里

$$a_i = \begin{cases} (p-l+2) \prod_{j=l}^p \frac{e_i - \varepsilon}{(\beta_j - 1)\varepsilon}, & e_i \in (\beta_{l-1}\varepsilon, \beta_l\varepsilon], \\ \vdots & \vdots \\ 1, & e_i \in (\beta_p\varepsilon, +\infty). \end{cases}$$

综上所述, 计算这种具有多分段损失函数的多输出支持向量机回归的步骤为:

第1步 设置初始值 γ^j , b^j , e_i 和参数值 C , σ , ε , β_l , 这里 γ^j 和 b^j 可以设置初始值为 0, e_i 可以设置初始值为一个较大的数.

第2步 对于每个 $i = 1, \dots, n$, 计算式(18)中 α_i 的值.

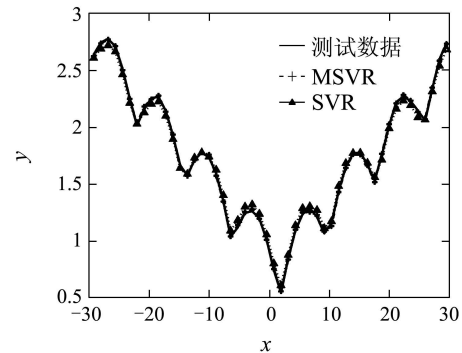
第3步 利用式(17)计算 γ^j , b^j , 更新 e_i 的值.

第4步 如果所有的 $e_i \leq \varepsilon$, $i = 1, \dots, n$, 则停止计算, 否则返回第2步.

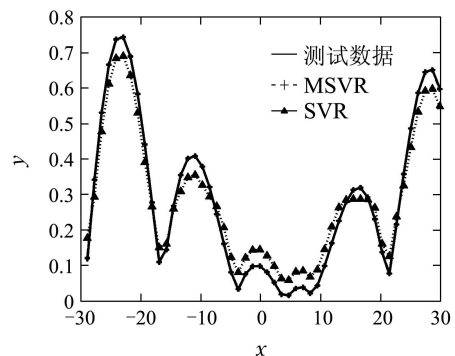
3 实验(Experiments)

本节进行仿真实验, 分析前面介绍的 MSVR 算法的性能. 实验是在 P4 1.49G CPU, 256MB 内存, Windows XP 环境下, MATLAB 6.5 上进行的, 核函数采用高斯核 $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$, 损失函数采用式(4)所定义的分段函数.

取数据 $\{(x_i, y_i) \in \mathbb{R} \times \mathbb{R}^3, i = 1, \dots, 150\}$, 其中输入数据 $x_i \in [-30, 30]$, 三维输出数据为: (a) $y_i^1 = 0.125|0.5x_i - 1| + 0.5|\sin[\pi(0.75 + 0.125x_i)]| + 0.5 + r_i$, (b) $y_i^2 = |0.25x_i - 1|[1 + 10|\sin(0.25x_i + 1)|]/100 + r_i$, (c) $y_i^3 = 2\sin(x_i/2)/x_i + r_i$, r_i 为服从均值为 0、标准差为 0.01 的正态分布的随机噪声. 从中任取 100 组数据作为训练数据, 其余的 50 组数据作为测试数据, 分别用本文介绍的 MSVR 算法和 3 个标准的单输出 SVR 算法进行训练和测试, 测试结果如图 1 所示, 训练和测试结果的均方误差如表 1 所示, 所用的参数为 $C = 1$, $\sigma = 1$, $\varepsilon = 0.05$, $\beta = 2$.



(a)



(b)

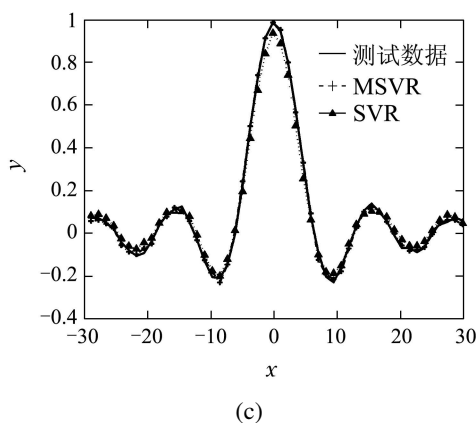


图1 用MSVR和单输出SVR算法所得到的测试结果比较
Fig. 1 Comparison of test results based on MSVR and single-output SVR

表1 利用MSVR和单输出SVR算法分别进行训练和测试的均方误差

Table 1 Mean-square errors of training and testing based on MSVR and single-output SVR

	(a)	(b)	(c)
SVR训练	0.0034	0.0035	0.0022
SVR测试	0.0066	0.0071	0.0047
MSVR训练	5×10^{-12}	9×10^{-13}	8×10^{-12}
MSVR测试	0.0008	0.0002	0.0009

MATLAB运行上述MSVR算法的时间为8.063 s, 运行3个单输出SVR算法的时间分别为58.671 s, 61.734 s, 58.985 s.

改变噪声 r_i 的标准差, 对应于不同的 β 值, MSVR的测试均方误差见表2所示, 其中均方误差是进行10次实验的平均值, 其余参数和前面的实验相同.

从表2可以看出, 如果选取 $\beta \sim 2Cr^2/\varepsilon + 1$, 则均方误差要小一些, 因而可以选取恰当的 β 值来提高预测的精确性.

表2 对应于不同的 β 值和噪声标准差, MSVR的测试均方误差

Table 2 Mean-square errors of testing based on MSVR for different β values and standard deviation of noise

	$r_i = 0.01$	$r_i = 0.1$	$r_i = 0.5$
$\beta = 1.01$	0.00021	0.0183	0.5366
$\beta = 1.2$	0.00022	0.0177	0.5778
$\beta = 5.01$	0.00023	0.0188	0.4829
$\beta = 10.01$	0.00025	0.0191	0.5335

4 结论(Conclusions)

从上述的仿真实验可以看出, 对于多维输出数据, 本文介绍的具有多分段损失函数的MSVR算法的性能比标准的单输出SVR算法要稍好一些, 而且其计算工作量也比使用多个单输出SVR算法要小. 在使用该MSVR算法时, 调整损失函数里不同的 β_i 值, 对计算结果有一定的影响, 如果噪声分布是高斯分布, 不同的 β_i 值适用于不同方差的噪声分布. 同时可以看到, 用MSVR做预测, 测试结果的波动比较大, 如何减小其波动性, 使之具有更好的泛化性能, 是今后要研究的重点之一.

参考文献(References):

- [1] VAPNIK V. *Statistical Learning Theory*[M]. New York: Wiley, 1998.
- [2] PEREZ-CRUZ F, NAVIA-VAZQUEZ A, ALARCON-DIANA P L, et al. An IRWLS procedure for SVR[C] // *Proc of the European Signal Processing Conf.* Tampere, Finland: [s.n.], 2000.
- [3] SUYKENS J A K, VANDEWALLE J. Multiclass least squares support vector machines[C] // *Int Joint Conf on Neural Networks*. Washington: [s.n.], 1999: 900 – 903.
- [4] HUBER P J. *Robust Statistics*[M]. New York: Wiley, 1981.
- [5] PEREZ-CRUZ F, CAMPS-VALLS G, SORIA-OLIVAS E, et al. Multi-dimensional function approximation and regression estimation[C] // *Proc of Int Conf on Artificial Neural Networks*. Berlin: Springer-Verlag, 2002, 2415: 757 – 762.
- [6] ROJO-ALVAREZ J, MARTINEZ-RAMON M, PRADO-CUMPLIDO M. Support vector method for robust ARMA system identification[J]. *IEEE Trans on Signal Processing*, 2004, 52(1): 155 – 164.
- [7] CHU W, KEERTHI S S, ONG C J. Bayesian support vector regression using a unified loss function[J]. *IEEE Trans on Neural Networks*, 2004, 15(1): 29 – 44.
- [8] YAN W, SHAO H, WANG X. Soft sensing modeling based on support vector machine and bayesian model selection[J]. *Computers & Chemical Engineering*, 2004, 28(8): 1489 – 1498.
- [9] PEREZ-CRUZ F, BOUSONO-CALZON C, ARTES-RODRIGUEZ A. Convergence of the IRWLS procedure to the support vector machine solution[J]. *Neural Computation*, 2005, 17(1): 7 – 18.
- [10] SCHOLKOPT B, SMOLA A J. *Learning with Kernels*[M]. Cambridge, MA: The MIT Press, 2002.

作者简介:

胡根生 (1971—), 男, 博士, 主要研究方向为支持向量机理论及应用, E-mail: hugs2906@sina.com;

邓飞其 (1962—), 男, 教授, 博士生导师, 主要研究方向为大系统、随机系统的控制理论、系统工程, 支持向量机及数理金融理论, E-mail: aufqdeng@163.com.