第 19 卷 第 12 期 控 制 与 决 策 2004 年 12 月
Vol 19 No. 12 Control and Decision Dec 2004

文章编号: 1001-0920(2004)12-1383-04

鲁棒 r-支持向量回归机中参数 r 的选择研究

朱嘉钢[1,2], 王士同[2,3,4], 杨静宇[1]

(1. 南京理工大学 计算机科学与工程系, 江苏 南京 210094; 2. 江南大学 信息工程学院, 江苏 无锡 214036; 3. 南京大学 软件新技术国家重点实验室, 江苏 南京 210016; 4. 中国科学院 软件研究所 计算机科学重点实验室, 北京 100080)

摘　要: 为使 r-支持向量回归机更具鲁棒性, 研究了 r-支持向量回归机 r-SVR 中参数 r 与输入噪声之间的关系. 将 r-SVR 的优化问题转换成最大后验估计问题, 推导出 r-SVR 后验估计最大化的条件, 得出了输入噪声为高斯分布时 r-SVR 中参数 r 与 σ 之间的近似线性反比关系, 为已知输入高斯噪声方差 σ 时合理选择 r 提供了理论依据.

关键词: 支持向量机; 支持向量回归机; r 范数损失函数

中图分类号: TP18; TP391　　　**文献标识码**: A

On the choice of the parameter r in robust r-support vector regression

ZHU Jia-gang[1,2], *WANG Shi-tong*[2,3,4], *YANG Jing-yu*[1]

(1. Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; 2. School of Information Engineering, Southern Yangtz University, Wuxi 214036, China; 3. The National Key Lab of Novel Software, Nanjing University, Nanjing 210016, China; 4. The Key Lab of Computer Science, Institute of Software, Chinese Academy of Science, Beijing 100080, China. Correspondent: WANG Shi-tong, E-mail: wxwangst@yahoo.com.cn)

Abstract: The dependency relationship between r and the input noise in r-SVR is studied using the idea that SVR can be interpreted into an equivalent MAP problem. The dependency relationships are derived by maximizing the posteriori estimation. Accordingly, focus is paid on the case of Gaussian noise, and the linear inverse proportional dependency between r and the variance of Gaussian noise is then derived. Such a dependency relationship is useful to determine the optimal choice for r in r-loss function in the existence of Gaussian noise.

Key words: support vector machines (SVM); support vector regression (SVR); r-loss function

1 引　　言

近年来, 支持向量机(SVM) 在各种分类问题和回归(又可称为建模) 问题中得到广泛应用. 各种 SVM 通常基于 3 类损失函数, 即 ϵ-不敏感损失函数, Huber 损失函数[1,2]和 r 范数损失函数. 其中, r 范数损失函数当 $r=2$ 时为平方损失函数, 而当 $r=1$ 时则为Laplace 损失函数.

当 SVM 用于建模时, 通常称其为支持向量回归机(SVR). 由于真实数据经常含有噪声, 在用 SVR 进行建模时, 如何选择合适的参数使SVR 具有更好的鲁棒性便成为一个重要课题. Kwok 等人和本文作者最近研究了当噪声已知时, ϵ-支持向量回

收稿日期: 2004-01-05; **修回日期**: 2004-04-02

基金项目: 国家自然科学基金资助项目(60225015); 江苏省自然科学基金资助项目(BK2003017); 江苏省计算机技术重点实验室开放课题资助项目.

作者简介: 朱嘉钢(1957—), 男, 上海人, 博士生, 副教授, 从事人工神经网络、模式识别等研究; 王士同(1964—), 男, 江苏扬州人, 教授, 博士生导师, 从事人工神经网络、神经模糊系统等研究.

归机[3,4]和Huber-支持向量回归机参数的最佳选择方法. 但对于噪声已知时, r 范数-支持向量回归机中参数 r 的选择问题目前尚未见到研究结果. r 范数损失函数统一了Laplace 损失函数和平方损失函数, 因此, 从理论上研究 r 范数损失函数中参数 r 与输入噪声之间的关系, 具有普遍的理论意义和实际指导意义.

本文研究 r 范数-支持向量回归机中参数 r 的选择问题. 基于将SVR 的优化问题转换成最大后验估计问题的思想, 推导出 r 范数-支持向量回归机后验估计最大化的条件, 研究了输入噪声为高斯噪声的情形, 并求出了该情形下鲁棒的 r 范数-支持向量回归机中参数 r 的选择方法.

2 r-SVR 与最大后验估计

2.1 r 范数-支持向量回归机 r-SVR

首先给定 r 范数-支持向量回归机 r-SVR 的基本概念. 考虑用线性函数

$$f(x) = \langle w, x\rangle + b$$

来逼近下列数据集:

$$D = \{(x^1, y^1), \cdots, (x^n, y^n)\}, x \in R^d, y \in R,$$

其中 n 为样本数目. 若使用 r 范数损失函数

$$L_r(f(x) - y) = |f(x) - y|^r,$$

则最优回归函数可通过下式求得:

$$\min \Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_i \xi_i, \tag{1}$$

$$\text{s.t.} \begin{cases} |f(x_i) - y_i|^r \leq \xi_i, i = 1, 2, \cdots, l; \\ \xi_i \geq 0. \end{cases}$$

其中: C 为设定常数, $\xi = (\xi_1, \xi_2, \cdots, \xi_l)$, $\xi_i (i = 1, 2, \cdots, l)$ 表示约束系统输出的松弛变量.

于是, r 范数损失函数与式(1) 共同构成了 r 范数-支持向量回归机 r-SVR.

2.2 最大后验估计问题(MAP)

形如式(1) 的最小化问题可以解释成对于给定值 r 和 β, 求 w 的最大后验估计问题[3].

对于已知数据集

$$D = \{(x_i, y_i)\}_{i=1}^n,$$

考虑用

$$y_i = \tilde{w}^{\mathrm{T}} x_i + \eta, i = 1, 2, \cdots, n \tag{2}$$

估计权参数 $\tilde{w}$ 的回归问题. 这里, $x_i \in \Omega$ 服从 $p(\cdot)$ 分布, η 表示噪声服从 $\phi(\cdot)$ 分布, 则相应的 y 的密度函数可记为 $p(y|x) = \phi(y - \tilde{w}^{\mathrm{T}}x)$. 注意, 式(2) 中也可以用 $\Psi(x_i)$ 代替 x_i, 为简单起见, 这里选用式(2) 的表示法, 但不影响本文的结论.

此外, 对于 r 范数-支持向量回归机, 如果令

$$p(y_i | x_i, w, \beta, r) = C(\beta, r)\exp(-\beta L_r(y_i - w^{\mathrm{T}}x_i)),$$

$$p(w|\alpha) = \sqrt{\frac{\alpha}{2\pi}}\exp\left(-\frac{\alpha}{2}\|w\|^2\right),$$

由Bayes 公式, 有

$$p(w|D, \beta, r) \propto p(D|w, \beta, r)p(w),$$

再对两边取对数, 可得

$$\ln p(w|D, w, \beta, r) = -\frac{\alpha}{2}\|w\|^2 - \beta\sum_{i=1}^n L_r(y_i - w^{\mathrm{T}}x_i) + n\ln C(\beta, r) + \text{常数} \tag{3}$$

与式(1) 比较, 则与之所对应的 r 范数-支持向量回归机 r-SVR 便可解释成式(3) 表示的对于给定参数 r 和 β 求 w 的最大后验估计问题.

式(3) 中, $C(\beta, r)$ 表示归一化系数(当 r 较小时 $\Gamma(1/r) \approx r$), 即

$$C(\beta, r) = \left[\int_{-\infty}^{\infty}\exp(-\beta L_r(t))\mathrm{d}t\right]^{-1} = \frac{r\beta^{1/r}}{2\Gamma(1/r)} \approx \frac{1}{2}\beta^{1/r}.$$

3 r 范数-支持向量回归机后验估计最大化的必要条件

对于式(3), 令

$$E_{XY}(|y - w^{\mathrm{T}}x|^r) = \iint_{\Omega}\int_{-\infty}^{\infty}|y - w^{\mathrm{T}}x|^r p(y|x)p(x)\mathrm{d}y\,\mathrm{d}x,$$

则式(3) 可近似地表示成

$$M(w, \beta, r) = -\frac{1}{2}\alpha\|w\|^2 - \beta n E_{XY}(|y - w^{\mathrm{T}}x|^r) + \frac{n}{r}\ln\beta - n\ln 2 + \text{常数} \tag{4}$$

于是, r 范数-支持向量回归机近似地对应于使 $M(\hat{w}, \beta, r)$ 达到最大, 此时式(4) 对 w 的偏导数应等于 0, 即

$$\alpha\hat{w} + \beta n\left[\iint_{\Omega}\int_{-\infty}^{\hat{w}^{\mathrm{T}}x} r(\hat{w}^{\mathrm{T}}x - y)^{r-1}xp(y|x)\mathrm{d}y - \int_{\hat{w}^{\mathrm{T}}x}^{\infty} r(y - \hat{w}^{\mathrm{T}}x)^{r-1}xp(y|x)\mathrm{d}y\right)p(x)\mathrm{d}x\Big] = 0. \tag{5}$$

使 r 范数-支持向量回归机后验估计最大就是使

$$M(\hat{w}, \beta, r) = -\frac{1}{2}\alpha\|\hat{w}\|^2 - \beta n E_{XY}(|y - \hat{w}^{\mathrm{T}}x|^r) +$$

$$\frac{n}{r}\ln\beta - n\ln 2 + \text{常数} \tag{6}$$

最大. 为求得 $M(\hat{w},\beta,\mu)$ 的最大值, 在式(4) 中, 使 $w=\hat{w}$, 对 β 求偏导数并令其等于 0, 得

$$\frac{\partial M(\hat w,\beta,r)}{\partial\beta} = \frac{\partial M}{\partial \hat w}\cdot\frac{\partial \hat w}{\partial\beta} + \frac{\partial M}{\partial\beta} =$$

$$-nE_{XY}(|y-\hat w^{\mathrm T} x|^r) + \frac{1}{\beta r}n = 0,$$

即

$$E_{XY}(|y-\hat w^{\mathrm T} x|^r) = \frac{1}{\beta r}. \tag{7}$$

同理可得

$$\frac{\partial E_{XY}(|y-\hat w^{\mathrm T} x|^r)}{\partial r} \approx \frac{1}{r^2\beta}\ln\beta \tag{8}$$

于是, 式(5), (7) 和(8) 共同构成了 r 范数 - 支持向量回归机后验估计最大化的必要条件. 对于不同的输入噪声 $\phi(y-\tilde w^{\mathrm T} x)$, 只需令 $p(y|x)=\phi(y-\tilde w^{\mathrm T} x)$ 并代入式(7) 和(8), 即可求得 β, r 与输入噪声的关系.

4 输入噪声为高斯模型时 r 与输入噪声方差的近似线性关系

在实际中, 大量的噪声服从高斯分布. 本节专门研究当输入噪声为高斯型时 r 范数 - 支持向量回归机中参数 r 与噪声方差的关系. 当输入噪声服从方差 σ 的高斯分布时, 即

$$\phi(\eta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\eta^2}{2\sigma^2}\right),$$

则有

$$E_{XY}(|y-\hat w^{\mathrm T} x|^r) =$$

$$\int_\Omega\left(\int_{-\infty}^{\infty}|y-\hat w^{\mathrm T} x|^r p(y|x)p(x)\mathrm{d}y\right)\mathrm{d}x =$$

$$\int_\Omega\left(\int_{-\infty}^{\hat w^{\mathrm T} x}(\hat w^{\mathrm T} x - y)^r p(y|x)\mathrm{d}y +\right.$$

$$\left.\int_{\hat w^{\mathrm T} x}^{\infty}(y-\hat w^{\mathrm T} x)^r p(y|x)\mathrm{d}y\right)p(x)\mathrm{d}x.$$

令

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\tilde w^{\mathrm T} x)^2}{2\sigma^2}\right),$$

$$t = (y-\tilde w^{\mathrm T} x)^2/(2\sigma^2),$$

$$y-\tilde w^{\mathrm T} x \approx \hat w^{\mathrm T} x - \tilde w^{\mathrm T} x = \delta(x),$$

得

$$E_{XY}(|y-\tilde w^{\mathrm T} x|^r) =$$

$$\int_\Omega \frac{1}{\sqrt\pi}\left[\int_{-\infty}^{\frac{\delta(x)}{\sqrt2\sigma}}(\delta(x)-\sqrt2\sigma t)^r\exp(-t^2)\mathrm{d}t +\right.$$

$$\left.\int_{\frac{\delta(x)}{\sqrt2\sigma}}^{\infty}(\sqrt2\sigma t-\delta(x))^r\exp(-t^2)\mathrm{d}t\right]p(x)\mathrm{d}x \approx$$

$$\int_\Omega\frac{1}{\sqrt\pi}\left[\int_{-\infty}^{0}(\delta(x)-\sqrt2\sigma t)^r\exp(-t^2)\mathrm{d}t +\right.$$

$$\left.\int_0^{\infty}(\sqrt2\sigma t-\delta(x))^r\exp(-t^2)\mathrm{d}t\right]p(x)\mathrm{d}x.$$

再令 $t = u + \frac{\delta(x)}{\sqrt2\sigma}$ 并进一步简化, 得

$$E_{XY}(|y-\hat w^{\mathrm T} x|^r)\approx$$

$$\int_\Omega\frac{1}{\sqrt\pi}\left[\int_0^\infty(\sqrt2\sigma t)^r\exp\left(-\left(t-\frac{\delta(x)}{\sqrt2\sigma}\right)^2\right)\mathrm{d}t +\right.$$

$$\left.\int_0^\infty(\sqrt2\sigma t)^r\exp\left(-\left(t+\frac{\delta(x)}{\sqrt2\sigma}\right)^2\right)\mathrm{d}t\right]p(x)\mathrm{d}x.$$

分别将 $\exp\left(-\left(t-\frac{\delta(x)}{\sqrt2\sigma}\right)^2\right)$ 和 $\exp\left(-\left(t+\frac{\delta(x)}{\sqrt2\sigma}\right)^2\right)$ 在 t 处展开至 $\frac{\delta(x)}{\sqrt2\sigma}$ 的 6 次方项, 得

$$E_{XY}(|y-\hat w^{\mathrm T} x|^r)\approx$$

$$\int_\Omega\frac{2}{\sqrt\pi}p(x)\left[\int_0^\infty(\sqrt2\sigma t)^r\left(\exp(-t^2)\left(1-\right.\right.\right.$$

$$\left.\frac{\delta^2(x)}{2\sigma^2}+\frac{\delta^4(x)}{8\sigma^2}-\frac{\delta^6(x)}{48\sigma^2}\right)+$$

$$t^2\exp(-t^2)\left(\frac{\delta^2(x)}{\sigma^2}-\frac{\delta^4(x)}{2\sigma^2}+\frac{\delta^6(x)}{8\sigma^2}\right)+$$

$$t^4\exp(-t^2)\left(\frac{\delta^4(x)}{6\sigma^2}-\frac{\delta^6(x)}{12\sigma^2}\right)+$$

$$\left.\left.\frac{\delta^6(x)}{90\sigma^2}t^6\exp(-t^2)\right)\mathrm{d}t\right]\mathrm{d}x.$$

注意到

$$\int_\Omega p(x)\mathrm{d}x = 1,$$

$$\int_0^\infty t^r\exp(-t^2)\mathrm{d}t = \frac12\Gamma\left(\frac{r+1}{2}\right)$$

$$\int_\Omega\delta(x)^k p(x)\mathrm{d}x \approx \sigma^k(k-1)!!,$$

则有

$$E_{XY}(|y-\hat w^{\mathrm T} x|^r)\approx$$

$$\frac{1}{\sqrt\pi}(\sqrt2\sigma)^r\left[\frac{7}{16}\Gamma\left(\frac{r+1}{2}\right)+\frac78\Gamma\left(\frac{r+3}{2}\right)-\right.$$

$$\left.\frac34\Gamma\left(\frac{r+5}{2}\right)+\frac16\Gamma\left(\frac{r+7}{2}\right)\right]. \tag{9}$$

而

$$\partial E_{XY}(|y-\hat w^{\mathrm T} x|^r)/\partial r\approx$$

$$\frac{1}{\sqrt\pi}(\sqrt2\sigma)^r\ln(\sqrt2\sigma)^r\left[\frac{7}{16}\Gamma\left(\frac{r+1}{2}\right)+\right.$$

$$\left.\frac78\Gamma\left(\frac{r+3}{2}\right)-\frac34\Gamma\left(\frac{r+5}{2}\right)+\frac16\Gamma\left(\frac{r+7}{2}\right)\right]+$$

$$\frac{1}{2\sqrt\pi}(\sqrt2\sigma)^r\ln(\sqrt2\sigma)^r\left[\frac{7}{16}\Gamma\left(\frac{r+1}{2}\right)\times\right.$$

$$\Psi\left(\frac{r+1}{2}\right)+\frac{7}{8}\Gamma\left(\frac{r+3}{2}\right)\Psi\left(\frac{r+3}{2}\right)-\frac{3}{4}\Gamma\left(\frac{r+5}{2}\right)\Psi\left(\frac{r+5}{2}\right)+\frac{1}{6}\Gamma\left(\frac{r+7}{2}\right)\Psi\left(\frac{r+7}{2}\right)\Big], \tag{10}$$

其中

$$\frac{\partial\Gamma(x)}{\partial r}=\Gamma(x)\Psi(x),$$

$$\Psi(x)=\sum_{m=0}^{\infty}\left(\frac{1}{m+1}-\frac{1}{m+x}\right)-\gamma,$$

γ是欧拉常数[5].

由式(7) 和(8) 消去β,得

$$\frac{\partial E_{XY}(|y-\hat{w}^{\mathrm{T}}x|^{r})}{\partial r}-\frac{1}{r}E_{XY}(|y-\hat{w}^{\mathrm{T}}x|^{r})\ln(rE_{XY}(|y-\hat{w}^{\mathrm{T}}x|^{r}))=0. \tag{11}$$

将式(9) 和(10) 代入上式,并令其左边为$f(r,\sigma)$,得

$$f(r,\sigma)=0. \tag{12}$$

显然,从式(12) 不容易直接看出r与σ的关系,因此,图1给出了式(12) 的数值计算结果. 数值计算中,取

$$\Psi(x)\approx\sum_{m=0}^{1\,000}\left(\frac{1}{m+1}-\frac{1}{m+x}\right)-0.577\,215\,67,$$

对σ在0.5附近的r值作了平滑处理.

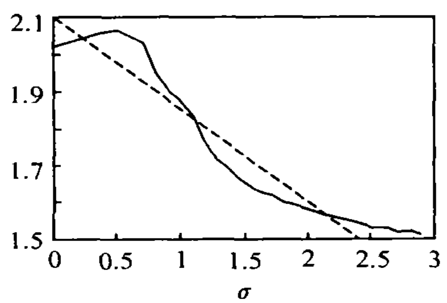

图1 对式(12) 作数值计算处理后的曲线

图1的曲线表明: σ与r基本呈近似线性反向变动的关系(如图中的虚线所示). 另外从图1可以看出,当噪声较小时($\sigma<1$),建模的失真并不太大,故r范数取2附近的值为宜,这也验证了人们在无噪音时通常取范数为2的损失函数这一事实. 随着噪声的加大,r范数几乎呈线性下降,因为噪声太大会导致建模失真,故σ分析到2.5～3已足够. 图中的虚线是为了有助于我们看出线性反向变动关系.

考虑到推导中用了一些近似,因此可以推断: 1) 当无噪音或噪音很小时,r应取2或其附近的值; 2) 当噪音较大时,σ与r范数的取值呈近似线性反向变动关系. 也就是说,在有噪音时,r范数的取值应小于2,并与σ呈近似线性反向变动关系. 顺便指出,当前馈神经网络用于噪音数据时,r通常应取小于2[6]. 在一定条件下,支持向量回归机与前馈神经网络具有一定的等价性,这也充分验证了本文结论的合理性.

5 结 语

本文通过将r范数-支持向量回归机r-SVR的优化问题转换成最大后验估计问题,研究了r范数-支持向量回归机中参数r的选择问题. 结果显示,在输入噪声为高斯模型时,r与σ存在某种近似的线性反向变动关系. 这一理论结果的意义在于: 可以指导在已知输入噪声分布的情况下合理选择r,从而使r范数-支持向量回归机更具鲁棒性. 此外,本文推导出的r范数-支持向量回归机后验估计最大化的必要条件也可应用于其他输入噪声模型,从而相应地导出如何合理地选取r. 为将上述结果应用于实际问题,必须估计噪声的参数σ. 文献[7]提出了Bayes估计方法,作者也将在这方面作进一步的研究.

参考文献(References):

[1] Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines*[M]. Cambridge: Cambridge University Press, 2000.

[2] Vapnik V. *Statistical Learning Theory* [M]. New York: Wiley, 1998.

[3] James T K, Ivor W T. Linear dependency between ε and the input noise in ε-support vector regression[J]. *IEEE Trans on Neural Networks*, 2003, 14(5): 544-553.

[4] Smola A J, Murata N, Schölkopf B, et al. Asymptotically optimal choice of ε-loss for support vector machines[A]. *Proc of the Int Conf on Artificial Neural Networks*[C]. 1998.

[5] 沈永欢,梁在中,许履瑚,等. 实用数学手册[M]. 北京: 科学出版社, 2002. 657.

[6] 阎平凡,张长水. 人工神经网络和模拟进化计算[M]. 北京: 清华大学出版社, 2001.

[7] Vladimir Cherkassky, Yunqian Ma. Practical selection of SVM parameters and noise estimation for SVM regression[J]. *Neural Networks*, 2003, 17(1): 113-126.

[8] Smola A J, Schölkopf B. A tutotial on support vector regression[R]. London: University of London, 1998.

[9] Law M H, Kwok J T. Bayesian support vector regression[A]. *Proc of the English Int Workshop on Artificial Intelligence and Statistics*[C]. Florida: Key West, 2001. 239-244.

[10] Gao J B, Gunn S R, Harris C J. A probabilistic framework for SVM regression and error bar estimation[J]. *Machine Learning*, 2002, 46(2): 71-89.