Huber-SVR 中参数 μ 与输入噪声间的近似线性关系*)

周晓剑 朱嘉钢 王士同

(江南大学信息工程学院 无锡 214122)

摘 要 为使 Huber-SVR 更具鲁棒性,深入研究了 Huber-SVR 中参数与输入噪声之间的关系。运用 SVR 的贝叶斯框架,分别推导出了鲁棒的 Huber-SVR 中参数 μ 与拉斯噪声和均匀噪声之间呈近似线性关系,并结合仿真结果和已有的相关结论,得到了更为一般的结论,即鲁棒的 Huber-SVR 中参数 μ 与输入噪声之间呈近似线性关系。这一结论为输入样本含有分布未知噪声的情况下 Huber-SVR 参数的选择提供了理论依据。

关键词 支持向量机,支持向量回归机,Huber 损失函数

Approximately Linear Dependency between μ and the Input Noise in Huber-Support Vector Regression

ZHOU Xiao-Jian ZHU Jia-Gang WANG Shi-Tong

(School of Information Engineering, Southern Yangtze University, Wuxi 214122)

Abstract The dependency relationship between μ and the input noise in Huber-SVR is studied using SVR Bayesian evidence framework. First, focus is paid on the cases of Laplacian noise and Uniform noise, and the approximately linear dependencies between μ and the variances of the two noises are then respectively derived. Second, with the relevant conclusion on Huber-SVR and experimental study, the more general claim is then proposed that the approximately linear dependency is almost kept between μ and the input noise in Huber-SVR. Such a dependency relationship is useful to determine the optimal choice for μ in Huber-loss function in the existence of unknown input noise.

Keywords Support vector machines(SVM), Support vector regression(SVR), Huber loss function

1 引言

近年来,SVM 广泛地应用于各种分类和回归问题中。当 SVM 应用于回归分析和预测时,通常称之为 SVR。由于真实的数据经常含有噪声,因此在用 SVR 进行回归时,如何使 SVR 具有更好的鲁棒性,即如何选择损失函数的参数使得 SVR 具有尽可能强的抑制噪声能力,就成了一个重要的研究课题。为此一些学者进行了研究,重要成果有:文[1,2]提出贝叶斯框架理论,该理论指出 SVR 优化问题等价于求对应的最大后验概率问题(MAP 问题)。文[3]在贝叶斯框架下推导出重要结论,即 ε-SVR 中参数 ε 和输入噪声(高斯噪声、拉斯噪声、均匀噪声)间呈线性关系。文[4,5]在贝叶斯框架下推导出了 Huber-SVR 中的参数 μ 与高斯输入噪声间呈近似线性关系。实际中,除了高斯噪声外,拉斯噪声和均匀噪声也是典型的噪声。在文[4,5]的基础上,研究 Huber-SVR 中的参数 μ 与拉斯噪声和均匀噪声间的关系,可以进一步得出 Huber-SVR 中的参数 μ 与输入噪声间的更一般的关系,从而为输入噪声分布未知情形下 Huber-SVR 的参数选择提供理论依据,本文的目的就在于此。本文运用 SVR 的贝叶斯框架,推导出 Huber-SVR 中的参数 μ 与拉斯噪声和均匀噪声之间均呈近似线性关系,这一理论推导结果也与实验结果相吻合。进而,在文[4,5]的基础上,综合理论和实验两方面结果,得出 Huber-SVR 的参数与输入噪声间总是呈近似线性关系这一更一般的结论。

*)江苏省自然科学基金(BK2003017);教育部跨世纪优秀人才计划资助项目;江苏计算机技术重点实验室开放课题资助项目。**周晓剑** 硕士研究生,从事软件理论、人工神经网络等研究;**朱嘉钢** 博士,副教授,从事人工神经网络、模式识别等研究;**王士同** 教授,博士生导师,CCF 高级会员,从事人工神经网络、模式识别、生物信息、模糊系统等研究。

种新的解释,而且以上定理表明通过知识距离定义的粗糙熵与其它定义下的粗糙熵也同样具有粗糙不变性和随知识分辨能力增强而单调减少的性质。

结束语 基于知识即划分的观点,本文证明了论域中所有知识构成一个距离空间,研究了知识空间中知识距离的一些性质,并从一个全新的角度运用知识距离定义了信息系统中的知识粗糙熵,为信息系统中知识度量研究提供了新工具,有利于更加深刻地分析信息系统中知识间的关系和知识的粗糙性。

参考文献

1 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publisher, 1991
2 梁吉业,李德玉. 信息系统中的不确定性与知识获取. 北京:科学出版社,2005
3 管延勇,王洪凯,史开泉. 知识的粗识别及其评判. 计算机科学,2004,31(12):113~116
4 张钹,张铃. 问题求解理论及应用. 北京:清华大学出版社,1990
5 张钹,张铃. 模糊商空间理论(模糊粒度计算方法). 软件学报,2003,14(4):770~776
6 Beaubouef T, Petry F E, Arora G. Information theoretic measures of uncertainty for rough sets and rough relational databases. Information Science, 1998, 109(1-4): 185~195
7 Wierman M J. Measuring uncertainty in rough set theory. International Journal of General Systems, 1999, 28(4): 283~397
8 Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2004, 12(1): 37~46
9 黄兵,何新,周献中. 基于广义粗集覆盖约简的粗糙熵. 软件学报,2004,15(2):215~220
10 黄兵,周献中,史迎春. 基于一般二元关系的知识粗糙熵与粗集粗糙熵. 系统工程理论与实践,2004(1):93~96

本文组织如下：第 2 节介绍有关已有的研究结果，包括 Huber-支持向量回归机、贝叶斯框架和 Huber-支持向量回归机的参数与高斯输入噪声间的近似线性关系；第 3 节运用贝叶斯框架推导 Huber -SVR 的后验估计最大化条件；第 4 节运用这一后验估计最大化条件推导 Huber -SVR 中损失函数的参数 μ 与输入噪声之间的近似线性关系；第 5 节给出实验和实验结果。最后给出结论。

2 Huber-SVR 的参数与高斯输入噪声间的近似线性关系[1,2,4,5]

2.1 Huber-SVR

当 SVR 中的损失函数取 Huber 函数时，构成 Huber-SVR。其模型如下：

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

$$s.t.\begin{cases}L_{Huber}(y_i-f(x_i))\leqslant\xi_i^-,y_i\geqslant f(x_i)\\ L_{Hhber}(f(x_i)-y_i)\leqslant\xi_i^+,f(x_i)>y_i,i=1,\cdots,l\\ \xi_i^-,\xi_i^+\geqslant 0\end{cases}$$

将 Huber 函数

$$L_{Haber}(f(x)-y)=\begin{cases}(f(x)-y)^2/4\mu,|f(x)-y|\leqslant 2\mu\\ |f(x)-y|-\mu,其他\end{cases} \tag{2}$$

代入上述约束条件，则上述模型可以进一步化为如下的二次规划问题：

$$\min_{\beta}\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_jK(x_i,y_j)-\sum_{i=1}^{l}\alpha_iy_i+\frac{1}{2C}\sum_{i=1}^{l}\alpha_i^2\mu$$

$$\begin{cases}-C\leqslant\alpha_i\leqslant C,i=1,\cdots,l\\ \sum_{i=1}^{l}\alpha_i=0\end{cases}$$

其中 $K(.,.)$是核函数。在求出上式中的参数 α 后，即可求得回归函数 $f(x)$。

2.2 Huber-SVR 的贝叶斯框架[1,2]

在参数估计中，贝叶斯方法是一种把待估计参数看作具有某一已知分布的随机变量，从而进行参数估计的、将先验信息数学形式化的方法。通常称这一已知分布为先验分布。由于利用了先验信息，则当先验信息与实际情形比较一致时，贝叶斯估计要优于极大似然估计。当将这一思想应用于 SVR 时，就有了 SVR 的贝叶斯框架，由此可以推导出 SVR 损失函数的参数与输入样本噪声之间的关系。下面对此作简要介绍。

对于给定的样本集

$$D=\{(x_1,y_1),\cdots,(x_l,y_l)\},x\in R^n,y\in R$$

考虑用

$$y_i=\tilde{w}^Tx_i+\eta_i,i=1,\cdots,l$$

进行回归时的权参数 $\tilde{w}$ 估计问题。其中，$x_i\in\Omega$ 服从分布 $p(\cdot)$，η_i 服从分布 $\phi(\cdot)$。对应于 y 的密度函数记为 $p(y|x)=\phi(y-\tilde{w}^Tx)$。为简化问题，假定所有的 x_i 具有 0 均值。

求解上述问题的 Huber-SVR 贝叶斯框架要点如下：

(1)Huber 损失函数导致以下的对应于 y 的概率密度函数：

$$p(y_i|x_i,w,\beta,\mu)=C(\beta,\mu)\exp(-\beta L_{humber}(f(x)-y)) \tag{3}$$

其中，μ 是 Huber 损失函数的参数，β 是用于控制噪声水平的参数，$C(\beta,\mu)$是归一化系数。则

$$p(D|w,\beta,\mu)=(C(\beta,\mu))^l\exp(-\sum_{i=1}^{l}\beta L_{humber}(f(x_i)-y_i)) \tag{4}$$

(2)权参数 w 的先验分布为

$$p(w)=\sqrt{\frac{\alpha}{2\pi}}\exp\left(-\frac{\alpha}{2}\|w\|^2\right) \tag{5}$$

其中，α 为某一常数。

(3)权参数 w 的后验分布为

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

则

$$\log p(w|D,w,\beta,\mu)=-\frac{\alpha}{2}\|w\|^2$$

$$-\beta\sum_{i=1}^{l}L_{humber}(y_i-w^Tx_i)+l\log C(\beta,\mu)+常数 \tag{6}$$

于是，估计权参数 $\tilde{w}$ 的问题就变成了对于给定的 β、μ 的值求 w 的最大后验估计问题。由此可以进一步推导出损失函数的参数与输入噪声之间的关系。

2.3 Huber-SVR 的参数与高斯输入噪声间的近似线性关系 [4,5]

文[4,5]通过复杂的理论推导得到了如下关系式：

$$f(\delta,\mu)\geqslant 2\sqrt{\left[\frac{3\mu\sigma}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)+\frac{\mu^3}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)+\frac{\mu^3}{6\sqrt{2\pi}\sigma}-\frac{\mu^2}{\sqrt{\pi}}\int_{\frac{\mu}{\sqrt{2}\sigma}}^{\infty}\exp(-t^2)\mathrm{d}t\right]\frac{2}{\mu^2}}+\log\mu \tag{7}$$

明显地，上式与$\left(\frac{\mu}{\sigma}\right)$相关。当$\left(\frac{\mu}{\sigma}\right)$取一定值时，$f(\delta,\mu)$将达到最小。这一结果表明，$\mu$ 与标准差 δ 存在着近似线性关系。

3 Huber-SVR 后验估计最大化必要条件

下面推导：当 Huber 函数为式(2)时，Huber-支持向量回归机后验估计最大化的必要条件。

为便于分析，用期望

$$E_{XY}(L_{humber}(y-w^Tx))=\int_{\Omega}\int_{-\infty}^{\infty}L_{humber}(y-w^Tx)p(y|x)p(x)\mathrm{d}y\mathrm{d}x$$

$$=\int_{\Omega}\left(\int_{-\infty}^{w^Tx-2\mu}(w^Tx-y-\mu)p(y|x)\mathrm{d}y+\int_{w^Tx-2\mu}^{w^Tx+2\mu}\frac{(y-w^Tx)^2}{4\mu}p(y|x)\mathrm{d}y\right.$$

$$\left.+\int_{w^Tx+2\mu}^{\infty}(y-w^Tx-\mu)p(y|x)\mathrm{d}y\right)p(x)\mathrm{d}x$$

近似代替

$$\frac{1}{l}\sum_{i=1}^{l}L_{humber}(y_i-w^Tx_i)$$

则式(6)变为

$$M(w,\beta,\mu)=-\frac{\alpha}{2}\|w\|^2-\beta lE_{XY}(L_{humber}(y-w^Tx))+l\log C(\beta,m)+常数 \tag{8}$$

所谓后验估计最大化必要条件，就是将分别对式(8)的 3 个参数求导数，并令其等于 0 而得到的 3 个关系式。下面就推导这 3 个关系式。

将式(8)对 w 求偏导数并令其等于 0，得

$$\alpha\hat{w}+\beta l\int_{\Omega}\left(\int_{-\infty}^{\hat{w}^Tx-2\mu}xp(y|x)\mathrm{d}y+\int_{\hat{w}^Tx-2\mu}^{\hat{w}^Tx+2\mu}\frac{x(y-\hat{w}^Tx)}{2\mu}p(y|x)\mathrm{d}y-\int_{\hat{w}^Txk+2\mu}^{\infty}xp(y|x)\mathrm{d}y\right)p(x)\mathrm{d}x=0 \tag{9}$$

使 Huber-支持向量回归机后验估计最大，就是使

$$M(\hat{w},\beta,\mu)=-\frac{\alpha}{2}\|\hat{w}\|^2-\beta lE_{XY}(L_{humber}(y-\hat{w}^Tx))+l\log C(\beta,\mu)+常数 \tag{10}$$

最大。其中 $C(\beta,\mu)$是归一化系数

$$C(\beta,\mu)\approx\frac{3\beta}{6\beta\mu-4\beta^2\mu^2+6} \tag{11}$$

为求得 $M(\hat{w},\beta,\mu)$ 的最大值，在式(8)中，使 $w=\hat{w}$，对 β 求偏导数并令其等于0，得

$$\frac{\partial M(\hat{w},\beta,\mu)}{\partial\beta}=(\alpha\hat{w}+\beta l\int_\Omega(\int_{-\infty}^{\hat{w}^Tx-2\mu}xp(y|x)\mathrm{d}y+\int_{\hat{w}^Tx-2\mu}^{\hat{w}^Tx+2\mu}\frac{x(y-\hat{w}^Tx)}{2\mu}p(y|x)\mathrm{d}y-\int_{\hat{w}^Tx+2\mu}^{\infty}xp(y|x)\mathrm{d}y)p(x)\mathrm{d}x)\frac{\partial\hat{w}}{\partial\beta}+\frac{\partial M}{\partial\beta}=0$$

即得，

$$E_{XY}(L_{humber}(y-\hat{w}^Tx))=\frac{2\beta^2\mu^2+3}{\beta(3\beta\mu-2\beta^2\mu^2+3)} \tag{12}$$

同理，在式(8)中，使 $w=\hat{w}$ 并对 μ 求偏导数并令其等于0，得

$$\frac{\partial M(\hat{w},\beta,\mu)}{\partial\epsilon}=(\alpha\hat{w}+\beta l\int_\Omega(\int_{-\infty}^{\hat{w}^Tx-2\mu}xp(y|x)\mathrm{d}y+\int_{\hat{w}^Tx-2\mu}^{\hat{w}^Tx+2\mu}\frac{x(y-\hat{w}^Tx)}{2\mu}p(y|x)\mathrm{d}y-\int_{\hat{w}^Tx+2\mu}^{\infty}xp(y|x)\mathrm{d}y)p(x)\mathrm{d}x)^T\frac{\partial\hat{w}}{\partial\epsilon}+\frac{\partial M}{\partial\epsilon}=0$$

即得

$$\int_\Omega\left(\int_{-\infty}^{W^Tx-2\mu}p(y|x)\mathrm{d}y+\frac{1}{4\mu^2}\int_{W^TX-2\mu}^{W^TX+2\mu}(y-\hat{w}^Tx)^2p(y|x)\mathrm{d}y+\int_{W^TX+2\mu}^{\infty}p(y|x)\right)p(x)\mathrm{d}x=-\frac{3-4\beta\mu}{3\beta\mu-2\beta^2\mu^2+3} \tag{13}$$

式(9)、(12)和(13)构成了 Huber-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 $\phi(y-\tilde{w}^Tx)$ 噪声，只需令 $p(y|x)=\phi(y-\tilde{w}^Tx)$ 后代入式(12)和(13)，即可求得 β、μ 与输入噪声间的关系。

在实际中，拉斯噪声和均匀噪声也是典型的噪声。以下研究 Huber-支持向量回归机的参数 μ 与拉斯和均匀噪声的关系。

4 Huber-SVR 的参数与输入噪声间近似线性关系的理论推导结果

4.1 Huber-SVR 的参数与拉斯输入噪声间的近似线性关系

当输入噪声取以下拉斯分布时，

$$\phi(\eta)=\frac{1}{2\xi}\exp\left(-\frac{|\eta|}{\xi}\right)$$

只需分别在式(12)和(13)令

$$p(y|x)=\phi(y-\tilde{w}^Tx),$$

即

$$p(y|x)=\frac{1}{2\delta}\exp\left(-\frac{|y-\tilde{w}^Tx|}{\delta}\right)$$

解出参数 β，并利用式(10)，可以求得我们所要的 SVR 的参数 μ 和输入噪声标准差 σ 的关系。

为此，由式(12)得

$$E_{XY}(L_{humber}(y-\hat{w}^Tx))=\iint_{\Omega-\infty}^{\infty}L_{humber}(y-\hat{w}^Tx)p(y|x)p(x)\mathrm{d}y\mathrm{d}x$$

$$=\int_\Omega\left(\int_{-\infty}^{\hat{w}^Tx-2\mu}(\hat{w}^Tx-y-\mu)\frac{1}{2\xi}\exp\left(-\frac{|y-\hat{w}^Tx|}{\xi}\right)\mathrm{d}y+\int_{\hat{w}^Tx+2\mu}^{\hat{w}^Tx+2\mu}\frac{1}{4\mu}(y-\hat{w}^Tx)^2\frac{1}{2\xi}\exp\left(-\frac{|y-\hat{w}^Tx|}{\xi}\right)\mathrm{d}y+\int_{\hat{w}^Tx+2\mu}^{\infty}(y-\hat{w}^Tx-\mu)\frac{1}{2\xi}\exp\left(-\frac{|y-\hat{w}^Tx|}{\xi}\right)\mathrm{d}y\right)p(x)\mathrm{d}x \tag{14}$$

为便于推导，在此仅讨论一维的情形。对 X 是多维的情形，只要选择合适的每一维的积分子区间，也不难得到类似的结论。

对一维的情形，当 X 服从 $[-L,L]$ 上的均匀分布(即 $X\sim U[-L,L]$)时，注意到 $L(\tilde{w}-\hat{w})=\sqrt{6}\xi$[3]，经推导，式(14)可化为

$$\frac{2\beta^2\mu^2+3}{\beta(3\beta\mu-2\beta^2\mu^2+3)}=\frac{(\mu+\xi)(\xi-2\mu)}{\xi}+\frac{1}{4\xi\mu}[4\xi^3-(4\mu^2+4\xi\mu+2\xi^2)(\xi-2\mu)] \tag{15}$$

同理，由式(13)经推导得

$$\frac{3-4\beta\mu}{3\beta\mu-2\beta^2\mu^2+3}=\frac{\xi-2\mu}{\xi}+\frac{[4\xi^3-(4\mu^2+4\xi\mu+2\xi^2)(\xi-2\mu)]}{4\mu^2\xi} \tag{16}$$

为消去 β，对式(15)、(16)右边分别令

$$\xi f_1\left(\frac{\mu}{\xi}\right)=\frac{(\mu+\xi)(\xi-2\mu)}{\xi}+\frac{[4\xi^3-(4\mu^2+4\xi\mu+2\xi^2)(\xi-2\mu)]}{4\mu\xi}$$

$$f_2\left(\frac{\mu}{\xi}\right)=\frac{\xi-2\mu}{\xi}+\frac{[4\xi^3-(4\mu^2+4\xi\mu+2\xi^2)(\xi-2\mu)]}{4\mu^2\xi}$$

由式(15)、(16)解出 β，得

$$\beta=\frac{3\left(\xi f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)}{2\left(2\mu^2+4\mu\left(\xi f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)\right)}\pm\frac{\sqrt{9\left(\xi f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)^2-12\left(3\mu^2+4\mu\left(\xi f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)\right)}}{2\left(2\mu^2+4\mu\left(\xi f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)\right)}$$

为表示得更简洁，令

$$f_3\left(\frac{\mu}{\xi}\right)=-\frac{3\left(f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)}{4\left(\frac{\mu}{\xi}\right)^2+8\left(\frac{\mu}{\xi}\right)\left(f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)}\pm\frac{\sqrt{9\left(f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)^2-12\left(3\left(\frac{\mu}{\xi}\right)^2+4\left(\frac{\mu}{\xi}\right)\left(f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)\right)}}{4\left(\frac{\mu}{\xi}\right)^2+8\left(\frac{\mu}{\xi}\right)\left(f_1\left(\frac{\mu}{\xi}\right)\Big/f_2\left(\frac{\mu}{\xi}\right)\right)}$$

则 β 可表示为

$$\beta=\frac{1}{\xi}f_3\left(\frac{\mu}{\xi}\right) \tag{17}$$

以下通过式(10)求出参数 μ 和输入噪声标准差的关系。

由式(10)，求式(10)中的 M 的最大值，也就是求 $f(\beta,\mu)$

$$=\beta E_{XY}(L_{humber}(y-\hat{w}^Tx))-\log C(\beta,\mu) \tag{18}$$

的最小值。

将式(17)代入式(18)，综合式(14)、(15)，并令

$$\varphi\left(\frac{\mu}{\xi}\right)=\frac{3f_3\left(\frac{\mu}{\xi}\right)}{6\frac{\mu}{\xi}f_3\left(\frac{\mu}{\xi}\right)-4\left(\frac{\mu}{\xi}\right)^2f_3^2\left(\frac{\mu}{\xi}\right)+6}$$

可得

$$f(\xi,\mu)=\frac{1}{\xi}f_3\left(\frac{\mu}{\xi}\right)\cdot\xi f_1\left(\frac{\mu}{\xi}\right)-\log\varphi\left(\frac{\mu}{\xi}\right)-\log\xi \tag{19}$$

明显地，上式与$\left(\frac{\mu}{\xi}\right)$相关。当$\left(\frac{\mu}{\xi}\right)$取一定值时，$f(\xi,\mu)$将达到最小。这一结果表明，在拉斯噪声中，$\mu$与$\xi$存在着近似线性关系。又因为拉斯噪声的标准差$\delta=\sqrt{2}\xi$，所以$\mu$与$\delta$也是近似线性的关系。

4.2 Huber-SVR的参数与均匀输入噪声间的近似线性关系

同理，当输入噪声为均匀分布时，

$$\varphi(\eta)=\frac{1}{2\xi},\ \eta\in[-\xi,\xi]$$

类似式(14)，有

$$E_{XY}(L_{hamber}(y-\hat{w}^Tx))=$$
$$\iint_{\Omega\,-\infty}^{\infty} L_{humber}(y-\hat{w}^Tx)p(y|x)p(x)\mathrm{d}y\mathrm{d}x$$
$$=\int_\Omega\left(\int_{-\infty}^{\hat{w}^Tx-2\mu}(\hat{w}^Tx-y-\mu)\frac{1}{2\xi}\mathrm{d}y+\int_{\hat{w}^Tx-2\mu}^{\hat{w}^Tx+2\mu}\frac{1}{4\mu}(y-\hat{w}^Tx)^2\ \frac{1}{2\xi}\mathrm{d}y\right.$$
$$\left.+\int_{\hat{w}^Tx+2\mu}^{\infty}(y-\hat{w}^Tx-\mu)\frac{1}{2\xi}\mathrm{d}y\right)p(x)\mathrm{d}x$$
$$=\frac{\mu^3(\xi-2\mu)}{4\xi^2}+\frac{\mu^2}{4\xi}-\frac{\mu^3-(2\xi-\mu)^3}{12\xi^2}+\frac{\mu^2}{3\xi}-\frac{\varepsilon(2\mu-\xi)}{3\xi^2}\quad(20)$$

同样，对式(13)的左边，有

$$\int_\Omega\left(\int_{-\infty}^{w^TX-2\mu}\frac{1}{2\xi}\mathrm{d}y+\frac{1}{4\mu^2}\int_{w^TX-2\mu}^{w^TX+2\mu}(y-\hat{w}^Tx)^2\ \frac{1}{2\xi}\mathrm{d}y\right.$$
$$\left.+\int_{w^TX+2\mu}^{\infty}\frac{1}{2\xi}\right)p\ (x)\mathrm{d}x$$
$$=\frac{(\xi-\mu)^2}{\xi^2}+\frac{\mu}{3\xi}-\frac{\varepsilon\ (2\mu-\xi)}{3\xi^2}\quad(21)$$

所以综合式(12)、(20)和式(13)、(21)可得

$$\frac{2\beta^2\mu^2+3}{\beta(3\beta\mu-2\beta^2\mu^2+3)}=\frac{\mu^3(\xi-2\mu)}{4\xi^2}+\frac{\mu^2}{4\xi}$$
$$-\frac{\mu^3-(2\xi-\mu)^3}{12\xi^2}+\frac{\mu^2}{3\xi}-\frac{\mu(2\mu-\xi)}{3\xi^2}\quad(22)$$
$$\frac{3-4\beta\mu}{3\beta\mu-2\beta^2\mu^2+3}=\frac{(\xi-\mu)^2}{\xi^2}+\frac{\mu}{3\xi}-\frac{\mu(2\mu-\xi)}{3\xi^2}\quad(23)$$

类似式(19)，可得到

$$F(\xi,\mu)=\frac{1}{\xi}F_3\left(\frac{\mu}{\xi}\right)\cdot\xi F_1\left(\frac{\mu}{\xi}\right)-\log\varphi\left(\frac{\mu}{\xi}\right)-\log\xi$$

显然，上式与$\left(\frac{\mu}{\xi}\right)$相关。当$\left(\frac{\mu}{\xi}\right)$取一定值时，$F(\xi,\mu)$将达到最小。这一结果表明，在均匀噪声中，$\mu$与$\xi$同样也存在着近似线性关系。又因均匀分布的标准差$\delta=\xi/\sqrt{3}$，所以$\mu$与$\delta$也是近似线性关系。

4.3 Huber-SVR的参数与输入噪声间的近似线性关系

由于Huber-支持向量回归机中参数μ分别与高斯输入噪声[4,5]、拉斯输入噪声和均匀输入噪声之间均呈近似线性关系，依据文[3]的同样逻辑，可以得出更一般的结论，即Huber-支持向量回归机中参数μ与输入噪声间呈近似线性关系。

5 实验和实验结果

5.1 实验目的

本实验的目的有二：第一，验证Huber-SVR中参数μ与拉斯输入噪声间的近似线性关系；第二，验证Huber-SVR中参数μ与均匀输入噪声间的近似线性关系。

5.2 实验步骤

分别对拉斯噪声、均匀噪声做二组实验，每组都按以下步骤进行。

取函数$y=\frac{\sin(x)}{x}$，$x\in[-10,10]$，为基准函数。令x从-10至10，步长为0.5产生其样本(x_i,y_i)，$i=1,\cdots,41$。用Huber-SVR对这些样本做回归，得到对应于$y=\frac{\sin(x)}{x}$的回归曲线$r=r(x)$。对应地，取函数$y'=\frac{\sin(x)}{x}+k\cdot\eta$，$x\in[-10,10]$为加入噪声后的信号，其中为信噪比系数。令$x$从$-10$至10，步长为0.5产生其样本$(x_i,y'_i)$，$i=1,\cdots,41$。用Huber-SVR对这些样本做回归，得到对应于$y'=\frac{\sin(x)}{x}+k\cdot\eta$的回归曲线$r'=r(x)$。在进行Huber-SVR回归时，令$\delta$在[0.1,2.0]上变化。对于一个给定的$\delta$，取使$\sum_{i=1}^{41}\mathrm{Huber}(r'_i-r_i)$最小的$\mu$，观察$\mu$与$\delta$之间的关系。由于$\eta$是随机变量，实际计算时，对于每一个给定的$\delta$，产生20组$\eta$的值，每一组都计算出一个$\mu$，取该20个$\mu$的平均值作为最后的$\mu$，以抑制$\mu$的随机性。

5.3 实验结果与分析

在图2和图4中，符号"+"是μ。由于μ有随机性，对其进行回归并用实线表示，实线反映了μ随着δ的变化而变化的趋势。

图1、图2给出了第一组实验结果。图1是对加有拉斯噪声的基准函数用Huber-SVR回归的结果，图2给出拉斯输入噪声下使Huber-SVR鲁棒性最佳时μ与δ的关系。该组实验结果表明：当δ在0到1.2时，Huber-SVR中参数μ与拉斯输入噪声间呈近似线性关系。

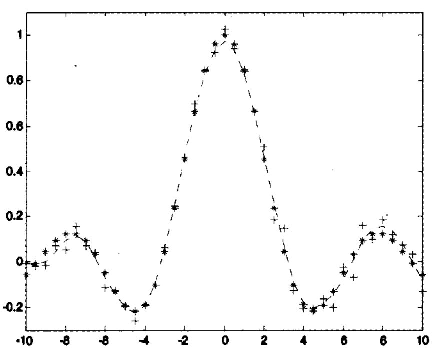

图1 对加有拉斯噪声的基准函数用Huber-SVR回归的结果($k=0.05$)

"* * *"：$y=\frac{\sin(x)}{x}$，"—"：$r'=r'(x)$，

"+++"：$y'=\frac{\sin(x)}{x}+k\cdot\eta$

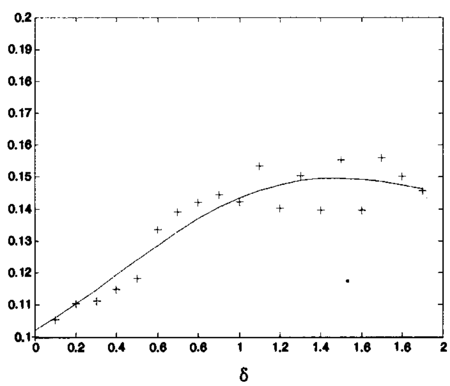

图2 拉斯输入噪声下μ与δ间的近似线性关系

图 3、图 4 给出了第二组实验结果。图 3 是对加有均匀噪声的基准函数用 Huber-SVR 回归的结果，图 4 给出均匀输入噪声下使 Huber-SVR 鲁棒性最佳时的 μ 与 δ 的关系。该组实验结果表明：当 δ 在 0 到 1.4 时，Huber-SVR 中参数 μ 与拉斯输入噪声间呈近似线性关系。

结束语 本文运用 SVR 的贝叶斯框架，推导 Huber-SVR 中的参数 μ 与拉斯噪声和均匀噪声之间均呈近似线性关系，这一理论推导结果也与实验结果相吻合。由于高斯噪声、拉斯噪声和均匀噪声的典型性，在文[4,5]的基础上，综合理论和实验两方面结果，可以进一步得 Huber-SVR 的参数与输入噪声间呈近似线性关系这个更为一般的结论。这一结论可以为鲁棒的 Huber-SVR 的参数选择提供更完善的理论依据。

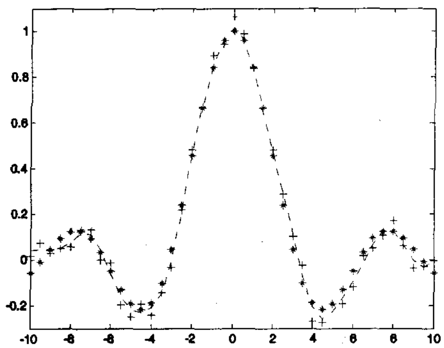

图 3 对加有均匀噪声的基准函数用 Huber-SVR 回归的结果(k=0.05)

$$“***”: y=\frac{\sin(x)}{x}, “—”: r'=r(x),$$

$$“+++”: y'=\frac{\sin(x)}{x}+k\cdot\eta_2$$

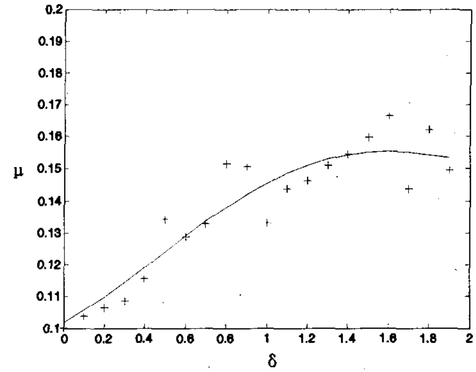

图 4 均匀噪声下 μ 与 δ 间的近似线性关系

参 考 文 献

1 Gao J B, Gunn S R, Ham's C J. A probabilistic framework for SVM regression and Error Bar Estimation [J]. Machine Learning, 2002, 46: 71～89

2 Kwok J T. The evidence framework applied to support vector machines [J]. IEEE Transactions on Neural Networks, 2000, 11(5): 1162～1173

3 Kwok J T, Tsang I W. Linear dependency between ε and the input noise in ε-support vector regression. IEEE Transaction on Neural Networks [J], 2003, 14(3): 544～553

4 Wang Shitong, Zhu Jiagang, F. L. Chung, et al. Theoretically optimal parameter choices for support vector regression machines Huber-SVR and Norm_r-SVR with noisy input. Soft Computing [J], 2005, 9(10): 732～741

5 朱嘉钢，王士同. Huber-SVR 中参数 μ 与输入噪声间关系的研究 [J]. 复旦学报, 2004, 43(5): 793～796

6 Cherkassky V, Ma Yunqian. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks [J], 2004, 17(1): 113～126

(上接第 5 页)

10 Callon R. The twelve networking truths. RFC1925, April 1996

11 Kempf J, Austein R. The rise of the middle and the future of End-to-End: reflections on the evolution of the Internet architecture. RFC 3724, March 2004

12 Coulouris G, Dollimore J, Kindberg T. Distributed systems concepts and design. 4th Edition. Addison Wesley, 2005

13 El-Sayed A, Roca V, Mathy L. A survey of proposals for an alternative group communication service. IEEE Network, 2003, 17 (1): 46～51

14 McCabe J D. Network analysis, architecture, and design. Second Edition. San Francisco, CA, USA: Morgan Kaufmann. Publishers, 2003

15 Tanenbaum A S. Computer networks. Fourth Edition. NJ: Prentice Hall Inc, 2002

16 Peterson L L, Davie B S. Computer networks: a systems approach. 3rd Edition. San Francisco CA: Morgan Kaufmann Publishers, 2003

17 Braden R, Clark D, Shenker S, et al. Developing a next-generation Internet architecture. July 2000. http://www.isi.edu/newarch/WhitePaper.pdf

18 Blake S, Black D, Carlson M, et al. An architecture for Differentiated Services. RFC 2475, December 1998

19 Molva R. Internet security architecture. Computer Networks, 1999, 31(8): 787～804

20 Leiner B M, Cerf V G, Clark D D, et al. A brief history of the Internet. http://www.isoc.org/internet/history/brief.shtml. 2003-12-10

21 Roberts J W. Traffic theory and the Internet. IEEE Communication Magazine, 2001, 39(1): 94～99

22 Leland W E, Taqqu M S, Willinger W, et al. On the self-similar nature of Ethernet traffic (Extended Version). IEEE/ACM Transactions on Networking, 1994, 2(1): 1～15

23 Paxson V, Floyd S. Wide area traffic: the failure of Poisson modeling. IEEE/ACM Transactions on Networking, 1995, 3(3): 226～244

24 Beran J, Sherman R, Taqqu M S, et al. Long-range dependence in variable-bit-rate video traffic. IEEE Transactions Communications, 1995, 43: 1566～1579

25 Crovella M E, Bestavros A. Self-similarity in World Wide Web traffic: evidence and Possible causes. IEEE/ACM Transactions on Networking, 1997, 5(6): 836～846

26 Carlson J M, Doyle J. Highly optimized tolerance: a mechanism for power laws in designed systems. Physics Review E, 1999, 60 (2): 1412～1427

27 Albert R, Barabasi A-L. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002, 74: 47～97

28 Wang X F, Chen G R. Complex networks: Small-World, scale-free and beyond. IEEE Circuits and Systems Magazine, 2003, 3 (1): 6～20

29 Barabasi A-L, Bonabeau E. Scale-free networks. Scientific American, May 2003. 50～59

30 Clark D, Tennenhouse D. Architectural considerations for a new generation of protocols. Computer Communications Review, 1990, 20(4): 200～208

31 Clark D D, Sollins K R, et al. Addressing reality: an architectural response to real-world demands on the evolving internet. In: ACM SIGCOMM 2003 FDNA Workshop, Karlsruhe, August 2003

32 Lampson B W. Hints for computer system design. ACM Operating Systems Review, SIGOPS, 1983, 15(5): 33～48

33 Carpenter B. Architectural principles of the Internet. RFC1958, June 1996

34 Braden B. Architectural principles of the Internet. IPAM Tutorial, 2002. http://www.ipam.ucla.edu/publications/cntut/cntut_1494.pdf

35 Bush R, Meyer D. Some Internet architectural guidelines and philosophy. RFC 3439, December 2002

36 Floyd S. General architectural and policy considerations. RFC3426, November 2002