http://ansinet.com/itj

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 10 (12): 2276-2283, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.2276.2283


# Inverse Proportional Dependency Between the Parameter H and the Input Noise in Possibilistic Linear Model

Hongwei Ge, Shitong Wang and Wei Song
School of Internet of Things, Jiangnan University, Wuxi, Jiangsu, China

**Abstract:** In order to determine the optimal choice of the parameter h (i.e., the threshold value used to measure degree of fit) in Possibilistic Linear Model (PLM) with the existence of noisy input, the dependency between h and the input noise is studied. PLM is first extended to its regularized version, i.e., Regularized Possibilistic Linear Model RPLM; then RPLM is interpreted as the corresponding equivalent Maximum a Posteriori (MAP) problem; finally, the approximately inverse proportional dependency relationships that the parameter h with Laplacian noisy input and Uniform noisy input should follow are derived, respectively. At the same time, the emulation results also proved the conclusion. Obviously, this important conclusion is very helpful for the practical applications of both PLM and RPLM.

**Key words:** Possibility theory, possibilistic linear model, MAP

## INTRODUCTION

In the analysis of fuzzy data, fuzzy regression analysis plays a pivotal role and has been successfully applied to various problems such as forecasting and engineering. Fuzzy regression models can be classified into two classes. The first class includes Tanaka's method and its extension, i.e., possibilistic regression model which is based on a possibility distribution that reflects the membership values of the dependent variable rather than a probability distribution. Also, the relation between the dependent variable and independent variables is defined by using fuzzy concept rather than statistical concept. In this class, the objective is to minimize the total spread of the fuzzy regression coefficients subject to the constraint that the regression model needed to satisfy a pre-specified membership value in estimating the fuzzy responses. The second class adopts Fuzzy Least Square Method (FLSM) to minimize the total square of errors in the estimated value. The advantage of Tanaka's model is its simplicity in programming and computation, while that of FLSM is its minimum degree of fuzziness between the observed and estimated values. Here, possibilistic regression model is our main concern.

Since possibilistic linear regression analysis was first introduced by Tanaka, the literature dealing with possibilistic regression has grown rapidly. For example, a modified version of Tanaka's possibilistic regression model was given by Savic and Pedrycz (1991) where possibilistic regression for fuzzy input-output data was considered. Yen *et al*. (1999) extended the results of a possibilistic linear regression model that uses symmetric triangular coefficients to one with non-symmetric fuzzy triangular coefficients successfully. Peters (1994) introduced fuzzy linear programming into the modified Tanaka's model. The important properties of possibilistic regression have been studied by Tanaka (1987), Tanaka and Watada (1988) and Redden and Woodall (1994). Especially Hong and Hwang (2003) introduced the support vector technique into possibilistic regression analysis to enhance its generalization capability. However, up to date, very little attention has been paid on how to choose the parameter h (i.e., the threshold value used to measure degree of fit) for possibilistic linear regression model with noisy input. In practice, the input data often contain noise. With the existence of noise, one very interesting but challenging issue is how to determine the parameter h in possibilistic regression models. A bad choice for threshold h will heavily deteriorate the performances of these models. Ge *et al*. (2008) investigated the theoretically optimal choices of the free parameter h with Gaussian noisy input. Although, Gaussian noise is typically adopted in most robustness analyses, there remain other types of noise such as Laplacian distribution noise and uniform noise in real datasets. So, in this study, particular attention is paid on the dependency relationships between free threshold h and the standard deviations of Laplacian noisy input and uniform noisy input.

In this study, we first extend the PLM to its regularized version, i.e., Regularized Possibilistic Linear Model (RPLM), so as to enhance its generalization

**Corresponding Author:** Hongwei Ge, School of Internet of Things, Jiangnan University, Wuxi, Jiangsu, China

capability. Then, we explain this regularized model using maximum a posteriori MAP framework (Gao *et al*., 2002; Kwok and Tsang, 2003; Shitong *et al*., 2005). Finally, with the help of this MAP framework, we prove that there exists an approximately inverse proportional dependency between the parameter h and the standard deviation of input noise, whether the input noise is Laplacian noise or uniform noise.

## POSSIBILISTIC LINEAR MODEL AND MAP

**Possibilistic linear model:** A Possibilistic Linear Model (PLM) can be stated as follows:

$$\underset{\sim}{Y}=f(x,\underset{\sim}{A})=\underset{\sim}{A}_0+\underset{\sim}{A}_1x_1+\underset{\sim}{A}_2x_2+...+\underset{\sim}{A}_nx_n \quad (1)$$

where, $x = (x_1, x_2,..., x_n)^T$ is a vector of non-fuzzy inputs and

$$\underset{\sim}{A}=(\underset{\sim}{A}_0,\underset{\sim}{A}_1,\underset{\sim}{A}_2,...,\underset{\sim}{A}_n)$$

is a vector of fuzzy model parameters. $A_i$ $(0\le i\le n)$ have the symmetrical triangular membership functions:

$$\mu_{\underset{\sim}{A}_i}(a_i)=\begin{cases}1-\dfrac{|a_i-\xi_i|}{\eta_i}, & \xi_i-\eta_i\le a_i\le\xi_i+\eta_i,\\ 0, & \text{otherwise},\end{cases}$$

where, $\xi_i$ is the center and $\eta_i$ is the spread of $A_i$. In vector notation, the fuzzy parameter $\underset{\sim}{A}$ can be written as $A = (\xi, \eta)$, where, $\xi\ (\xi_0, \xi_1, \xi_2,..., \xi_n)^T$ and $\eta = (\eta_0, \eta_1, \eta_2,..., \eta_n)^T$. The estimated output Y can be obtained by using the extension principle. It has the membership function:

$$\mu_{\underset{\sim}{Y}}(y)=\begin{cases}1-\dfrac{|y-\xi^Tx|}{\eta^T|x|}, & x\neq 0,\\ 1, & x=0, y=0,\\ 0, & x=0, y\neq 0,\end{cases}$$

where, $|x| = (1, |x_1|, |x_2|,..., |x_n|)^T$. The center of Y is $\xi^T x$ and the spread of Y is $\eta^T |x|$ with $\eta^T |x|\ge 0$.

Assume we have a dataset with n-dimensional non-fuzzy input x and one-dimensional non-fuzzy output variable y:

$$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}, x_i\in R^{n+1}, y_i\in R, i=1,2,\cdots,N \quad (2)$$

We can establish the possibilistic linear model $\underset{\sim}{Y}_i^*=\underset{\sim}{A}^{*T}x_i$, where $x_i = (1, x_{i1}, x_{i2},..., x_{in})^T$, $\underset{\sim}{A}=(\underset{\sim}{A}_0,\underset{\sim}{A}_1,\underset{\sim}{A}_2,...,\underset{\sim}{A}_n)^T$, $\underset{\sim}{A}^*=(\underset{\sim}{A}_0^*,\underset{\sim}{A}_1^*,\underset{\sim}{A}_2^*,...,\underset{\sim}{A}_n^*)^T$, i = 1,2,..., N and $\underset{\sim}{Y}_i^*, \underset{\sim}{A}^*$ denote the estimates of $\underset{\sim}{Y}_i, \underset{\sim}{A}$, respectively.

In possibilistic linear regression model, we often need to choose an appropriate free parameter (i.e., threshold h ($h\in[0,1]$)), such that $\mu_{\underset{\sim}{Y}_i^*}(Y_i)\ge h$, i.e.,

$$\mu_{\underset{\sim}{Y}_i^*}(y_i)=1-\frac{|y_i-\xi^Tx_i|}{\eta^T|x_i|}\ge h$$

The parameter h plays a key role in the possibilistic linear regression model. If we take,

$$\sum_{i=1}^{N}\eta^T|x_i|$$

as the objective index, then possibilistic linear regression analysis in Eq. 1 will become the following optimization problem (Yen *et al*., 1999; Shitong *et al*., 2004; Jinpei, 1999):

$$\min\Phi(\eta,\xi,\Delta)=\sum_i\eta^T|x_i|+C\sum_i(\Delta_i^-+\Delta_i^+) \quad (3)$$

$$\text{s.t.}\begin{cases}\dfrac{y_i-\xi^Tx_i}{\eta^T|x_i|}\le(1-h)+\Delta_i^-,\\ \dfrac{\xi^Tx_i-y_i}{\eta^T|x_i|}\le(1-h)+\Delta_i^+,\\ \Delta_i^-,\Delta_i^+\ge 0,\\ i=1,2,\cdots,N,\end{cases}$$

where, C is a predefined constant. $\Delta^-$ and $\Delta^+$ denote the latent variables of the upper/lower bounds of the output, respectively. Of course, the PLM without delta in Eq. 3 corresponds to the original fuzzy regression model proposed by Tanaka (1987) and Tanaka and Watada (1988).

In order to enhance PLM's generalization capability and following the spirit of support vector regression techniques in (Hong and Hwang, 2003; Shitong *et al*., 2005; Law and Kwok, 2001; Cherkassky and Ma, 2004), we extend FLR model to RFLR model, i.e., introduce a regularized term $1/2\xi^T\xi$ into Eq. 3, then we have:

$$\min\Phi(\eta,\xi,\Delta)=\sum_i\eta^T|x_i|+\frac{1}{2}\xi^T\xi+C\sum_i(\Delta_i^-+\Delta_i^+) \quad (4)$$

In the rest of present study, we will study how to determine an appropriate threshold h for PLM, particularly in the cases of having Laplacian noisy input and uniform noisy input. Obviously, the conclusion we draw will also be very helpful for apply PLM regression techniques in the practical applications.

**The regularized model and MAP:** Here, by using the evidence theory (Gao *et al*., 2002; Kwok and Tsang, 2003), we demonstrate that RPLM is equivalent to Maximum A Posteriori (MAP) estimation.

Now, we are interested in obtaining a weight vector $\xi$ in RPLM such that:

$$y_i = \xi^T x_i + n_i, i = 1, 2, \cdots, N$$

for all the data in dataset (2), where all the data $x_i$ follow distribution p(.) and all $n_i$ are i.i.d noise following some distribution $\phi$(.). Thus, the corresponding density function on y can be denoted as:

$$p(y \mid x) = \phi(y - \xi^T x)$$

The degree of such an approximation may be measured by the following loss function (L ($\xi^T$ x, $\eta^T$ x,y):

$$L(\xi^T x, \eta^T x, y) = \begin{cases} 0, & \frac{|\xi^T x - y|}{\eta^T |x|} \le 1 - h, \\ \frac{|\xi^T x - y|}{\eta^T |x|} - 1 + h, & \text{otherwise,} \end{cases}$$

Assume L ($\xi^T$ x, $\eta^T$ x, y) leads to the following Gaussian probability density function on y:

$$p(y_i \mid x_i, \xi, \eta, \beta, h) = \frac{1}{C(\beta, h)} \exp[-\beta L(\xi^T x_i, \eta^T x_i, y_i)]$$

where, $\beta$ is a parameter and:

$$C(\beta, h) = \iint_D \exp[-\beta L(\xi^T x, \eta^T x, y)] dx dy$$

Just as the mathematical analysis in (Gao *et al*., 2002; Kwok and Tsang, 2003) does, we adopt the Gaussian prior in the following analytical framework. With the Gaussian prior on $\eta$ and $\xi$, i.e.,

$$p(\eta \mid \alpha, x_i) = \frac{1}{M(\alpha)} \exp(-\alpha \eta^T \mid x_i \mid) \text{ , where } M(\alpha) = \int \exp(-\alpha \eta^T \mid x_i \mid) d\eta$$

$$p(\xi \mid \gamma) = \frac{1}{F(\gamma)} \exp\left(-\frac{\gamma}{2} \xi^T \xi\right) \text{, where } F(\gamma) = \int \exp\left(-\frac{\gamma}{2} \xi^T \xi\right) d\xi$$

and by applying the Bayes rule:

$$p(\eta, \xi \mid D, \beta, h) \propto p(D \mid \eta, \xi, \beta, h) p(\xi \mid \gamma) \prod_{i=1}^{N} p(\eta \mid \alpha, x_i) p(x_i)$$

we have,

$$\ln p(\eta, \xi \mid D, \beta, h) =$$

$$-\alpha \sum_{i=1}^{N} \eta^T |x_i| - \frac{\gamma}{2} \xi^T \xi - \beta \sum_{i=1}^{N} L(\eta^T x_i, \xi^T x_i, y_i) - N \ln C(\beta, h) + \text{const} \quad (5)$$

When $\gamma = \alpha$ and $c = \beta/\alpha$, optimizing Eq. 4 can be interpreted as finding the MAP estimates of $\eta,\xi$ for the given values of $\beta$,h. In other words, RPLM in Eq. 4 is equivalent to the maximum a posteriori MAP problem in Eq. 5.

In order to make the analysis easier,

$$\frac{1}{N} \sum_{i=1}^{N} L(\eta^T x_i, \xi^T x_i, y_i)$$

in Eq. 5 is replaced by its expectation:

$$\begin{aligned} E(L(\eta^T x, \xi^T x, y)) &= \iint_D L(\eta^T x, \xi^T x, y) p(y \mid x) p(x) dx dy \\ &= \int_D \left[ \int_{-\infty}^{\xi^T x - (1-h)\eta^T |x|} \left( \frac{\xi^T x - y}{\eta^T |x|} - 1 + h \right) p(y \mid x) dy + \right. \\ &\left. \int_{\xi^T x + (1-h)\eta^T |x|}^{+\infty} \left( \frac{y - \xi^T x}{\eta^T |x|} - 1 + h \right) p(y \mid x) dy \right] p(x) dx \end{aligned} \quad (6)$$

and

$$\frac{1}{N} \sum_{i=1}^{N} \eta^T |x_i|$$

in Eq. 5 can also be approximated by:

$$E(\eta^T |x|) = \int_D \eta^T |x| p(x) dx$$

Thus, Eq. 5 becomes:

$$\ln p(\eta, \xi \mid D, \beta, h) =$$

$$-\alpha N * E(\eta^T |x|) - \frac{\gamma}{2} \xi^T \xi - \beta N * E(L(\eta^T x, \xi^T x, y)) - N \ln C(\beta, h) + \text{const} \quad (7)$$

According to the corresponding MAP estimation, C ($\beta$,h) in Eq. 7 can be rewritten as:

$$\begin{aligned} C(\beta, h) &= 2\left( \int_0^{1-h} \exp(0) dt + \int_{1-h}^{+\infty} \exp(-\beta(t - 1 + h)) dt \right) \\ &= \frac{2(1 + (1-h)\beta)}{\beta} \end{aligned} \quad (8)$$

In order to maximize Eq. 7, its derivatives with respect to $\eta$, $\xi$, $\beta$, h must be zeros. After a little tedious computation, we have:

$$E(L(\hat{\eta}^T x, \hat{\xi}^T x, y)) = \frac{1}{\beta(1 + (1-h)\beta)} \quad (9)$$

$$\int_D\left[\int_{-\infty}^{\hat{\xi}^T x-(1-h)\hat{\eta}^T|x|} p(y\mid x)dy+\int_{\hat{\xi}^T x+(1-h)?^T|x|}^{+\infty} p(y\mid x)dy\right]p(x)dx=\frac{1}{1+(1-h)\beta} \tag{10}$$

Furthermore, when $\eta=\hat{\eta},\xi=\hat{\xi}$, it can be shown that maximizing Eq. 5 actually becomes the following optimization problem:

$$\arg\min_{\alpha,\beta,h} \alpha E(\hat{\eta}^T\,|x|)+\beta E(L(\hat{\eta}^T x,\hat{\xi}^T x,y))-\ln\frac{1}{1+(1-h)\beta}+\ln\frac{1}{\beta} \tag{11}$$

## THE DEPENDENCY BETWEEN H AND THE INPUT NOISE

**Laplacian noise:** For the Laplacian noise model:

$$p(y\mid x)=\frac{1}{2\mu}\exp\left[-\frac{\left|y-\tilde{\xi}^T x\right|}{\mu}\right]$$

Here, we only consider the simpler case when x is one-dimensional with uniform density over the range [-L,L] (i.e., X~U([-L,L]). When x is multi-dimensional, we can also arrive at the similar conclusion, if we select suitable integral region for each dimension.

Let $\varepsilon=1-h, E=\int_D \hat{\eta}\,|x|\,p(x)dx$. After tedious computation, it can be shown that Eq. 10 reduce to:

$$\frac{1}{1+\varepsilon\beta}=\int_D\left[\int_{-\infty}^{\hat{\xi}x-\varepsilon\hat{\eta}|x|}\frac{1}{2\mu}\exp\left(-\frac{\left|y-\tilde{\xi}x\right|}{\mu}\right)dy+\int_{\hat{\xi}x+\varepsilon\hat{\eta}|x|}^{+\infty}\frac{1}{2\mu}\exp\left(-\frac{\left|y-\tilde{\xi}x\right|}{\mu}\right)dy\right]p(x)dx$$

$$=2\int_{-L}^{0}\left(\int_{\hat{\xi}x+\varepsilon\hat{\eta}x}^{+\infty}\frac{1}{2\mu}\exp\left(-\frac{y-\tilde{\xi}x}{\mu}\right)dy\right)\frac{1}{2L}dx+2\int_{0}^{L}\left(\int_{\tilde{\xi}x}^{+\infty}\frac{1}{2\mu}\exp\left(-\frac{y-\tilde{\xi}x}{\mu}\right)dy\right)\frac{1}{2L}dx$$

$$+2\int_{-L}^{0}\left(\int_{\tilde{\xi}x}^{\hat{\xi}x-\varepsilon\hat{\eta}x}\frac{1}{2\mu}\exp\left(-\frac{y-\tilde{\xi}x}{\mu}\right)dy\right)\frac{1}{2L}dx$$

$$=1+\frac{\mu}{2L(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})}\left(1-\exp\left(\frac{-(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})L}{\mu}\right)\right)+\frac{\mu}{2L(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})}\left(\exp\left(\frac{-(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})L}{\mu}\right)-1\right) \tag{12}$$

Similarly, from Eq. 6, 9 and 10, we obtain:

$$\frac{1}{\beta}=\frac{(\tilde{\xi}-\hat{\xi})L}{2E}+\frac{\mu^2}{LE}\cdot\frac{(\tilde{\xi}-\hat{\xi})^3-3(\tilde{\xi}-\hat{\xi})\varepsilon^2\hat{\eta}^2}{\left((\tilde{\xi}-\hat{\xi})^2-\varepsilon^2\hat{\eta}^2\right)^2}+\frac{\mu}{2E}\exp\left(-\frac{(\tilde{\xi}-\hat{\xi})L}{\mu}\right)$$
$$\left(\frac{(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})\varepsilon\hat{\eta}L-(\tilde{\xi}-\hat{\xi})\mu+2\mu\varepsilon\hat{\eta}}{(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})^2L}\exp\left(\frac{\varepsilon\hat{\eta}L}{\mu}\right)\right.$$
$$\left.-\frac{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})\varepsilon\hat{\eta}L+(\tilde{\xi}-\hat{\xi})\mu+2\mu\varepsilon\hat{\eta}}{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})^2L}\exp\left(-\frac{\varepsilon\hat{\eta}L}{\mu}\right)\right) \tag{13}$$

By using Eq. 9, 12 and 13, we have:

$$E(L(\hat{\eta}x,\hat{\xi}x,y))=\frac{(\tilde{\xi}-\hat{\xi})L-2\varepsilon E}{2E}$$
$$+\frac{\mu}{LE}\cdot\frac{(\tilde{\xi}-\hat{\xi})^3\mu-(\tilde{\xi}-\hat{\xi})^2\varepsilon^2\hat{\eta}E-3(\tilde{\xi}-\hat{\xi})\varepsilon^2\hat{\eta}^2\mu+\varepsilon^4\hat{\eta}^3E}{\left((\tilde{\xi}-\hat{\xi})^2-\varepsilon^2\hat{\eta}^2\right)^2}+\frac{\mu}{2E}\exp\left(-\frac{(\tilde{\xi}-\hat{\xi})L}{\mu}\right)$$
$$\left(\frac{(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})\varepsilon\hat{\eta}L-(\tilde{\xi}-\hat{\xi})\mu+2\mu\varepsilon\hat{\eta}+(\tilde{\xi}-\hat{\xi})\varepsilon E-\varepsilon^2\hat{\eta}E}{(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})^2L}\exp\left(\frac{\varepsilon\hat{\eta}L}{\mu}\right)\right.$$
$$\left.-\frac{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})\varepsilon\hat{\eta}L+(\tilde{\xi}-\hat{\xi})\mu+2\mu\varepsilon\hat{\eta}+(\tilde{\xi}-\hat{\xi})\varepsilon E+\varepsilon^2\hat{\eta}E}{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})^2L}\exp\left(-\frac{\varepsilon\hat{\eta}L}{\mu}\right)\right) \tag{14}$$

We have defined $E=\int_D\hat{\eta}|x|p(x)dx, C=\beta/\alpha$, so we obtain:

$$\alpha E(\hat{\eta}|x|)=\beta EC^{-1} \tag{15}$$

and

$$\varepsilon\hat{\eta}L=2\varepsilon E \tag{16}$$

Define $y-\tilde{\xi}x\approx\hat{\xi}x-\tilde{\xi}x=\delta(x)$. It can be shown that:

$$E_X\left(\delta^2(x)\right)\approx E_X\left((y-\tilde{\xi}x)^2\right)=2\mu^2 \tag{17}$$

With X~U ([l,l]), we also obtain var (X) = $L^2/3$. Consequently:

$$E_X\left(\delta^2(x)\right)\approx \mathrm{var}\left((\tilde{\xi}-\hat{\xi})X\right)\approx(\tilde{\xi}-\hat{\xi})^2L^2/3 \tag{18}$$

and hence,

$$\mu=(\tilde{\xi}-\hat{\xi})L/\sqrt{6} \tag{19}$$

After substituting Eq. 12, 13, 14, 15, 16 and 19 into Eq. 11, minimizing Eq. 11 becomes minimizing the following:

$$g\left(\frac{\varepsilon}{\mu}\right)=\left(\frac{\sqrt{6}}{2E}+f_1\left(\frac{\varepsilon}{\mu}\right)+\frac{\exp\left(-\sqrt{6}\right)}{4E}f_2\left(\frac{\varepsilon}{\mu}\right)\right)^{-1}\left(\frac{\sqrt{6}-2\frac{\varepsilon E}{\mu}}{2E}+\frac{3\sqrt{6}-\left(6+6\sqrt{6}\right)\frac{\varepsilon^2E^2}{\mu^2}+4\frac{\varepsilon^4E^4}{\mu^4}}{18E-24\frac{\varepsilon^2E^3}{\mu^2}+8\frac{\varepsilon^4E^5}{\mu^4}}+\frac{\exp\left(-\sqrt{6}\right)}{4E}f_3\left(\frac{\varepsilon}{\mu}\right)+\frac{E}{\mu C}\right)$$
$$-\ln\left(1+\frac{\frac{\varepsilon E}{\mu}}{3-2\frac{\varepsilon^2E^3}{\mu^2}}+\frac{\exp(-\sqrt{6})}{2}\left(\frac{\exp\left(-2\frac{\varepsilon E}{\mu}\right)}{\sqrt{6}+2\frac{\varepsilon E}{\mu}}-\frac{\exp\left(2\frac{\varepsilon E}{\mu}\right)}{\sqrt{6}-2\frac{\varepsilon E}{\mu}}\right)\right)+\ln\left(\frac{\sqrt{6}}{2E}+f_1\left(\frac{\varepsilon}{\mu}\right)+\frac{\exp\left(-\sqrt{6}\right)}{4E}f_2\left(\frac{E}{\mu}\right)\right)+\mathrm{const} \tag{20}$$

Where:

$$f_1\left(\frac{\varepsilon}{\sigma}\right)=\frac{3\sqrt{6}-6\sqrt{6}\frac{\varepsilon^2E^2}{\mu^2}}{18E-24\frac{\varepsilon^2E^3}{\mu^2}+8\frac{\varepsilon^4E^5}{\mu^4}}$$

$$f_2\left(\frac{\varepsilon}{\sigma}\right)=\frac{-\sqrt{6}+\left(4+2\sqrt{6}\right)\frac{\varepsilon E}{\mu}-4\frac{\varepsilon^2E^2}{\mu^2}}{3-2\sqrt{6}\frac{\varepsilon E}{\mu}+2\frac{\varepsilon^2E^2}{\mu^2}}\exp\left(2\frac{\varepsilon E}{\mu}\right)-\frac{\sqrt{6}+\left(4+2\sqrt{6}\right)\frac{\varepsilon E}{\mu}+4\frac{\varepsilon^2E^2}{\mu^2}}{3+2\sqrt{6}\frac{\varepsilon E}{\mu}+2\frac{\varepsilon^2E^2}{\mu^2}}\exp\left(-2\frac{\varepsilon E}{\mu}\right)$$

$$f_3\left(\frac{\varepsilon}{\sigma}\right)=\frac{-\sqrt{6}+\left(4+3\sqrt{6}\right)\frac{\varepsilon E}{\mu}-6\frac{\varepsilon^2E^2}{\mu^2}}{3-2\sqrt{6}\frac{\varepsilon E}{\mu}+2\frac{\varepsilon^2E^2}{\mu^2}}\exp\left(2\frac{\varepsilon E}{\mu}\right)-\frac{\sqrt{6}+\left(4+3\sqrt{6}\right)\frac{\varepsilon E}{\mu}+6\frac{\varepsilon^2E^2}{\mu^2}}{3+2\sqrt{6}\frac{\varepsilon E}{\mu}+2\frac{\varepsilon^2E^2}{\mu^2}}\exp\left(-2\frac{\varepsilon E}{\mu}\right)$$

Obviously, when $\frac{\varepsilon}{\mu}$ takes some fixed value:

$$g\left(\frac{\varepsilon}{\mu}\right)$$

will achieve its minimum which indicates that there is a linear dependency between $\varepsilon$ and $\mu$. Thus, there is an inverse linear dependency between the parameter h and the standard variance $\sigma$ of the Laplacian input noise because of $\varepsilon$ = 1-h and $\sigma=\sqrt{2}\mu$.

**Uniform noise:** Next, we consider the uniform noise model:

$$\phi(n)=\frac{1}{2\mu},\quad n\in[-\mu,\mu]$$

Again, we only consider the simpler case when x is one-dimensional, with uniform density over the range [-L,L]. Let $\varepsilon=1-h, E=\int_D \hat{\eta}|x|p(x)dx$. Eq. 10 becomes:

$$\begin{aligned}\frac{1}{1+\varepsilon\beta}&=\int_D\left[\int_{-\infty}^{\hat{\xi}x-\varepsilon\hat{\eta}|x|}\frac{1}{2\mu}dy+\int_{\hat{\xi}x+\varepsilon\hat{\eta}|x|}^{+\infty}\frac{1}{2\mu}dy\right]p(x)dx\\&=2\int_0^L\left(\int_{\hat{\xi}x+\varepsilon\hat{\eta}x}^{\tilde{\xi}x+\mu}\frac{1}{2\mu}dy\right)\frac{1}{2L}dx+2\int_0^{\frac{\mu}{\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta}}}\left(\int_{\tilde{\xi}x-\mu}^{\hat{\xi}x-\varepsilon\hat{\eta}x}\frac{1}{2\mu}dy\right)\frac{1}{2L}dx\\&=\frac{(\tilde{\xi}-\hat{\xi})^2L^2-\varepsilon^2\hat{\eta}^2L^2+2\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})L+\mu^2}{4\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})L}\end{aligned}\tag{21}$$

Using Eq. 6, 9, 10 and 21, after a little tedious computation, it can be shown that:

$$\begin{aligned}\frac{1}{\beta}&=\frac{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})^2L^3+3((\tilde{\xi}-\hat{\xi})^2-\varepsilon^2\hat{\eta}^2)L^2(\mu+\varepsilon E)}{12\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})LE}\\&+\frac{3\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})L(\mu+2\varepsilon E)+3\mu^2\varepsilon E+\mu^3}{12\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})LE}\end{aligned}\tag{22}$$

$$\begin{aligned}E(L(\hat{\eta}x,\hat{\xi}x,y))&=\frac{(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})^2L^3+3\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})(\tilde{\xi}-\hat{\xi}-\varepsilon\hat{\eta})L^2}{12\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})LE}\\&+\frac{3\mu^2(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})L+\mu^3}{12\mu(\tilde{\xi}-\hat{\xi}+\varepsilon\hat{\eta})LE}\end{aligned}\tag{23}$$

As in Laptacian Noise, consider $\delta(x)=(\hat{\xi}-\tilde{\xi})x$ and assume that:

$$E_X\left(\delta^2(x)\right)\approx E_X\left((y-\tilde{\xi}x)^2\right)=\mu^2/3\tag{24}$$

Using Eq. 18 again, we obtain:

$$\mu=(\tilde{\xi}-\hat{\xi})L\tag{25}$$

Analogous to Laptacian Noise, by plugging in Eq. 15, 16, 21, 22, 23 and 25, it can be shown that the problem reduces to minimizing:

$$\begin{aligned}g\left(\frac{\varepsilon}{\mu}\right)=&\left(\frac{3E+6\frac{\varepsilon E^2}{\mu}}{2+4\frac{\varepsilon E}{\mu}-\frac{\varepsilon^2E^2}{\mu^2}-\frac{\varepsilon^3E^3}{\mu^3}}\right)\left(\frac{2+\frac{\varepsilon E}{\mu}-4\frac{\varepsilon^2E^2}{\mu^2}+2\frac{\varepsilon^3E^3}{\mu^3}}{3E+6\frac{\varepsilon E^2}{\mu}}+\frac{E}{\mu C}\right)\\&-\ln\left(\frac{1+\frac{\varepsilon E}{\mu}-\frac{\varepsilon^2E^2}{\mu^2}}{1+2\frac{\varepsilon E}{\mu}}\right)+\ln\left(\frac{2+4\frac{\varepsilon E}{\mu}-\frac{\varepsilon^2E^2}{\mu^2}-\frac{\varepsilon^3E^3}{\mu^3}}{3E+6\frac{\varepsilon E^2}{\mu}}\right)+\text{const}\end{aligned}\tag{26}$$

Again, we can see from Eq. 26 that there is a linear dependency between $\varepsilon$ and $\mu$. Obviously, because the standard variance $\sigma$ of the uniform input noise equals to $\mu/\sqrt{3}$ and $\varepsilon$ = 1-h, there is an the inverse linear dependency relationship between the parameter h and $\sigma$.

**Approximately inverse proportional dependency between h and the input noise:** From Laptacian and uniform noise, we come to the conclusion that there are approximately inverse proportional dependency relationships between the parameter h and the standard variance $\sigma$ of Laplacian input noise and uniform input noise. At the same time, in (Ge *et al*., 2008), we have derived the analogical conclusion with Gaussian noisy input. Following the same spirit in (Kwok and Tsang, 2003), we can draw the more generic conclusion that the dependency between h and input noise is approximately inverse proportion.

It should be mentioned here that although, we investigate the dependency relationship between the parameter h in PLM and the input noise, the conclusion is very helpful for the practical applications of both PLM and RPLM, for PLM is just a special case of RPLM.

## EXPERIMENTAL RESULTS

In this experiment, we apply RPLM to a real dataset taken from (Coppi *et al*., 2006) (Table 1). Coppi *et al.*

Table 1: Input-output data in (16)

| Number | Relative Humidity (RH) (%) | Rain R (mm) | Wind speed WS (m $sec^{-1}$) | Concentrations of carbon mmonoxide (CO) mg $m^{-3}$ |
|---|---|---|---|---|
| 1 | 82.27 | 0.90 | 2.87 | 1.15 |
| 2 | 88.70 | 0.90 | 1.12 | 2.98 |
| 3 | 82.51 | 0.02 | 0.85 | 3.92 |
| 4 | 79.46 | 0.00 | 0.45 | 4.65 |
| 5 | 68.85 | 0.00 | 0.91 | 3.98 |
| 6 | 79.39 | 0.02 | 1.07 | 3.35 |
| 7 | 88.87 | 1.30 | 0.69 | 3.13 |
| 8 | 88.92 | 0.09 | 0.40 | 4.15 |
| 9 | 83.52 | 0.00 | 0.83 | 3.96 |
| 10 | 87.51 | 0.05 | 2.09 | 4.07 |
| 11 | 86.04 | 0.04 | 0.09 | 3.30 |
| 12 | 91.77 | 0.93 | 1.22 | 4.02 |
| 13 | 83.62 | 0.02 | 3.00 | 2.06 |
| 14 | 73.10 | 0.02 | 1.75 | 1.37 |
| 15 | 80.38 | 0.09 | 0.73 | 3.35 |
| 16 | 87.98 | 0.24 | 1.87 | 1.45 |
| 17 | 90.13 | 0.02 | 2.39 | 2.74 |
| 18 | 64.95 | 0.00 | 1.25 | 2.44 |
| 19 | 80.16 | 0.01 | 1.02 | 2.79 |
| 20 | 86.14 | 0.44 | 0.70 | 3.31 |
| 21 | 89.12 | 0.01 | 1.17 | 4.02 |

(2006) used fuzzy linear regression successfully to analyze the dependence relationship of the atmospheric concentration of Carbon monoxide (CO) on a set of meteorological variables, including Relative Humidity (RH) (%), Rain (R) (mm) and Wind Speed (WS) (m $sec^{-1}$).

By using the same method from Eq. 16, we also take CO as the output data, RH, R and WS as the input data. Let h = 0.5. Now the corresponding regression values $(\xi_i, \eta_i)$, = 1,2,...,21 can be obtained, where $\xi_i$ denotes the center and $\eta_i$ denotes the spread. Next, in order to investigate the dependency relationship between h and the noisy input, let $y'_i = y_i + k.n$, i = 1,2,..., 21, where k is a noise-signal ratio and n represents the Laplacian noise. Then the corresponding sampling dataset $(x_i, y'_i)$, i = 1,2,...,21 can be generated. Similarly, its corresponding regression values $(\xi'_i, \eta'_i)$ can be obtained by using the same RPLM. In order to make the experimental results fair, σ is taken from [0.1,2.0] with the step length 0.1 and the Laplacian noise distribution is used to generate 20 groups of the corresponding sampling datasets for each given σ. For each given σ, h is taken as the average result of all 20 h values which can minimize:

$$\sum_{i=1}^{21}\sqrt{(\xi_i-\xi_i^{'})^2+(\eta_i-\eta_i^{'})^2}$$

respectively for each group of the sampling datasets.

Figure 1a-c depict the dependency relationships between h and σ for all 20 σ values with different k (see + in the figures), where the curves are used to roughly indicate the change tendencies between h and σ, respectively. With the same experimental method as above and the uniform noise distribution, Fig. 2 a-c shows the experimental results which again validate our theoretical conclusion.

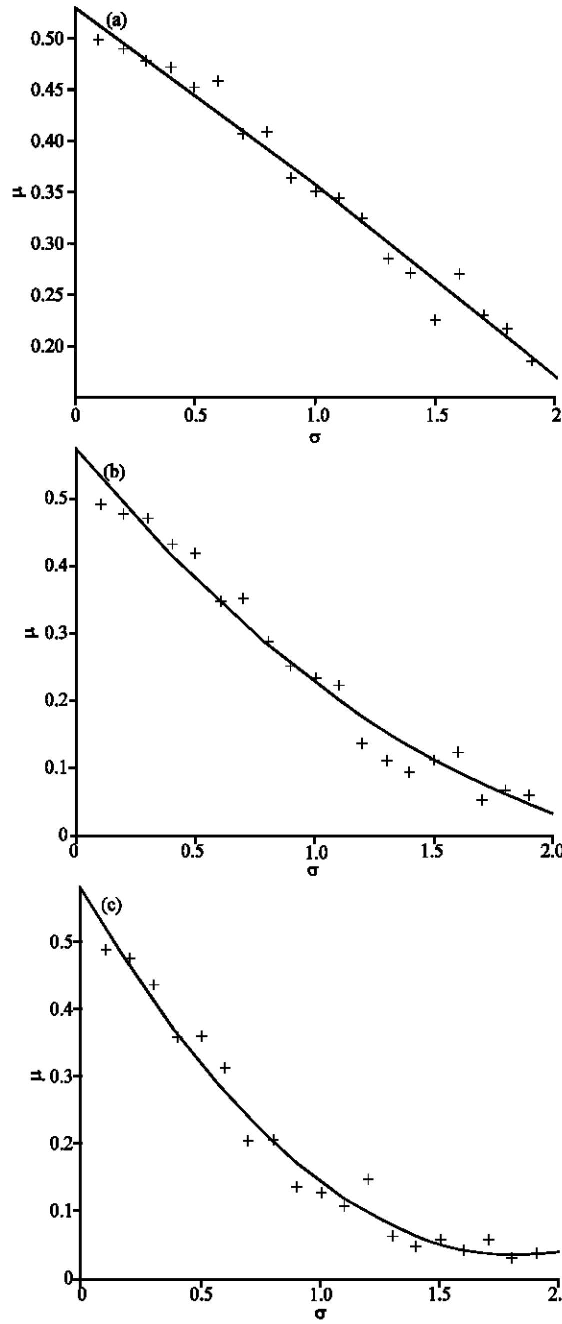


Fig. 1(a-c): With Laplacian noisy input, the relationship between h and σ when (a) k = 0.05, (b) k = 0.1, (c) k = 0.2

It can be easily seen from Fig. 1and 2 that when noise is small, i.e., k and σ is comparatively small,

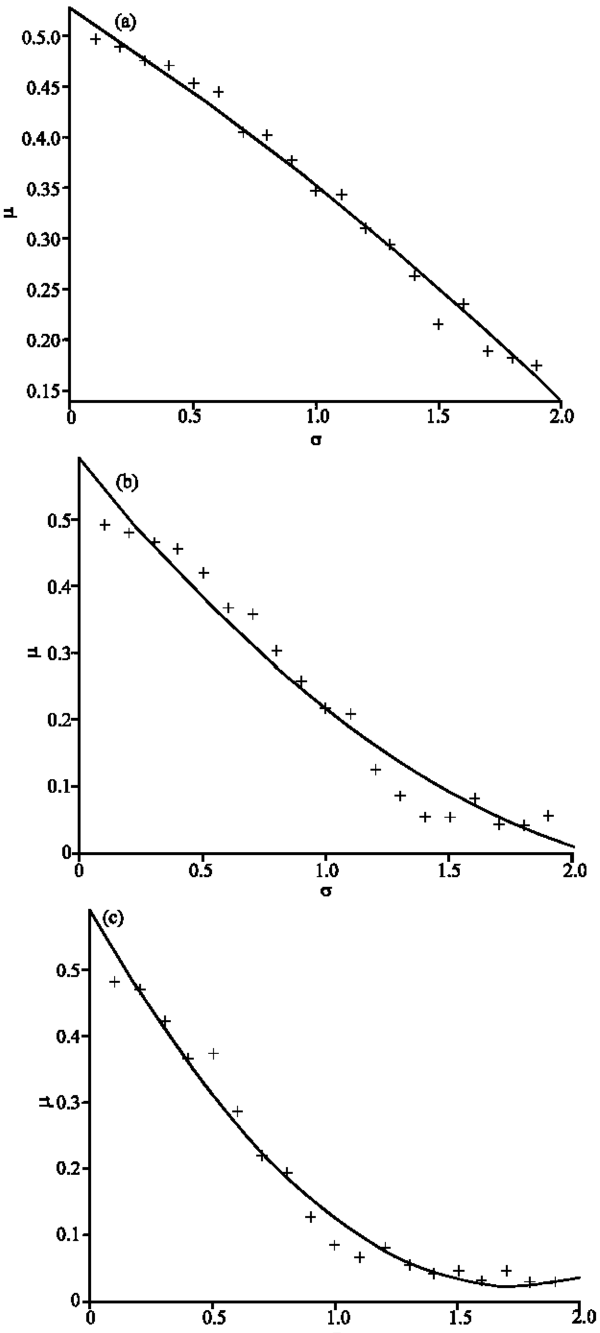


Fig. 2(a-c): With uniform noisy input, the relationship between h and σ when (a) k = 0.05, (b) k = 0.1 and (c) k = 0.2

there is an obvious inverse linear dependency relationship between h and σ. However, when k and/or σ are comparatively large, i.e., the datasets are seriously distorted, the inverse linear relationship between h and σ does not exist anymore. In other words, RPLM may become ineffective for seriously distorted datasets. Of course, although, we have arranged many experiments and all these experimental results validate the above32079obtained conclusion on RPLM, we just showed several experimental results here so as to save the paper's space.

## CONCLUSION

In present study, based on the Bayesian frame, the approximately inversely linear dependency relationships between the optimal h and the standard variance σ of Laplacian input noise and uniform input noise are derived. The experimental results also validate our theoretical conclusion. In fact, the optimal choices of the parameter h are actually dependent on the variance of the input noise. Because Gaussian noise, Laplacian noise and uniform noise are all typical noise and with the theoretical and experimental results from (Law and Kwok, 2001), we can arrive at the more generic conclusion that there is an approximately inverse proportional dependency between h and input noise. The theoretical result here is very useful for the optimal choices of the parameter h. Of course, in order to apply the inversely linear dependency result in practical applications, one has to first arrive at an estimate of the noise level σ. One way to obtain this is by using Bayesian methods (Law and Kwok, 2001). In the future, we will investigate better methods to solve the problem.

## ACKNOWLEDGMENTS

This study is supported by National Natural Science Foundation of China (Grant No. 60975027), New century Outstanding Young Scholar Grant of Ministry of Education of China.

## REFERENCES

Cherkassky, V. and Y. Ma, 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks, 17: 113-126.

Coppi, R., P. D'Urso, P. Giordani and A. Santoro, 2006. Least squares estimation of a linear regression model with LR fuzzy response. Comput. Stat. Data Anal., 51: 267-286.

Gao, J.B., S.R. Gunn, C.J. Harris and M. Brown, 2002. A probabilistic framework for SVM regression and error bar estimation. Mach. Learn., 46: 71-89.

Ge, H., F.L. Chung and S. Wang, 2008. Theoretical choice of the optimal threshold for possibilistic linear model with noisy input. IEEE Trans. Fuzzy Syst., 16: 1027-1037.

Hong, D.H. and C. Hwang, 2003. Support vector fuzzy regression machines. Fuzzy Sets Syst., 138: 271-281.

Jinpei, W., 1999. Device diagnosis based on possibility theory. Fuzzy Syst. Math., 13: 25-32.

Kwok, J.T. and I.W. Tsang, 2003. Linear dependency between ε and the input noise in ε-support vector regression. IEEE Trans. Neural Networks, 14: 544-553.

Law, M.H. and J.T. Kwok, 2001. Bayesian support vector regression. Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics, (AIS'01), Key West, FL., USA., pp: 239-244.

Peters, G., 1994. Fuzzy linear regression with fuzzy intervals. Fuzzy Sets Syst., 63: 45-55.

Redden, D.T. and W.H. Woodall, 1994. Properties of certain fuzzy linear regression methods. Fuzzy Sets Syst., 64: 361-375.

Savic, D.A. and W. Pedrycz, 1991. Evaluation of fuzzy linear regression models. Fuzzy Sets Syst., 39: 51-63.

Shitong, W., K.F. Chung, S. Hongbin and Z. Ruiqiang, 2004. Note on the relationship between probabilistic/fuzzy clustering. Soft Camput., 8: 366-369.

Shitong, W., Z. Jiagang, F.L. Chung, L. Qing and H. Dewen, 2005. Theoretically optimal parameter choices for support vector regression machines with noisy input. Soft Comput., 9: 732-741.

Tanaka, H. and J. Watada, 1988. Possibilistic linear systems and their application to the linear regression model. Fuzzy Sets Syst., 27: 275-289.

Tanaka, H., 1987. Fuzzy data analysis by possibilitic linear models. Fuzzy Sets Syst., 24: 363-375.

Yen, K.K., G. Gheoshary and G. Roig, 1999. A linear regression model using triangular fuzzy number coefficients. Fuzzy Sets Syst., 106: 167-177.