

Study of Dependency between the Input Noise and the Parameter in Fuzzy Linear Regression Model

Hongwei Ge, Shitong Wang and Wei Song
School of Internet of Things, Jiangnan University
Wuxi, Jiangshu, People's Republic of China
ghw8601@163.com

Abstract—When noise exists in data, it is a very meaningful topic to reveal the dependency between the parameter h (i.e. the threshold value used to measure degree of fit) in Fuzzy linear regression (FLR) model and the input noise. In this paper, the FLR model is first extended to its regularized version, i.e. regularized fuzzy linear regression (RFLR) model, so as to enhance its generalization capability; then RFLR model is explained 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. Our experimental results also confirm this theoretical claim. We believe that this conclusion provides an important reference for us to determine h in FLR model with noisy input.

Keywords—Fuzzy linear regression model; MAP; Optimal parameter choice

I. INTRODUCTION

Fuzzy linear regression model plays a pivotal role in fuzzy modeling. However, how to choose the parameter h for FLR model with noisy input still keeps an open problem. A bad choice for threshold h will heavily deteriorate the performances of these models. In recent years, how to choose the loss function and the corresponding parameters for support vector regression (SVR) machine with noisy input has been studied well. J.B.Gao and S.R.Gunn pointed out that SVR problem could be transformed into the equivalent MAP problem [1]. Based on this idea, J.T.Kwok derived the linear dependency between ε and the input noise in ε -SVR [2], Shitong Wang et al. investigated the theoretically optimal parameter choices for Huber-SVR and r-SVR [3], and Hongwei Ge et al. investigated the optimal parameter choices for FLR model with Gaussian noisy input [4,5]. Although Gaussian noise is typically adopted in most robustness analyses, there are still other types of noise such as Laplacian distribution noise and uniform noise in real datasets. So, in this paper, particular attention is paid on the dependency relationships between free threshold h and the standard deviations of Laplacian and uniform noisy input. We first extend the traditional FLR model to its regularized version, i.e. regularized fuzzy linear regression (RFLR) model, so as to enhance its generalization capability; and then explain this regularized model by using MAP framework [1-3]; and finally, with the help of this MAP framework, study the dependency between h and the standard deviation of input

noise.

The rest of this paper is organized as follows. In section 2, we will introduce the FLR model and the corresponding RFLR model and show the equivalent relationship between RFLR model and MAP. The analysis of the inverse linear dependency between h and the standard deviation of Laplacian noisy input and Uniform noisy input are theoretically derived in section 3, respectively, while section 4 gives the experimental results and last section concludes this paper.

II. FUZZY LINER REGRESSION MODEL AND MAP

A. Fuzzy liner regression model

The FLR model can be summarized as follows. Consider the function:

$$Y = f(\mathbf{x}, \mathbf{A}) = A_0 + A_1x_1 + A_2x_2 + \cdots + A_nx_n, \quad (1)$$

where $\mathbf{x} = (1, x_1, x_2, \dots, x_n)^T$ is a vector of non-fuzzy inputs, and $\mathbf{A} = (A_0, A_1, \dots, A_n)^T$ is a vector of fuzzy model parameters. Parameters $A_i (0 \leq i \leq n)$ are symmetric triangular fuzzy coefficients, and they can be described by $\{\xi_i, \eta_i\}$, where ξ_i is the center and η_i is the spread of A_i .

In vector notation, the fuzzy parameters $\mathbf{A}$ can be written as $\mathbf{A} = (\xi, \eta)$ where $\xi = (\xi_0, \xi_1, \xi_2, \dots, \xi_n)^T$ and $\eta = (\eta_0, \eta_1, \eta_2, \dots, \eta_n)^T$. Following the principle of extension, the fuzzy membership function for the output can be obtained by

$$\mu_Y(y) = \begin{cases} 1 - \frac{|y - \xi^T \mathbf{x}|}{\eta^T |\mathbf{x}|}, & \mathbf{x} \neq 0, \\ 1, & \mathbf{x} = 0, y = 0, \\ 0, & \mathbf{x} = 0, y \neq 0, \end{cases} \quad (2)$$

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

Assume we have a dataset with n -dimensional non-fuzzy input $\mathbf{x}$ and one-dimensional non-fuzzy output variable y :

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}, \\ \mathbf{x}_i \in R^{n+1}, y_i \in R, i = 1, 2, \dots, N. \quad (3)$$

We can establish the following FLR model $Y_i^* = \mathbf{A}^{*T} \mathbf{x}_i$, where $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{in})^T$, $\mathbf{A} = (A_0, A_1, A_2, \dots, A_n)^T$, $\mathbf{A}^* = (A_0^*, A_1^*, A_2^*, \dots, A_n^*)^T$, $i = 1, 2, \dots, N$, and $Y_i^*, \mathbf{A}^*$ denote the

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

estimates of Y_i , $\mathbf{A}$ respectively. The objective is to minimize the fuzziness of the fuzzy linear model and the following linear programming problem can be formulated [4,5]:

$$\min \Phi(\boldsymbol{\eta}, \boldsymbol{\xi}, \Delta) = \sum_i \boldsymbol{\eta}^T | \mathbf{x}_i | + C \sum_i (\Delta_i^- + \Delta_i^+), \quad (4)$$

$$s.t. \begin{cases} \frac{y_i - \boldsymbol{\xi}^T \mathbf{x}_i}{\boldsymbol{\eta}^T | \mathbf{x}_i |} \leq (1-h) + \Delta_i^-, \\ \frac{\boldsymbol{\xi}^T \mathbf{x}_i - y_i}{\boldsymbol{\eta}^T | \mathbf{x}_i |} \leq (1-h) + \Delta_i^+, \\ \Delta_i^-, \Delta_i^+ \geq 0, \\ i = 1, 2, \dots, N, \end{cases}$$

where C is a predefined constant, Δ_i^- and Δ_i^+ denote the latent variables of the upper/lower bounds of output respectively, h ($h \in [0,1)$) is the threshold of fitness.

In order to enhance the generalization capability of FLR model and to be easy to analyze, just like SVR techniques in [3, 7, 8], a regularized term $\frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\xi}$ can be introduced into Eq. (4), i.e.

$$\min \Phi(\boldsymbol{\eta}, \boldsymbol{\xi}, \Delta) = \sum_i \boldsymbol{\eta}^T | \mathbf{x}_i | + \frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\xi} + C \sum_i (\Delta_i^- + \Delta_i^+). \quad (5)$$

Equation (5) formulates the RFLR model. In the rest of this paper, how to choose the parameter h for RFLR model with Laplacian noisy input and Uniform noisy input will be studied. Obviously, the conclusion will also be very helpful for the practical applications of FLR model.

B. The regression model and MAP

In this subsection, in terms of the evidence theory [1,2], we will demonstrate the RFLR model is equivalent to *maximum a posteriori* MAP estimation, based on maximum likelihood estimation.

Now, the focus is on obtaining a weight vector $\tilde{\boldsymbol{\xi}}$ in the generalized linear regression model such as

$$y_i = \tilde{\boldsymbol{\xi}}^T \mathbf{x}_i + n_i, i = 1, 2, \dots, N, \quad (6)$$

for all the data in the dataset (3), where all the data $\mathbf{x}_i$ follow distribution $p(\cdot)$ and all n_i are i.i.d noise following some distribution $\phi(\cdot)$. Thus, the corresponding density function on y can be denoted as $p(y | \mathbf{x}) = \phi(y - \tilde{\boldsymbol{\xi}}^T \mathbf{x})$. The degree of such an approximation can be measured by a loss function $L(\tilde{\boldsymbol{\xi}}^T \mathbf{x}, \boldsymbol{\eta}^T \mathbf{x}, y)$, i.e.

$$L(\tilde{\boldsymbol{\xi}}^T \mathbf{x}, \boldsymbol{\eta}^T \mathbf{x}, y) = \begin{cases} 0, & \frac{|\tilde{\boldsymbol{\xi}}^T \mathbf{x} - y|}{\boldsymbol{\eta}^T | \mathbf{x} |} \leq 1-h, \\ \frac{|\tilde{\boldsymbol{\xi}}^T \mathbf{x} - y|}{\boldsymbol{\eta}^T | \mathbf{x} |} - 1 + h, & \text{otherwise,} \end{cases} \quad (7)$$

Assume the loss function $L(\tilde{\boldsymbol{\xi}}^T \mathbf{x}, \boldsymbol{\eta}^T \mathbf{x}, y)$ leads to the following Gaussian probability density function on y :

$$p(y_i | \mathbf{x}_i, \tilde{\boldsymbol{\xi}}, \boldsymbol{\eta}, \beta, h) = \frac{1}{C(\beta, h)} \exp[-\beta L(\tilde{\boldsymbol{\xi}}^T \mathbf{x}_i, \boldsymbol{\eta}^T \mathbf{x}_i, y_i)] \quad (8)$$

where $C(\beta, h) = \iint_D \exp[-\beta L(\tilde{\boldsymbol{\xi}}^T \mathbf{x}, \boldsymbol{\eta}^T \mathbf{x}, y)] d\mathbf{x} dy$, and β is a

parameter.

Just as the mathematical analysis in [1,2] does, we adopt the Gaussian prior in the following analytical framework. With the Gaussian prior on $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, i.e.

$$p(\boldsymbol{\eta} | \alpha, \mathbf{x}_i) = \frac{1}{M(\alpha)} \exp(-\alpha \boldsymbol{\eta}^T | \mathbf{x}_i |), \quad (9)$$

where $M(\alpha) = \int \exp(-\alpha \boldsymbol{\eta}^T | \mathbf{x}_i |) d\boldsymbol{\eta}$,

$$p(\boldsymbol{\xi} | \gamma) = \frac{1}{F(\gamma)} \exp\left(-\frac{\gamma}{2} \boldsymbol{\xi}^T \boldsymbol{\xi}\right), \quad (10)$$

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

and by applying the Bayes rule:

$$p(\boldsymbol{\eta}, \boldsymbol{\xi} | D, \beta, h) \propto p(D | \boldsymbol{\eta}, \boldsymbol{\xi}, \beta, h) p(\boldsymbol{\xi} | \gamma) \prod_{i=1}^N p(\boldsymbol{\eta} | \alpha, \mathbf{x}_i) p(\mathbf{x}_i), \quad (11)$$

we have

$$\ln p(\boldsymbol{\eta}, \boldsymbol{\xi} | D, \beta, h) = -\alpha \sum_{i=1}^N \boldsymbol{\eta}^T | \mathbf{x}_i | - \frac{\gamma}{2} \boldsymbol{\xi}^T \boldsymbol{\xi} - \beta \sum_{i=1}^N L(\boldsymbol{\eta}^T \mathbf{x}_i, \boldsymbol{\xi}^T \mathbf{x}_i, y_i) - N \ln C(\beta, h) + \text{const}. \quad (12)$$

With setting $\gamma = \alpha$ and $C = \beta/\alpha$, optimizing (5) can be interpreted as finding for the MAP estimate of $\boldsymbol{\eta}, \boldsymbol{\xi}$ at given values of β, h . That is to say, the RFLR model in (5) is equivalent to *maximum a posteriori* MAP estimation.

In order to make the analysis easier, $\frac{1}{N} \sum_{i=1}^N L(\boldsymbol{\eta}^T \mathbf{x}_i, \boldsymbol{\xi}^T \mathbf{x}_i, y_i)$

in (12) is replaced by its expectation:

$$\begin{aligned} E(L(\boldsymbol{\eta}^T \mathbf{x}, \boldsymbol{\xi}^T \mathbf{x}, y)) &= \iint_D L(\boldsymbol{\eta}^T \mathbf{x}, \boldsymbol{\xi}^T \mathbf{x}, y) p(y | \mathbf{x}) p(\mathbf{x}) d\mathbf{x} dy \\ &= \int_D \left[\int_{-\infty}^{\boldsymbol{\xi}^T \mathbf{x} - (1-h)\boldsymbol{\eta}^T | \mathbf{x} |} \left(\frac{\boldsymbol{\xi}^T \mathbf{x} - y}{\boldsymbol{\eta}^T | \mathbf{x} |} - 1 + h \right) p(y | \mathbf{x}) dy \right. \\ &\quad \left. + \int_{\boldsymbol{\xi}^T \mathbf{x} + (1-h)\boldsymbol{\eta}^T | \mathbf{x} |}^{+\infty} \left(\frac{y - \boldsymbol{\xi}^T \mathbf{x}}{\boldsymbol{\eta}^T | \mathbf{x} |} - 1 + h \right) p(y | \mathbf{x}) dy \right] p(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (13)$$

and $\frac{1}{N} \sum_{i=1}^N \boldsymbol{\eta}^T | \mathbf{x}_i |$ in (12) can also be approximated by

$$E(\boldsymbol{\eta}^T | \mathbf{x}) = \int_D \boldsymbol{\eta}^T | \mathbf{x} | p(\mathbf{x}) d\mathbf{x} \quad (14)$$

Thus, Eq.(12) becomes

$$\begin{aligned} \ln p(\boldsymbol{\eta}, \boldsymbol{\xi} | D, \beta, h) &= -\alpha N * E(\boldsymbol{\eta}^T | \mathbf{x}) - \frac{\gamma}{2} \boldsymbol{\xi}^T \boldsymbol{\xi} \\ &\quad - \beta N * E(L(\boldsymbol{\eta}^T \mathbf{x}, \boldsymbol{\xi}^T \mathbf{x}, y)) - N \ln C(\beta, h) + \text{const}. \end{aligned} \quad (15)$$

In order to maximize (15), its derivatives with respect to $\boldsymbol{\eta}, \boldsymbol{\xi}, \beta, h$ must be zero. Thus, we can obtain

$$E(L(\hat{\boldsymbol{\eta}}^T \mathbf{x}, \hat{\boldsymbol{\xi}}^T \mathbf{x}, y)) = -\frac{\partial C(\beta, h) / \partial \beta}{C(\beta, h)}, \quad (16)$$

$$\int_D \left[\int_{-\infty}^{\xi^T \mathbf{x} - (1-h)\hat{\eta}^T \mathbf{x}} p(y | \mathbf{x}) dy + \int_{\xi^T \mathbf{x} + (1-h)\hat{\eta}^T \mathbf{x}}^{+\infty} p(y | \mathbf{x}) dy \right] p(\mathbf{x}) d\mathbf{x} = \frac{-\partial C(\beta, h)/\partial h}{\beta C(\beta, h)}. \quad (17)$$

When $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}, \xi = \hat{\xi}$, maximizing (15) actually becomes the following optimization problem:

$$\arg \min_{\alpha, \beta, h} \alpha E(\hat{\boldsymbol{\eta}}^T | \mathbf{x}) + \beta E(L(\hat{\boldsymbol{\eta}}^T \mathbf{x}, \hat{\xi}^T \mathbf{x}, y)) + \ln C(\beta, h). \quad (18)$$

Please note $C(\beta, h)$ in (18). According to its corresponding MAP, we have

$$C(\beta, h) = 2 \left(\int_0^{1-h} \exp(0) dt + \int_{1-h}^{+\infty} \exp(-\beta(t-1+h)) dt \right) = \frac{2(1+(1-h)\beta)}{\beta}. \quad (19)$$

By substituting (19) into (16) and (17), we obtain

$$E(L(\hat{\boldsymbol{\eta}}^T \mathbf{x}, \hat{\xi}^T \mathbf{x}, y)) = \frac{1}{\beta(1+(1-h)\beta)}, \quad (20)$$

$$\int_D \left[\int_{-\infty}^{\xi^T \mathbf{x} - (1-h)\hat{\eta}^T \mathbf{x}} p(y | \mathbf{x}) dy + \int_{\xi^T \mathbf{x} + (1-h)\hat{\eta}^T \mathbf{x}}^{+\infty} p(y | \mathbf{x}) dy \right] p(\mathbf{x}) d\mathbf{x} = \frac{1}{1+(1-h)\beta}. \quad (21)$$

III. THE DEPENDENCY BETWEEN h AND THE INPUT NOISE

A. Laplacian Noise

For the Laplacian noise model,

$$p(y | \mathbf{x}) = \frac{1}{2\mu} \exp \left[-\frac{|y - \xi^T \mathbf{x}|}{\mu} \right], \quad (22)$$

Here, we only consider the simpler case when $\mathbf{x}$ is one-dimensional, with uniform density over the range $[-L, L]$ (i.e. $X \sim U([-L, L])$). When $\mathbf{x}$ 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 (21) reduce to

$$\begin{aligned} \frac{1}{1+\varepsilon\beta} &= \int_D \left[\int_{-\infty}^{\xi x - \varepsilon\hat{\eta}|x|} \frac{1}{2\mu} \exp \left(-\frac{|y - \xi x|}{\mu} \right) dy + \int_{\xi x + \varepsilon\hat{\eta}|x|}^{+\infty} \frac{1}{2\mu} \exp \left(-\frac{|y - \xi x|}{\mu} \right) dy \right] p(x) dx \\ &= 2 \int_{-L}^0 \left(\int_{\xi x + \varepsilon\hat{\eta}|x|}^{+\infty} \frac{1}{2\mu} \exp \left(-\frac{y - \xi x}{\mu} \right) dy \right) \frac{1}{2L} dx \\ &\quad + 2 \int_0^L \left(\int_{\xi x}^{+\infty} \frac{1}{2\mu} \exp \left(-\frac{y - \xi x}{\mu} \right) dy \right) \frac{1}{2L} dx \\ &\quad + 2 \int_{-L}^0 \left(\int_{\xi x}^{\xi x - \varepsilon\hat{\eta}|x|} \frac{1}{2\mu} \exp \left(-\frac{y - \xi x}{\mu} \right) dy \right) \frac{1}{2L} dx \end{aligned}$$

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

(23)

Similarly, from (13), (20) and (21), we obtain

$$\begin{aligned} \frac{1}{\beta} &= \frac{(\xi - \hat{\xi})L}{2E} + \frac{\mu^2}{LE} \cdot \frac{(\xi - \hat{\xi})^3 - 3(\xi - \hat{\xi})\varepsilon^2\hat{\eta}^2}{((\xi - \hat{\xi})^2 - \varepsilon^2\hat{\eta}^2)^2} \\ &\quad + \frac{\mu}{2E} \exp \left(-\frac{(\xi - \hat{\xi})L}{\mu} \right) \\ &\quad \left(\frac{(\xi - \hat{\xi} - \varepsilon\hat{\eta})\varepsilon\hat{\eta}L - (\xi - \hat{\xi})\mu + 2\mu\varepsilon\hat{\eta}}{(\xi - \hat{\xi} - \varepsilon\hat{\eta})^2 L} \exp \left(\frac{\varepsilon\hat{\eta}L}{\mu} \right) \right. \\ &\quad \left. - \frac{(\xi - \hat{\xi} + \varepsilon\hat{\eta})\varepsilon\hat{\eta}L + (\xi - \hat{\xi})\mu + 2\mu\varepsilon\hat{\eta}}{(\xi - \hat{\xi} + \varepsilon\hat{\eta})^2 L} \exp \left(-\frac{\varepsilon\hat{\eta}L}{\mu} \right) \right). \quad (24) \end{aligned}$$

By using (20), (23) and (24), we have

$$\begin{aligned} E(L(\hat{\eta}x, \hat{\xi}x, y)) &= \frac{(\xi - \hat{\xi})L - 2\varepsilon E}{2E} \\ &\quad + \frac{\mu}{LE} \cdot \frac{(\xi - \hat{\xi})^3 \mu - (\xi - \hat{\xi})^2 \varepsilon^2 \hat{\eta} E - 3(\xi - \hat{\xi}) \varepsilon^2 \hat{\eta}^2 \mu + \varepsilon^4 \hat{\eta}^3 E}{((\xi - \hat{\xi})^2 - \varepsilon^2 \hat{\eta}^2)^2} \\ &\quad + \frac{\mu}{2E} \exp \left(-\frac{(\xi - \hat{\xi})L}{\mu} \right) \\ &\quad \left(\frac{(\xi - \hat{\xi} - \varepsilon\hat{\eta})\varepsilon\hat{\eta}L - (\xi - \hat{\xi})\mu + 2\mu\varepsilon\hat{\eta} + (\xi - \hat{\xi})\varepsilon E - \varepsilon^2 \hat{\eta} E}{(\xi - \hat{\xi} - \varepsilon\hat{\eta})^2 L} \exp \left(\frac{\varepsilon\hat{\eta}L}{\mu} \right) \right. \\ &\quad \left. - \frac{(\xi - \hat{\xi} + \varepsilon\hat{\eta})\varepsilon\hat{\eta}L + (\xi - \hat{\xi})\mu + 2\mu\varepsilon\hat{\eta} + (\xi - \hat{\xi})\varepsilon E + \varepsilon^2 \hat{\eta} E}{(\xi - \hat{\xi} + \varepsilon\hat{\eta})^2 L} \exp \left(-\frac{\varepsilon\hat{\eta}L}{\mu} \right) \right). \quad (25) \end{aligned}$$

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

$$\alpha E(\hat{\eta} | x) = \beta E C^{-1} \quad (26)$$

and

$$\varepsilon \hat{\eta} L = 2\varepsilon E \quad (27)$$

Define $y - \xi x \approx \xi x - \xi x = \delta(x)$. It can be shown that

$$E_X(\delta^2(x)) \approx E_X((y - \xi x)^2) = 2\mu^2. \quad (28)$$

With $X \sim U([-L, L])$, we also obtain $\text{var}(X) = L^2/3$.

Consequently

$$E_X(\delta^2(x)) \approx \text{var}((\xi - \hat{\xi})X) \approx (\xi - \hat{\xi})^2 L^2/3, \quad (29)$$

and hence

$$\mu = (\xi - \hat{\xi})L/\sqrt{6}. \quad (30)$$

After substituting (19), (23), (24), (25), (26), (27) and (30) into (18), minimizing (18) becomes minimizing

$$\begin{aligned} g\left(\frac{\varepsilon}{\mu}\right) &= \left(\frac{\sqrt{6}}{2E} + f_1\left(\frac{\varepsilon}{\mu}\right) + \frac{\exp(-\sqrt{6})}{4E} f_2\left(\frac{\varepsilon}{\mu}\right) \right)^{-1} \\ &\quad \left(\frac{\sqrt{6} - 2\frac{\varepsilon E}{\mu}}{2E} + \frac{3\sqrt{6} - (6 + 6\sqrt{6})\frac{\varepsilon^2 E^2}{\mu^2} + 4\frac{\varepsilon^4 E^4}{\mu^4}}{18E - 24\frac{\varepsilon^2 E^3}{\mu^2} + 8\frac{\varepsilon^4 E^5}{\mu^4}} \right. \\ &\quad \left. + \frac{\exp(-\sqrt{6})}{4E} f_3\left(\frac{\varepsilon}{\mu}\right) + \frac{E}{\mu C} \right) \end{aligned}$$

$$\begin{aligned}
& -\ln \left(1 + \frac{\frac{\varepsilon E}{\mu}}{3 - 2\frac{\varepsilon^2 E^3}{\mu^2}} + \frac{\exp(-\sqrt{6})}{2} \left(\frac{\exp\left(-2\frac{\varepsilon E}{\mu}\right)}{\sqrt{6} + 2\frac{\varepsilon E}{\mu}} - \frac{\exp\left(2\frac{\varepsilon E}{\mu}\right)}{\sqrt{6} - 2\frac{\varepsilon E}{\mu}} \right) \right) \\
& + \ln \left(\frac{\sqrt{6}}{2E} + f_1\left(\frac{\varepsilon}{\mu}\right) + \frac{\exp(-\sqrt{6})}{4E} f_2\left(\frac{\varepsilon}{\mu}\right) \right) + \text{const}, \quad (31)
\end{aligned}$$

where

$$\begin{aligned}
f_1\left(\frac{\varepsilon}{\sigma}\right) &= \frac{3\sqrt{6} - 6\sqrt{6} \frac{\varepsilon^2 E^2}{\mu^2}}{18E - 24 \frac{\varepsilon^2 E^3}{\mu^2} + 8 \frac{\varepsilon^4 E^5}{\mu^4}}, \\
f_2\left(\frac{\varepsilon}{\sigma}\right) &= \frac{-\sqrt{6} + (4 + 2\sqrt{6}) \frac{\varepsilon E}{\mu} - 4 \frac{\varepsilon^2 E^2}{\mu^2}}{3 - 2\sqrt{6} \frac{\varepsilon E}{\mu} + 2 \frac{\varepsilon^2 E^2}{\mu^2}} \exp\left(2 \frac{\varepsilon E}{\mu}\right) \\
& - \frac{\sqrt{6} + (4 + 2\sqrt{6}) \frac{\varepsilon E}{\mu} + 4 \frac{\varepsilon^2 E^2}{\mu^2}}{3 + 2\sqrt{6} \frac{\varepsilon E}{\mu} + 2 \frac{\varepsilon^2 E^2}{\mu^2}} \exp\left(-2 \frac{\varepsilon E}{\mu}\right) \\
f_3\left(\frac{\varepsilon}{\sigma}\right) &= \frac{-\sqrt{6} + (4 + 3\sqrt{6}) \frac{\varepsilon E}{\mu} - 6 \frac{\varepsilon^2 E^2}{\mu^2}}{3 - 2\sqrt{6} \frac{\varepsilon E}{\mu} + 2 \frac{\varepsilon^2 E^2}{\mu^2}} \exp\left(2 \frac{\varepsilon E}{\mu}\right) \\
& - \frac{\sqrt{6} + (4 + 3\sqrt{6}) \frac{\varepsilon E}{\mu} + 6 \frac{\varepsilon^2 E^2}{\mu^2}}{3 + 2\sqrt{6} \frac{\varepsilon E}{\mu} + 2 \frac{\varepsilon^2 E^2}{\mu^2}} \exp\left(-2 \frac{\varepsilon E}{\mu}\right)
\end{aligned}$$

Obviously, when $\left(\frac{\varepsilon}{\mu}\right)$ takes some fixed value, $g\left(\frac{\varepsilon}{\mu}\right)$ will achieve its minimum, which indicates that there is a linear dependency between ε and μ . Thus, there is an inverse linear dependency between the parameter h and the standard variance σ of the Laplacian input noise because of $\varepsilon = 1 - h$ and $\sigma = \sqrt{2}\mu$.

B. Uniform Noise

Next, we consider the uniform noise model

$$\phi(n) = \frac{1}{2\mu}, \quad n \in [-\mu, \mu] \quad (32)$$

Again, we only consider the simpler case when $\mathbf{x}$ 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. (21) becomes

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

$$= \frac{(\hat{\xi} - \hat{\xi})^2 L^2 - \varepsilon^2 \hat{\eta}^2 L^2 + 2\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})L + \mu^2}{4\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})L}. \quad (33)$$

Using (13), (20), (21) and (33), after a little tedious computation, it can be shown that

$$\begin{aligned}
\frac{1}{\beta} &= \frac{(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})(\hat{\xi} - \hat{\xi} - \varepsilon\hat{\eta})^2 L^3 + 3((\hat{\xi} - \hat{\xi})^2 - \varepsilon^2 \hat{\eta}^2) L^2 (\mu + \varepsilon E)}{12\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})LE} \\
&+ \frac{3\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})L(\mu + 2\varepsilon E) + 3\mu^2 \varepsilon E + \mu^3}{12\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})LE}, \quad (34)
\end{aligned}$$

$$\begin{aligned}
E(L(\hat{\eta}x, \hat{\xi}x, y)) &= \frac{(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})(\hat{\xi} - \hat{\xi} - \varepsilon\hat{\eta})^2 L^3 + 3\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})(\hat{\xi} - \hat{\xi} - \varepsilon\hat{\eta})L^2}{12\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})LE} \\
&+ \frac{3\mu^2(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})L + \mu^3}{12\mu(\hat{\xi} - \hat{\xi} + \varepsilon\hat{\eta})LE}. \quad (35)
\end{aligned}$$

As in Section 3.A, consider $\delta(x) = (\hat{\xi} - \hat{\xi})x$ and assume that

$$E_x(\delta^2(x)) \approx E_x((y - \hat{\xi}x)^2) = \mu^2/3 \quad (36)$$

Using (29) again, we obtain

$$\mu = (\hat{\xi} - \hat{\xi})L \quad (37)$$

Analogous to Section 3.A, by plugging in (19), (33), (34), (35), (26), (27) and (37), it can be shown that the problem reduces to minimizing

$$\begin{aligned}
g\left(\frac{\varepsilon}{\mu}\right) &= \left(\frac{3E + 6\frac{\varepsilon E^2}{\mu}}{2 + 4\frac{\varepsilon E}{\mu} - \frac{\varepsilon^2 E^2}{\mu^2} - \frac{\varepsilon^3 E^3}{\mu^3}} \right) \left(\frac{2 + \frac{\varepsilon E}{\mu} - 4\frac{\varepsilon^2 E^2}{\mu^2} + 2\frac{\varepsilon^3 E^3}{\mu^3}}{3E + 6\frac{\varepsilon E^2}{\mu}} + \frac{E}{\mu C} \right) \\
&- \ln \left(\frac{1 + \frac{\varepsilon E}{\mu} - \frac{\varepsilon^2 E^2}{\mu^2}}{1 + 2\frac{\varepsilon E}{\mu}} \right) + \ln \left(\frac{2 + 4\frac{\varepsilon E}{\mu} - \frac{\varepsilon^2 E^2}{\mu^2} - \frac{\varepsilon^3 E^3}{\mu^3}}{3E + 6\frac{\varepsilon E^2}{\mu}} \right) + \text{const}. \quad (38)
\end{aligned}$$

Again, we can see from (38) that there is a linear dependency between ε and μ . Obviously, because the standard variance σ 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 σ .

C. Dependency Between h and the Input Noise

From Section 3.A and 3.B, we come to the conclusion that there are approximately inverse proportional dependency relationships between the parameter h and the standard variance σ of Laplacian input noise and uniform input noise. At the same time, in [4,5], we have derived the analogical conclusion with Gaussian noisy input. Following the same spirit in [2], 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 RFLR model and the input noise, the conclusion is very helpful for the practical applications of both FLR and RFLR model, for FLR model is just a special case of RFLR model.

IV. EXPERIMENTAL STUDIES

In this experiment, we apply RFLR model to a real dataset taken from [6] (see Table1). In [6], fuzzy linear regression was used 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/s).

By using the same method from [6], 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 $(\xi_i, \eta_i), i=1,2,\dots,21$ can be obtained, where ξ_i denotes the center and η_i denotes the spread. Next, in order to investigate the dependency relationship between h and the noisy input, let $y'_i = y_i + k \cdot n$, $i=1,2,\dots,21$, where k is a noise-signal ratio and n represents the Laplacian noise. Then the corresponding sampling dataset $(x_i, y'_i), i=1,2,\dots,21$ can be generated. Similarly, its corresponding regression values $(\xi'_i, \eta'_i), i=1,2,\dots,21$ can be obtained by using the same RFLR model. 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.

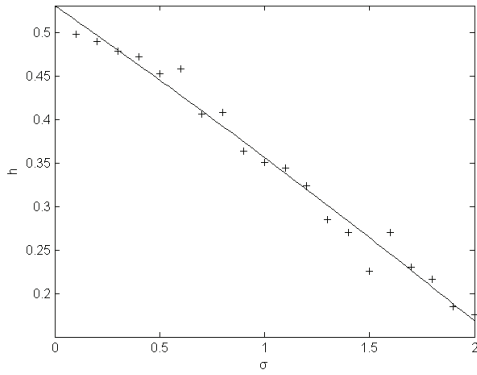


Fig. 1. With Laplacian noisy input, the relationship between h and σ when $k=0.05$

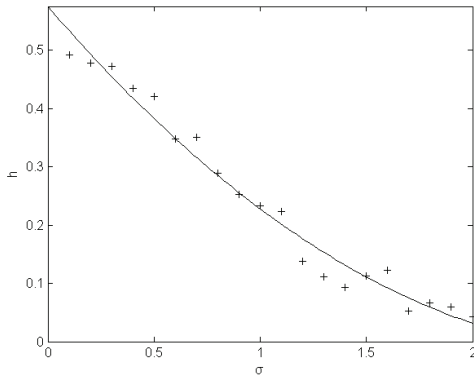


Fig. 2. With Laplacian noisy input, the relationship between h and σ when $k=0.1$

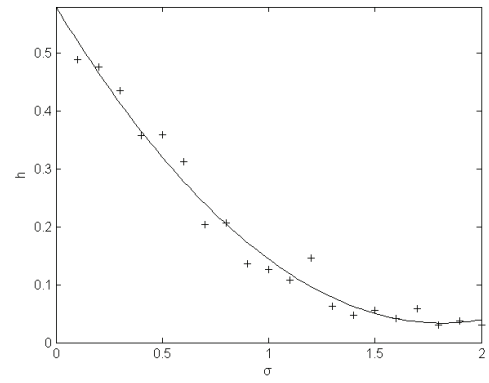


Fig. 3. With Laplacian noisy input, the relationship between h and σ when $k=0.2$

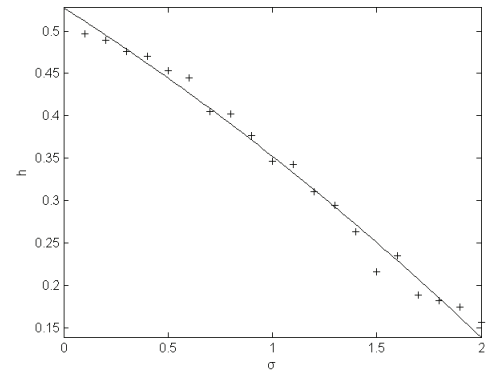


Fig. 4. With uniform noisy input, the relationship between h and σ when $k=0.05$

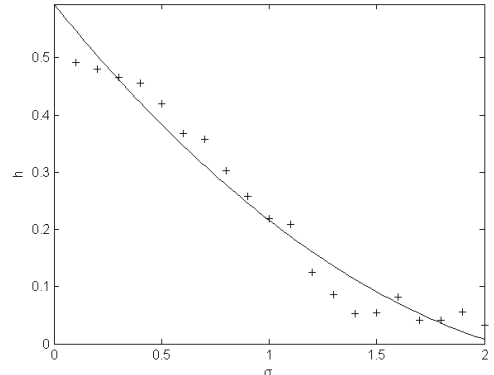


Fig. 5. With uniform noisy input, the relationship between h and σ when $k=0.1$

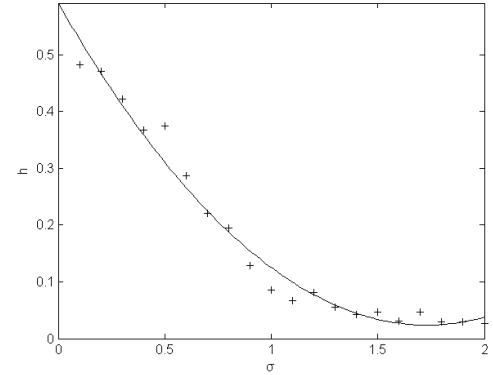


Fig. 6. With uniform noisy input, the relationship between h and σ when $k=0.2$

Fig.1, 2 and 3 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, Figures 4-6 show the experimental results, which again validate our theoretical conclusion.

It can be easily seen from Fig.1-6 that when noise is small, i.e. k and σ is comparatively small, 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, RFLR model may become ineffective for seriously distorted datasets. Of course, although we have arranged many experiments and all these experimental results validate the above obtained conclusion on RFLR model, we just show several experimental results here so as to save the paper's space.

V. CONCLUSIONS AND FUTURE WORK

In this paper, based on the Bayesian frame, the approximately inversely linear dependency relationships between the parameter h and the standard variance σ of Laplacian input noise and uniform input noise are derived. And 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 [4, 5], 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 [7]. In the future, we will investigate better methods to solve the problem.

REFERENCES

- [1] J.B.Gao, S.R.Gunn and C.J. Ham's, "A probabilistic framework for SVM regression and Error Bar Estimation," *Machine Learning*, vol. 46, pp.71-89, 2002
- [2] J. T. Kwok and Ivor W. Tsang, "Linear dependency between $\mathcal{E}$ and the input noise in $\mathcal{E}$ -support vector regression," *IEEE Transaction on Neural Networks*, vol.14, no.3, pp.544-553, 2003
- [3] Wang shitong, "Theoretically optimal parameter choices for support vector regression machines with noisy input," *Int. J. Soft Computing*, vol.9, no.10, pp.732-741, 2005
- [4] Hongwei Ge and Shitong Wang, "Dependency between degree of fit and input noise in fuzzy linear regression using non-symmetric fuzzy triangular coefficients," *Fuzzy Sets and Systems*, vol.158, no.19, pp.2189-2202, 2007
- [5] Hongwei Ge, Fu-lai Chung and Shitong Wang, "Theoretical Choice of the Optimal Threshold for Possibilistic Linear Model with Noisy Input," *IEEE Transactions on Fuzzy Systems*, vol.16, no.4, pp.1027-1037, 2008
- [6] Renato Coppi and Pierpaolo D'Urso, "Least squares estimation of a linear regression model with LR fuzzy response," *Computational Statistics & Data Analysis*, vol.51, pp.267-286, 2006

- [7] M.H.Law and J.T.Kwok, "Bayesian support vector regression," in *Proc English Int.Workshop Artif.Intell.Statist.*,2001, pp.239-244
- [8] D. H. Hong and C. Hwang, "Support vector fuzzy regression machines," *Fuzzy Sets and Syst.*, vol.138, pp.271-281, 2003
- [9] Mehran Hojati and C.R.Bector, "A simple method for computation of fuzzy linear regression," *European Journal of Operational Research*, vol.166, pp.172-184, 2005

TABLE I
INPUT-OUTPUT DATA IN [8]

Number	Relative Humidity (RH)(%)	Rain R(mm)	Wind Speed WS(m/s)	Concentrations of carbon monoxide (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	0.90	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