퍼지 및 지능시스템학회 논문지 2006, Vol. 16, No. 4, pp. 499-505

Fuzzy regression using regularlization method based on Tanaka's model

Dug Hun Hong* **and Kyung Tae Kim****

*Department of Mathematics, Myongji University, Yongin, Kyonggi, Korea
**Department of Electronics and Electrical Information Engineering, Kyungwon University, Sungnam Kyunggido, South Korea

Abstract

Regularlization approach to regression can be easily found in Statistics and Information Science literature. The technique of regularlization was introduced as a way of controlling the smoothness properties of regression function. In this paper, we have presented a new method to evaluate linear and non-linear fuzzy regression model based on Tanaka's model using the idea of regularlization technique. Especially this method is a very attractive approach to model non-linear fuzzy data.

Key words : Fuzzy inference systems, Fuzzy regression, Regularlization methods.

1. Introduction

Fuzzy linear regression provides means for tackling regression problems lacking a significance amount of data for determining regression models and with vague relationships between the dependent variables.

The concept of fuzzy regression analysis was introduced by Tanaka et al. in 1982[17], where an LP based method with symmetric triangular fuzzy parameters was proposed.

The method is recommend for practical situations where decisions often have to be made on the basis of imprecise and partially available data where human estimation is influential. This first attempt of applying fuzzy regression was done using non-fuzzy input experimental data. An extension of the idea was reported by Tanaka et al.[16] comparing the capability to process fuzzy input experimental data. Heshmaty and Kandel[8] applied this method to forecasting in uncertain environment and Watada[20] applied the idea of fuzzy regression to fuzzy time-series. Fuzzy data analysis, regarded as a non-statistical procedure for possibilistic systems, was reported by Tanaka[16], and the Tanaka et al.[18]. Fuzzy regression has been also investigated from the viewpoint of least square regression. Celminš[4,5] and Diamond[7] developed several models for fuzzy least squares fitting. A collection of recent papers dealing with several approach to fuzzy regression analysis can be found in [13].

In contrast to fuzzy linear regression, there have been only a few articles on fuzzy nonlinear regression. What researchers in fuzzy nonlinear regression were concerned with was data of the form with crisp inputs and fuzzy output. However, some papers, for example [2,3,6], were concerned with the data set with fuzzy inputs and fuzzy output. By the way, in this paper we will treat fuzzy nonlinear regression for data of the form with crisp inputs and fuzzy output.

In this paper, we will present a new method to evaluate fuzzy regression model based on Tanaka's model using the idea of regularlization method.

Utilizing regularlization method, we can extend Tanaka's model to non-linear case easily. Regularization techniques [10,11,12,14,15,19] have been extensively studied in the context of crisp linear regression models. The technique of regularlization encourages smoother regression function. One of the simplest forms of regularlizer is called weight decay. In fact, in this paper we use weight decay as regularlizer. This approach to regression is also known as ridge regression.

The main difference between our regularlization methods approach and the nonlinear approaches by Buckley et al. [2,3] and Celmins [6] is not crisp input-fuzzy output versus fuzzy input-fuzzy output, but model-free versus model-dependent.

The rest of this paper is organized as follows. In Section 2, we modify Tanaka's fuzzy linear regression model by utilizing regularlization technique. Section 3 provides details regarding how to extend to non-linear fuzzy regression model. In Section 4, we consider fuzzy linear regression model with fuzzy input-output data but real coefficients. Section 5 gives some conclusions.

2. Fuzzy linear regression models

First we need to briefly look at how to get solutions

접수일자 : 2006년 3월 18일
완료일자 : 2006년 8월 3일

for crisp multiple linear regression models using regularlization method. See for details [12,15,19]. Suppose we are given training data $\{(\mathbf{x}_1,y_1),\cdots,(\mathbf{x}_l,y_l)\}\subset X\times R$, where X denotes the space of the input patterns, for example R^d. For pedagogical reasons, we begin by describing the case of multiple crisp linear regression functions f, taking the form

$$f(\mathbf{x})=\langle\mathbf{w},\mathbf{x}\rangle+b \text{ with } \mathbf{x}\in X, b\in R \quad (1)$$

where $\langle\cdot,\cdot\rangle$ denotes the dot product in X. Flatness in the case of (1) means that one seeks small $\mathbf{w}$. One way to ensure this is to minimize the Euclidean norm $\|\mathbf{w}\|^2$. Hence we can write this problem as the convex optimization problem as follows.

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{l}\xi_i^2 \quad (2)$$
$$\text{subject to } y_i-\langle\mathbf{w},\mathbf{x}_i\rangle-b=\xi_i.$$

The parameter $C\geq 0$ controls the smoothness and degree of fit. To solve this convex optimization problem, we use a standard dualization method utilizing Lagrange multipliers, as described in Fletcher [10]. The key idea is to construct a Lagrange function from both the objective function and the corresponding constraint. For details see [11,14]. Hence we proceed as follows:

$$L=\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{l}\xi_i^2-\sum_{i=1}^{l}\alpha_i(\xi_i-y_i+\langle\mathbf{w},\mathbf{x}_i\rangle+b)$$

In this paper we will modify this idea for the purpose of deriving the convex optimization problems for fuzzy multiple linear regression models and fuzzy nonlinear regression model with numerical inputs. We now briefly review the Tanaka's model[18].

Definition 1. A symmetric fuzzy number A_i denoted as $(\alpha_i,c_i)_L$ is defined by

$$\mu_{A_i}(a_i)=L((a_i-\alpha_i)/c_i)$$

where reference function $L(x)$ satisfies (i) $L(x)=L(-x)$, (ii) $L(0)=1$ and (iii) L is strictly decreasing on $[0,+\infty)$.

As examples of Definition 1, $L(x)=\max(0,1-|x|^p)$, $L(x)=e^{-|x|^p}$ and $L(x)=1/1+|x|^p$ are shown in [9], where $p>0$.

Proposition 1. The possibilistic linear function with fuzzy parameters $A_i=(\alpha_i,c_i)_L$, $i=1,\cdots,d$,

$$Y=A_1x_1+\cdots+A_dx_d,$$

is obtained by

$$Y=\sum_{i=1}^{d}(\alpha_i,c_i)_L x_i=(\langle\boldsymbol{\alpha},\mathbf{x}\rangle,\langle\mathbf{c},|\mathbf{x}|\rangle)_L$$

where $|\mathbf{x}|=(|x_1|,\cdots,|x_d|)$, $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_d)$, and $\mathbf{c}=(c_1,\cdots,c_d)$.

Consider a fuzzy linear model,

$$Y=W_1x_1+\cdots+W_dx_d+B=\langle\mathbf{W},\mathbf{x}\rangle+B$$

where $\mathbf{W}=\left(W_1=(m_{W_1},\alpha_{W_1})_L,\cdots,W_d=(m_{W_d},\alpha_{W_d})_L\right)$ and $B=(m_B,\alpha_B)_L$ are fuzzy parameters. Let us consider fuzzy output $\mathbf{Y}_i=(m_{Y_i},\alpha_{Y_i})$(for the case of real output, $Y_i=y_i$) and crisp input $\mathbf{x}_i$, $i=1,\cdots,l$. The problem in the Tanaka's fuzzy linear regression model is to determine fuzzy parameters $W_1^*,\cdots,W_d^*$ and B^* which

$$\begin{aligned}
&\text{minimize } \; J(\mathbf{W})=\sum_{i=1}^{l}(\langle\boldsymbol{\alpha}_\mathbf{W},|\mathbf{x}|_i\rangle+\alpha_B)\\
&\text{subject to } \boldsymbol{\alpha}_\mathbf{W},\alpha_B\geq 0 \text{ and}\\
&\quad m_{Y_i}-(\langle\mathbf{m}_\mathbf{W},\mathbf{x}_i\rangle+m_B)\\
&\quad\leq|L^{-1}(h)|(\langle\mathbf{m}_\mathbf{W},|\mathbf{x}_i|\rangle+\alpha_B-\alpha_{Y_i}) \qquad (3)\\
&\quad\langle\mathbf{m}_\mathbf{W},\mathbf{x}_i\rangle+m_B-m_{Y_i}\\
&\quad\leq|L^{-1}(h)|(\langle\mathbf{m}_\mathbf{W},|\mathbf{x}_i|\rangle+\alpha_B-\alpha_{Y_i}),\\
&\quad i=1,\cdots,l
\end{aligned}$$

where $\mathbf{m}_\mathbf{W}=(m_{W_1},\cdots,m_{W_d})$ and $\boldsymbol{\alpha}_\mathbf{W}=(\alpha_{W_1},\cdots,\alpha_{W_d})$.

In this section, we present a new method to evaluate fuzzy regression model based on Tanaka's model using regularlization method.

Let

$$\mathbf{W}=(\mathrm{W}_1,\cdots,\mathrm{W}_d),$$
$$\text{for } W_i=(m_{W_i},\alpha_{W_i},\beta_{W_i})_L,\ i=1,\cdots,d,$$
$$\mathbf{m}_\mathbf{W}=(m_{W_1},\cdots,m_{W_d}),\ \boldsymbol{\alpha}_\mathbf{W}=(\alpha_{W_1},\cdots,\alpha_{W_d})$$

and $\boldsymbol{\beta}_\mathbf{W}=(\beta_{W_1},\cdots,\beta_{W_d})$.

We defined for the case of $\boldsymbol{\alpha}_\mathbf{W}=\boldsymbol{\beta}_\mathbf{W}$,

$$\begin{aligned}
\|\mathbf{W}\|^2&=\|\mathbf{m}_\mathbf{W}\|^2+\|\mathbf{m}_\mathbf{W}-\boldsymbol{\alpha}_\mathbf{W}\|^2+\|\mathbf{m}_\mathbf{W}+\boldsymbol{\beta}_\mathbf{W}\|^2\\
&=3\langle\mathbf{m}_\mathbf{W},\mathbf{m}_\mathbf{W}\rangle+2\langle\boldsymbol{\alpha}_\mathbf{W},\boldsymbol{\alpha}_\mathbf{W}\rangle.
\end{aligned}$$

Then, we arrive at the following convex optimization problem for model as follows:

$$\text{minimize } \frac{1}{2}\|\mathbf{W}\|^2+C\sum_{i=1}^{l}\xi_i \quad (4)$$

subject to

$$\begin{cases}
(\langle\boldsymbol{\alpha}_\mathbf{W},|\mathbf{x}_i|\rangle+\alpha_B)=\xi_i,\ \xi_i\geq 0,\ i=1,\cdots,l\\
m_{Y_i}-(\langle\mathbf{m}_\mathbf{W},\mathbf{x}_i\rangle+m_B)\\
\quad\leq|L^{-1}(h)|(\langle\boldsymbol{\alpha}_\mathbf{W},|\mathbf{x}_i|\rangle+\alpha_B-\alpha_{Y_i})\\
(\langle\mathbf{m}_\mathbf{W},\mathbf{x}_i\rangle+m_B)-m_{Y_i}\\
\quad\leq|L^{-1}(h)|(\langle\boldsymbol{\alpha}_\mathbf{W},|\mathbf{x}_i|\rangle+\alpha_B-\alpha_{Y_i})
\end{cases} \quad (5)$$

We construct a Lagrange function as follows:

$$
\begin{aligned}
L = &\frac{1}{2}\parallel \mathbf{W} \parallel^2 + C\sum_{i=1}^{l}\xi_i \\
&-\sum_{i=1}^{l}\alpha_{1i}(\xi_i - \langle \boldsymbol{\alpha}_{\mathbf{W}}, |\mathbf{x}_i| \rangle - \alpha_B) \\
&-\sum_{i=1}^{l}\alpha_{2i}\big(|L^{-1}(h)|\big(\langle \boldsymbol{\alpha}_{\mathbf{W}}, |\mathbf{x}_i|\rangle + \alpha_B - \alpha_{Y_i}\big) \\
&-m_{Y_i} + \big(\langle \mathbf{m_w}, \mathbf{x}_i\rangle + m_B\big)\big) \\
&-\sum_{i=1}^{l}\alpha_{2i}^{*}\big(|L^{-1}(h)|\big(\langle \boldsymbol{\alpha}_{\mathbf{W}}, |\mathbf{x}_i|\rangle + \alpha_B - \alpha_{Y_i}\big) \\
&+m_{Y_i} - \big(\langle \mathbf{m_w}, \mathbf{x}_i\rangle + m_B\big)\big) \\
&-\sum_{i=1}^{l}\eta_i\xi_i
\end{aligned} \tag{6}
$$

$\alpha_{2i}^{(*)}, \eta_i \geq 0, i = 1,\cdots,l$

It follows from the saddle point condition that the partial derivatives of L with respect to the primal variable (W,B,ξ) have to vanish for optimality.

$$\frac{\partial L}{\partial m_B} = \sum_{i=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right) = 0 \tag{7}$$

$$\frac{\partial L}{\partial \alpha_B} = \sum_{i=1}^{l}\alpha_{1i} - |L^{-1}(h)|\sum_{i=1}^{l}\left(\alpha_{2i} + \alpha_{2i}^{*}\right) = 0 \tag{8}$$

$$\frac{\partial L}{\partial \mathbf{m_W}} = 3\mathbf{m_W} - \sum_{i=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right)\mathbf{x}_i = 0 \tag{9}$$

$$\frac{\partial L}{\partial \boldsymbol{\alpha}_{\mathbf{W}}} = 2\boldsymbol{\alpha}_{\mathbf{W}} + \sum_{i=1}^{l}\alpha_{1i}|\mathbf{x}_i| - |L^{-1}(h)|\sum_{i=1}^{l}\left(\alpha_{2i} + \alpha_{2i}^{*}\right)|\mathbf{x}_i| = 0 \tag{10}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_{1i} - \eta_i = 0 \tag{11}$$

Eq.(8) (9) and (10) can be rewritten as follows, respectively,

$$\mathbf{m_W} = \frac{1}{3}\sum_{i=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right)\mathbf{x}_i$$

$$\boldsymbol{\alpha}_{\mathbf{W}} = \frac{1}{2}\sum_{i=1}^{l}\left\{\left(\alpha_{2i} + \alpha_{2i}^{*}\right)|L^{-1}(h)| - \alpha_{1i}\right\}|\mathbf{x}_i| \tag{12}$$

Substituting (7) ~ (11) into (6) yields the optimization problem

$$\text{maximize } -\frac{1}{2}\parallel \mathbf{W} \parallel^2 + \sum_{i=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right)m_{Y_i} + \sum_{i=1}^{l}\left(\alpha_{2i} + \alpha_{2i}^{*}\right)|L^{-1}(h)|\alpha_{Y_i} \tag{13}$$

where

$$
\begin{aligned}
\parallel \mathbf{W} \parallel^2 &= 3\langle \mathbf{m_W}, \mathbf{m_W}\rangle + 2\langle \boldsymbol{\alpha}_{\mathbf{W}}, \boldsymbol{\alpha}_{\mathbf{W}}\rangle \\
&= \frac{1}{3}\sum_{i,j=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right)\left(\alpha_{2j} - \alpha_{2j}^{*}\right)\langle \mathbf{x}_i, \mathbf{x}_j\rangle \\
&+ \frac{1}{2}\sum_{i,j=1}^{l}\left[\left(\left(\alpha_{2i} + \alpha_{2i}^{*}\right)|L^{-1}(h)| - \alpha_{1i}\right)\right. \\
&\left.\left(\left(\alpha_{2j} + \alpha_{2j}^{*}\right)|L^{-1}(h)| - \alpha_{1j}\right)\right]\langle |\mathbf{x}_i|, |\mathbf{x}_j| \rangle
\end{aligned}
$$

subject to

$$\sum_{i=1}^{l}\left(\alpha_{2i} - \alpha_{2i}^{*}\right) = 0,\ \alpha_{2i}^{(*)} \geq 0,\ i = 1,\cdots,d. \tag{14}$$

Define $\alpha'_{W_k} = \max\left\{\alpha_{W_k}, 0\right\}$ and $\boldsymbol{\alpha}'_{\mathbf{W}} = \left(\alpha'_{W_1}, \cdots, \alpha'_{W_k}\right)$

Then we have

$$Y = \langle \mathbf{W}', \mathbf{x}\rangle + B = \left(\langle \mathbf{m_W}, \mathbf{x}\rangle, \langle \boldsymbol{\alpha}'_{\mathbf{W}}, \mathbf{x}\rangle\right) + B,$$

where $\mathbf{W}' = (m_W, \alpha'_W)$.

Let

$$H_R^i(\alpha) = m_{Y_i} - \langle \mathbf{m_W}, \mathbf{x}_i\rangle + |L^{-1}(h)|\left(\langle \boldsymbol{\alpha}'_{\mathbf{W}}, |\mathbf{x}_i|\rangle + \alpha - \alpha_{Y_i}\right)$$

and

$$H_L^i(\alpha) = m_{Y_i} - \langle \mathbf{m_W}, \mathbf{x}_i\rangle - |L^{-1}(h)|\left(\langle \boldsymbol{\alpha}'_{\mathbf{W}}, |\mathbf{x}_i|\rangle + \alpha - \alpha_{Y_i}\right)$$

From (5), we have $H_L^i(\alpha_B) \leq m_B \leq H_R^i(\alpha_B)$ for all i. Hence, we should take m_B and α_B as follow ;

$$\alpha_B = \inf\left\{\alpha \geq 0 \mid \bigcap_{i=1}^{l}\left[H_L^i(\alpha), H_R^i(\alpha)\right] \neq \varnothing\right\} \text{ and}$$

$$m_B \in \bigcap_{i=1}^{l}\left[H_L^i(\alpha_B), H_R^i(\alpha_B)\right].$$

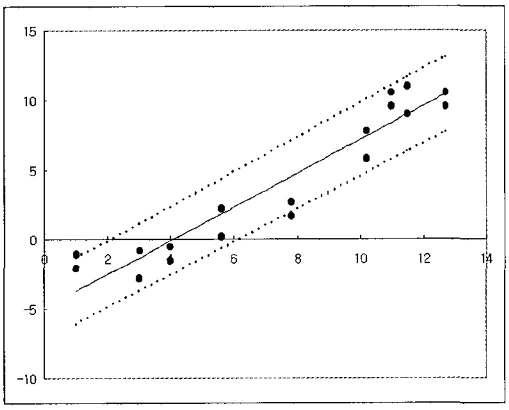
Table 1

x	$Y=(y,\alpha_y)$
1	(−1.6, 0.5)
3	(−1.8, 1.0)
4	(−1.0, 0.5)
5.6	(1.2, 1.0)
7.8	(2.2, 0.5)
10.2	(6.8, 1.0)
11.0	(10.0, 0.5)
11.5	(1.0, 1.0)
12.7	(10.0, 0.5)

Example 1. We consider a data from Gunn[12], which were constructed using the original and symmetric fuzzfied Y_i as shown in Table 1. Using the data in Table 1, the obtained results from Tanaka's model and our model with $h=0.5$ and $L(x)=1-x$ are shown in Table 2. In this example for our model, $C=500$ is used but it heuristically determined. In fact, the values of $C \geq 500$ give almost same results. Comparing the result of $J(\mathbf{W})$, index of fuzziness, for Tanaka's model, the resulting solutions are very close even though different methods are used.

Table 2

Model	Fuzzy coefiicient	$J(\mathbf{W})$: index of fuzziness
Tanaka's model	$m_W = 1.023$, $\alpha_W = 0.274$, $m_B = -3.735$, $\alpha_B = 2,453$	40.371
Our model	$m_W = 1.208$, $\alpha_W = 0.058$, $m_B = -4.917$, $\alpha_B = 4,660$	45.788



$\bullet : y \pm \alpha_y$, $- : \mu_Y^{-1}(1)$, $\cdots : \mu_Y^{-1}(0)$

Figure 1. The fuzzy linear regression model

3. Fuzzy nonlinear regression

In this section we consider fuzzy nonlinear regression for the case of numerical inputs and fuzzy output. In other words, we are interested in making the model nonlinear. To do this, we need to briefly look at again the idea used in support vector machine(SVM) for crisp nonlinear regression. See for details [12,15,19].

In the case where a linear regression function is inappropriate SVM makes algorithm nonlinear. This could be achieved by simply preprocessing input patterns $\mathbf{x}_i$ by a map $\Phi : R^d \rightarrow F$ into some feature space F, as described in [1] and then applying the standard SVM regression algorithm. A rather old trick [1] can be used to accomplish this in an astonishingly straightforward way. First notice that the only way in which the data appears in algorithm is in the form of dot products $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. The algorithm would only depend on the data through dot products in F, i.e. on functions of the form $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. Hence it suffices to know and use $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ instead of $\Phi(\cdot)$ explicitly. We can overcome the curse of dimensionality by using this kernel. The well used kernels for regression problem are given below.

$$K(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^p \quad \text{: Polynomial kernel}$$

$$K(\mathbf{x}, \mathbf{y}) = e^{\frac{-\| \mathbf{x} - \mathbf{y} \|^2}{2\sigma^2}} \quad \text{: Gaussian kernel}$$

$$K(\mathbf{x}, \mathbf{y}) = \tan h(\Phi \langle \mathbf{x}, \mathbf{y} \rangle + \theta) \quad \text{: Hyperbolic tangent kernel}$$

Hence we arrive at the following optimization problem for the nonlinear model by replacing $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ with $K(\mathbf{x}_i, \mathbf{x}_j)$ in the optimization problem for the model.

$$\text{maximize} \quad -\frac{1}{2} \| \mathbf{W} \|^2 + \sum_{i=1}^{l} (\alpha_{2i} - \alpha_{2i}^*) m_{Y_i} + \sum_{i=1}^{l} (\alpha_{2i} + \alpha_{2i}^*) |L^{-1}(h)| \alpha_{Y_i}$$

where

$$\| \mathbf{W} \|^2 = \frac{1}{3} \sum_{i,j=1}^{l} (\alpha_{2i} - \alpha_{2i}^*)(\alpha_{2j} - \alpha_{2j}^*) K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{2} \sum_{i,j=1}^{l} \left[\left((\alpha_{2i} + \alpha_{2i}^*) |L^{-1}(h)| - \alpha_{1i} \right) \left((\alpha_{2j} + \alpha_{2j}^*) |L^{-1}(h)| - \alpha_{1j} \right) \right] K(|\mathbf{x}_i|, |\mathbf{x}_j|)$$

and $K(\mathbf{x}, \mathbf{y})$ is the kernel function performing the nonlinear mapping into feature space.

Here, we should notice that the constraint are unchanged, $\sum_{i=1}^{l} (\alpha_{2i} - \alpha_{2i}^*) = 0$, $\alpha_{2i}^{(*)} \geq 0$.

In case of

$$\sum_{i=1}^{l} \left[\left((\alpha_{2i} + \alpha_{2i}^*) |L^{-1}(h)| - \alpha_{1i} \right) \right] K(|\mathbf{x}_i|, |\mathbf{x}_j|) \geq 0,$$

we have

$$Y = \left(\frac{1}{3} \sum_{i=1}^{l} (\alpha_{2i} - \alpha_{2i}^*) K(|\mathbf{x}_i|, \mathbf{x}), \; \frac{1}{2} \sum_{i=1}^{l} \left[(\alpha_{2i} + \alpha_{2i}^*) |L^{-1}(h)| - \alpha_{1i} \right] K(|\mathbf{x}_i|, \mathbf{x}) \right) + B$$

And m_B, α_B can be obtained as follow:

$$\alpha_B = \inf \left\{ \alpha \geq 0 \mid \bigcap_{i=1}^{l} \left[H_L^i(\alpha), H_R^i(\alpha) \right] \neq \varnothing \right\} \text{ and}$$

$$m_B \in \bigcap_{i=1}^{l} \left[H_L^i(\alpha_B), H_R^i(\alpha_B) \right].$$

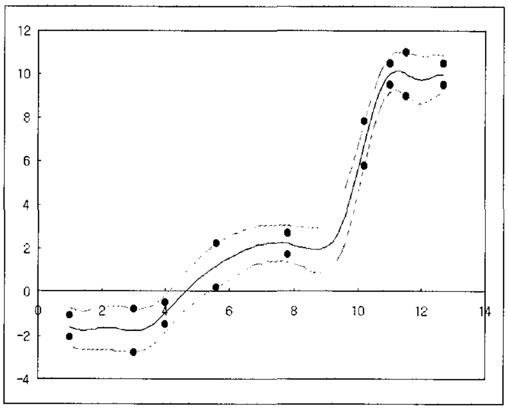
where

$$H_R^i(\alpha) = m_{Y_i} - \frac{1}{3} \sum_{j=1}^{l} (\alpha_{2j} - \alpha_{2j}^*) K(\mathbf{x}_j, \mathbf{x}_i) + |\mathrm{L}^{-1}(\mathrm{h})| \left\{ \frac{1}{2} \sum_{j=1}^{l} \left[(\alpha_{2j} + \alpha_{2j}^*) |L^{-1}(h)| - \alpha_{1j} \right] K(\mathbf{x}_j, \mathbf{x}_i) + \alpha - \alpha_{Y_i} \right\}$$

and

$$H_L^i(\alpha) = m_{Y_i} - \frac{1}{3} \sum_{j=1}^{l} (\alpha_{2j} - \alpha_{2j}^*) K(\mathbf{x}_j, \mathbf{x}_i) - |L^{-1}(h)| \left\{ \frac{1}{2} \sum_{j=1}^{l} \left[(\alpha_{2j} + \alpha_{2j}^*) |L^{-1}(h)| - \alpha_{1j} \right] K(\mathbf{x}_j, \mathbf{x}_i) + \alpha - \alpha_{Y_i} \right\}$$

Example 2. We now apply fuzzy nonlinear regression model to the data in Table 1 with $h=0.5$ and $L(x)=1-x$ again. According to Gunn [12], the non-linear regression model is appropriate for the original crisp data of Table 1. When we apply fuzzy nonlinear regression model to this data set, we have the residual sum 1.667. Hence we can recognize fuzzy nonlinear model is more appropriate than the linear model. For this data set we use Gaussian kernel with $\sigma=1.0$ and $C=500$. These parameters are determined in the heuristic way.



$\bullet : y \pm \alpha_y$, $-: \mu_Y^{-1}(1), \mu_Y^{-1}(0)$

Figure 2. The fuzzy non-linear regression model

4. Fuzzy input-output with real coefficients

In this section, we finally consider the following fuzzy linear regression model with fuzzy input-output data but real coefficients,

$$Y = < \mathbf{w}, \mathbf{X} > + B, \tag{15}$$

where $\mathbf{W} = (w_1, \cdots, w_d)$ is real vector, $|\mathbf{w}| = (|w_1|, \cdots, |w_d|)$ and $\mathbf{X}_i = \left(X_{i1} = (m_{X_{i1}}, \sigma_{X_{i1}}), \cdots, X_{id} = (m_{X_{id}}, \sigma_{X_{id}})\right)$, $\mathbf{Y}_i = (m_{Y_i}, \alpha_{Y_i})$, $i = 1, \cdots, l$.

We consider the following convex optimization problem for this model;

$$\text{minimize } \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{l} \xi_i \tag{16}$$

subject to

$$\begin{cases} \left(\langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle - \alpha_B \right) = \xi_i, \ \xi_i \geq 0, \ i = 1, \cdots, l \\ m_{Y_i} - \left(\langle \mathbf{w}, \mathbf{m}_{X_i} \rangle + m_B \right) \\ \qquad \leq |L^{-1}(h)| \left(\langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle + \alpha_B - \alpha_{Y_i} \right) \\ \langle \mathbf{w}, \mathbf{m}_{X_i} \rangle + m_B - m_{Y_i} \\ \qquad \leq |L^{-1}(h)| \left(\langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle + \alpha_B - \alpha_{Y_i} \right) \end{cases} \tag{17}$$

We construct a Lagrange function as follows:

$$\begin{aligned} L = & \frac{1}{2} \| \mathbf{W} \|^2 + C \sum_{i=1}^{l} \xi_i \\ & - \sum_{i=1}^{l} \alpha_{1i} \left(\xi_i - \langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle - \alpha_B \right) \\ & - \sum_{i=1}^{l} \alpha_{2i} \Big(|L^{-1}(h)| \left(\langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle + \alpha_B - \alpha_{Y_i} \right) \\ & - m_{Y_i} + \left(\langle \mathbf{w}, \mathbf{m}_{X_i} \rangle + m_B \right) \Big) \\ & - \sum_{i=1}^{l} \alpha_{2i}^* \Big(|L^{-1}(h)| \left(\langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle + \alpha_B - \alpha_{Y_i} \right) \\ & + m_{Y_i} - \left(\langle \mathbf{w}, \mathbf{m}_{X_i} \rangle + m_B \right) \Big) \\ & - \sum_{i=0}^{l} \eta_i \xi_i \end{aligned} \tag{18}$$

$$\alpha_{2i}^{(*)}, \eta_i \geq 0, \ i = 0, \cdots, l$$

It follows from the saddle point condition that the partial derivatives of L *w.r.t.* the primal variable $(\mathbf{w}, B, \xi)$ have to vanish for optimality.

$$\frac{\partial L}{\partial m_B} = \sum_{i=1}^{l} \left(\alpha_{2i} - \alpha_{2i}^* \right) = 0 \tag{19}$$

$$\frac{\partial L}{\partial \alpha_B} = \sum_{i=1}^{l} \alpha_{1i} - |L^{-1}(h)| \sum_{i=1}^{l} \left(\alpha_{2i} + \alpha_{2i}^* \right) = 0 \tag{20}$$

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = & \ \mathbf{w} - \sum_{i=1}^{l} \alpha_{1i} \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i} \\ & - |L^{-1}(h)| \sum_{i=1}^{l} \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i} \left(\alpha_{2i} + \alpha_{2i}^* \right) \\ & - \sum_{i=1}^{l} \left(\alpha_{2i} - \alpha_{2i}^* \right) \mathbf{m}_{X_i} = 0 \end{aligned} \tag{21}$$

where

$\mathbf{sgn}(\mathbf{w}) = \left(sgn(w_1), sgn(w_2), \cdots, sgn(w_d) \right)$,

and

$\mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{\mathbf{X}_i} = \left(sgn(w_1) \alpha_{X_{i1}}, \cdots, sgn(w_d) \alpha_{X_{id}} \right)$,

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_{1i} - \eta_i = 0, \ i = 0, \cdots, l. \tag{22}$$

Eq.(21) can be rewritten as follows,

$$\begin{aligned} \mathbf{w} = & \sum_{i=1}^{l} \alpha_{1i} \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i} \\ & + |L^{-1}(h)| \sum_{i=1}^{l} \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i} \left(\alpha_{2i} + \alpha_{2i}^* \right) \\ & + \sum_{i=1}^{l} \left(\alpha_{2i} - \alpha_{2i}^* \right) m_{X_i} \end{aligned} \tag{23}$$

and therefore

$$\begin{aligned} Y = & \sum_{i=1}^{l} \langle \alpha_{1i} \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i}, \mathbf{X} \rangle \\ & + |L^{-1}(h)| \sum_{i=1}^{l} \langle \mathbf{sgn}(\mathbf{w}) \boldsymbol{\alpha}_{X_i} \left(\alpha_{2i} + \alpha_{2i}^* \right), \mathbf{X} \rangle \\ & + \sum_{i=1}^{l} \langle \left(\alpha_{2i} - \alpha_{2i}^* \right) \mathbf{m}_{X_i}, \mathbf{X} \rangle + B \end{aligned} \tag{24}$$

Substituting (19) ~ (22) into (18) yields the optimization problem

$$\text{maximize} \quad -\frac{1}{2}\|\mathbf{W}\|^2 + \sum_{i=1}^{l}(\alpha_{2i}-\alpha_{2i}^*)m_{Y_i} + \sum_{i=1}^{l}(\alpha_{2i}+\alpha_{2i}^*)|L^{-1}(h)|\alpha_{Y_i} \quad (25)$$

where

$$\begin{aligned}\|\mathbf{w}\|^2 = &\sum_{i,j=1}^{l}\langle \alpha_{1i}\mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_i}, \alpha_{1j}\mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_j}\rangle \\ &+L^{-1}(h)^2\sum_{i,j=1}^{l}(\alpha_{2i}+\alpha_{2i}^*)(\alpha_{2j}+\alpha_{2j}^*) \langle \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_i}, \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_j}\rangle \\ &+\sum_{i,j=1}^{l}(\alpha_{2i}-\alpha_{2i}^*)(\alpha_{2j}-\alpha_{2j}^*)\langle \mathbf{m}_{X_i}, \mathbf{m}_{X_j}\rangle \\ &+2|L^{-1}(h)|\sum_{i,j=1}^{l}\alpha_{1i}(\alpha_{2j}+\alpha_{2j}^*) \langle \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_i}, \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_j}\rangle \\ &+2|L^{-1}(h)|\sum_{i,j=1}^{l}(\alpha_{2i}+\alpha_{2i}^*)(\alpha_{2i}-\alpha_{2i}^*) \langle \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_i}, \mathbf{m}_{X_j}\rangle \\ &+2\sum_{i,j=1}^{l}\alpha_{1i}(\alpha_{2j}-\alpha_{2j}^*)\langle \mathbf{sgn}(\mathbf{w})\boldsymbol{\alpha}_{X_i}, \mathbf{m}_{X_j}\rangle \end{aligned} \quad (26)$$

subject to

$$\sum_{i=1}^{l}(\alpha_{2i}-\alpha_{2i}^*)=0,\ \alpha_{2i}^{(*)}\ge 0,\ i=1,\cdots,l. \quad (27)$$

m_b and α_B can be computed similarly as previous case as follows;

$$\alpha_B = \inf\left\{\alpha \ge 0 \mid \bigcap_{i=1}^{l}\left[H_L^i(\alpha), H_R^i(\alpha)\right] \ne \varnothing\right\} \text{ and}$$

$$m_B \in \bigcap_{i=1}^{l}\left[H_L^i(\alpha_B), H_R^i(\alpha_B)\right].$$

where

$$H_R^i(\alpha) = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{X_i}\rangle + |L^{-1}(h)|(\langle \mathbf{w}, \boldsymbol{\alpha}_{X_i}\rangle + \alpha - \alpha_{Y_i})$$

and

$$H_R^i(\alpha) = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{X_i}\rangle - |L^{-1}(h)|(\langle \mathbf{w}, \boldsymbol{\alpha}_{X_i}\rangle + \alpha - \alpha_{Y_i})$$

5. Conclusion

In this paper, we have presented a new method to evaluate fuzzy linear and fuzzy non-linear regression models based on Tanaka's approach using the idea of regularlization technique. From the results in examples we realize that the derived fuzzy regression models derive the satisfying solutions and is a very attractive approach to modeling fuzzy date, especially for fuzzy non-linear regression model. The main formulation results in a global quadratic optimization problem with box constraints. However this is not a computationally expensive way. Kernel parameter σ and control parameter C determined in a heuristic way. The obvious question that arises is which are the best for a particular problem? hence we need model selection method to determine these parameters.

References

[1] A. Aizerman, E. M. Braverman and L. I. Rozoner, Theoretical foundations of the potential function method in pattern recognition learning, Automation and Remote Control 25 (1964), 821-837.

[2] J. Buckley, T. Feuring and Y. Hayashi, Multivariate non-linear fuzzy regression: An evolutionary algorithm approach, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 7 (1999) 83-98.

[3] J. Buckley and T. Feuring, Linear and non-linear fuzzy regression: Evolutionary algorithm solutions, Fuzzy Sets and Systems 112 (2000) 381-394.

[4] A. Celminš, Least model fitting to fuzzy vector data, Fuzzy Sets and Systems 22 (1987) 245-269.

[5] A. Celminš, Multidimensional least-squares fitting of fuzzy models, Math. Modelling 9 (1987) 669-690.

[6] A. Celminš, A practical approach to nonlinear fuzzy regression, SLAM J. Sci. Stat. Comput., Vol. 12 No. 3 (1991) 521-546.

[7]] P. Diamond, Fuzzy least squares, Inform. Sci. 46 (1988) 141-157.

[8] B. Heshmaty, A. Kandel, Fuzzy linear regression and its applications to forecasting in uncertain environment, Fuzzy Sets and Systems 15 (1985) 159-191.

[9] D. Dubois and H. Prade, Theory and Applications, Fuzzy Sets and Systems, Academic Press, New York, 1980.

[10] R. Fletcher, Practical Methods of Optimization, John Wiley and Sons, Inc., 2nd edition, 1987.

[11] H. Goldstein, Classical Mechanis, Addison-Wesley, Reading, MA, 1986.

[12] S. Gunn, Support Vector Machines for Classification and Regression, ISIS Technical Report, U. of Southampton, 1998.

[13] J. Kacprzyk and M. Fedrizzi, Fuzzy Regression Analysis(Physica-Verlag, Heidelberg, 1992)

[14] G. P. McCormick, Nonlinear Programing: Theory, Algorithms and Applications, Wiley-Interscience, New York, NY, 1983.

[15] A. J. Smola and B. Scholkopf, A Tutorial on Support Vector Regression, NeuroCOLT2 Technical Report, NeuroCOLT, 1998.

[16] H. Tanaka, Fuzzy data analysis by possibilistic linear models, Fuzzy Sets and Systems 24 (1987) 363-375.
[17] v H. Tanaka, S. Uejima and K. Asia, Linear regression analysis with Fuzzy model, IEEE Trans. Man. Cybernet. 12 (6) (1982) 903-907.
[18] H. Tanaka, J. Watada, Possibilistic linear systems and their applications to linear regression model, Fuzzy Sets and Systems 27 (1988) 275-289.
[19] V. Vapnik, The Nature of Statistical learning Theory, Springer, 1995.
[20] J. Watada, theory of fuzzy multivariate analysis and its applications, PH. D. Dissertation, University of Osaka Prefecture, 1983.

저 자 소 개

Dug Hun Hong
received the B.S., M.S. degrees in mathematics from Kyungpook National University , Taegu, Korea and Ph. D degree in mathematics from University of Minnesota, Twin City in 1981, 1983 and 1990, respectively. From 1991 to 2003, he worked with department of Statistics and School of Mechanical and Automotive Engineering, Catholic University of Daegu, Daegu, Korea, Since 2004, he has been a Professor in Department of Mathematics, Myongji University, Korea. His research interests include general fuzzy theory with application and probability theory.

e-mail : dhhong@mju.ac.kr

Kyung Tae Kim
received the B.S. degree in Electrical Engineering from Kyungpook National University, Taegu, Korea and M.E and Ph. D degrees in Electrical Engineering from Yonsei University, Seoul, Korea in 1978, 1980 and 1987, respectively. Since 1987, he has been a Professor in Department of Electronics and Electrical Information Engineering, Kyungwon University, Korea. His research interests include mobile communications and optical communications.

e-mail : ktkim@kyungwon.ac.kr