Support vector fuzzy regression machines

Dug Hun Hong[a,*], Changha Hwang[b]

[a] *School of Mechanical and Automotive Engineering, Catholic University of Daegu, 330 Keumrak 1-ri Hayang-up Kyongsan, Kyungbuk 712 - 702, South Korea*
[b] *Department of Statistical Information, Catholic University of Daegu, Kyungbuk 712 - 702, South Korea*

Abstract

Support vector machine (SVM) has been very successful in pattern recognition and function estimation problems. In this paper, we introduce the use of SVM for multivariate fuzzy linear and nonlinear regression models. Using the basic idea underlying SVM for multivariate fuzzy regressions gives computational efficiency of getting solutions.
ⓒ 2002 Published by Elsevier Science B.V.

Keywords: Fuzzy inference systems; Support vector machine; Fuzzy regression; Fuzzy system models

1. Introduction

Linear regression models are widely used today in business, administration, economics, engineering, as well as in many other traditionally nonquantitative fields such as social, health, and biological sciences. In all cases of fuzzy regression, the linear regression is recommended for practical situations when decisions often have to be made on the basis of imprecise and/or partially available data. Many different fuzzy regression approaches have been proposed. Fuzzy regression, as first developed by Tanaka et al. [14] in a linear system, is based on the extension principle. Tanaka et al. [14] initially applied their fuzzy linear regression procedure to nonfuzzy experimental data. In the experiments that followed this pioneering effort, Tanaka et al. [15] used fuzzy input experimental data to build fuzzy regression models. Fuzzy input data used in these experiments were given in the form of triangular fuzzy numbers. The process is explained in more detail by Dubois and Prade [10]. A technique for linear least-squares fitting of fuzzy variable was developed by Diamond [7,8] giving the solution to an analog of the normal equation of classical least squares. A collection of recent papers dealing with several approaches to fuzzy regression analysis can be found [12].

---

* Corresponding author. Tel.: +82-53-850-2712; fax: +82-53-850-2799.
E-mail address: dhhong@cuth.cataegu.ac.kr (D.H. Hong).

1   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
3 form with crisp inputs and fuzzy output. However, some papers, for example [2,3,5], were concerned with the data set with fuzzy inputs and fuzzy output. By the way, in this paper we will treat fuzzy
5 nonlinear regression for data of the form with crisp inputs and fuzzy output.

    In this paper, we discuss multivariate fuzzy linear and nonlinear regression by support vector
7 machine (SVM). SVM has been recently introduced for solving pattern recognition and function estimation problems [1,4,6,9,11,13,16–18]. SVM is a nonlinear generalization of the Generalized
9 Portrait algorithm developed in Russia in the 1960s. In its present form, the SVM was developed at AT&T Bell Laboratories by Vapnik and co-workers. Due to this industrial context, SVM research has
11 up to date had a sound orientation towards real-world applications. SVM learning has now evolved into an active area of research. Moreover, it is in the process of entering the standard methods
13 toolbox of machine learning. SVM is based on the idea of structural risk minimization, which shows that the generalization error is bounded by the sum of the training set and a term depending on the
15 Vapnik–Chervonenkis dimension. By minimizing this bound, high generalization performance can be achieved. Moreover, unlike other machine learning methods, SVMs generalization error is not
17 related to the problem's input dimensionality. This explains why SVM can have good performance even in high dimensional problem.
19   The main difference between our SVM approach and the nonlinear approaches by Buckley et al. [2,3] and Celmins [5] is not crisp input–fuzzy output versus fuzzy input–fuzzy output, but model-free
21 versus model-dependent.

    The rest of this paper is organized as follows. Section 2 illustrates the SVM regression procedures
23 for fuzzy multivariate linear models. Section 3 describes how to apply this idea to the fuzzy multivariate nonlinear model with numerical inputs and fuzzy output. Section 4 gives some conclusions.

25 ## 2. SVM fuzzy linear regression

    In this section, we will modify the underlying idea of SVM for the purpose of deriving the
27 convex optimization problems for multivariate fuzzy linear regression models. In Section 3 we will consider multivariate fuzzy nonlinear regression model for numerical inputs and fuzzy output. The
29 basic idea of SVM gives computational efficiency in finding solutions of fuzzy regression models particularly for multivariate case. Here, we consider three different models for multivariate fuzzy
31 linear regression.

    To do this we need some preliminaries. Let $X = (m, \alpha, \beta)$ be a triangular fuzzy numbers when m
33 is the model value of X and α and β are the left and right spreads respectively. If $\alpha = \beta$, we can write $X = (m, \alpha)$. On the space $T(R)$ of all triangular fuzzy numbers we use the metric d defined by

35 $$d(X, Y)^2 = (m_X - m_Y)^2 + ((m_X - \alpha_X) - (m_Y - \alpha_Y))^2 + ((m_X + \beta_X) - (m_Y + \beta_Y))^2,$$

where $X = (m_X, \alpha_X, \beta_X)$ and $Y = (m_Y, \alpha_Y, \beta_Y)$ are any two vectors of triangular fuzzy numbers in
37 $T(R)$. A linear structure is defined on $T(R)$ by $(m_X, \alpha_X, \beta_X) + (m_Y, \alpha_Y, \beta_Y) = (m_X + m_Y, \alpha_X + \alpha_Y, \beta_X + \beta_Y), t(m, \alpha, \beta) = (tm, t\alpha, t\beta)$ if $t \geqslant 0$, and $t(m, \alpha, \beta) = (tm, t\beta, t\alpha)$ if $t < 0$.

1 *2.1. Case of fuzzy inputs and fuzzy output*

Suppose that observations consist of data pairs $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, l$, where $\mathbf{X}_i = ((m_{X_{i1}}, \alpha_{X_{i1}}, \beta_{X_{i1}}), \ldots,$
3 $(m_{X_{id}}, \alpha_{X_{id}}, \beta_{X_{id}})) \in T(R)^d$ and $Y_i = (m_{Y_i}, \alpha_{Y_i}, \beta_{Y_i}) \in T(R)$. Here $T(R)$ and $T(R)^d$ are the set of triangular fuzzy numbers and the set of d-vectors of triangular fuzzy numbers, respectively. Let
5 $\mathbf{m}_{\mathbf{X}_i} = (m_{X_{i1}}, \ldots, m_{X_{id}})$, $\boldsymbol{\alpha}_{X_i} = (\alpha_{X_{i1}}, \ldots, \alpha_{X_{id}})$ and $\boldsymbol{\beta}_{X_i} = (\beta_{X_{i1}}, \ldots, \beta_{X_{id}})$.

Now, we consider the following two models $M1$ and $M2$:

$$M1: \ f(\mathbf{X}) = B + \langle \mathbf{w}, \mathbf{X} \rangle, \quad B \in T(R), \quad \mathbf{w} \in R^d,$$

$$M2: \ f(\mathbf{X}) = b + \langle \mathbf{w}, \mathbf{X} \rangle, \quad b \in R, \quad \mathbf{w} \in R^d.$$

7 For computational simplicity we assume $\mathbf{X}_i, Y_i$ and B are symmetric triangular fuzzy numbers. Then model $M1$ can be rewritten as

9 $$M1: \ f(\mathbf{X}) = (\langle \mathbf{w}, m_{\mathbf{X}} \rangle + m_B, \langle |\mathbf{w}|, \alpha_{\mathbf{X}} \rangle + \alpha_B),$$

where $|\mathbf{w}| = (|w_1|, |w_2|, \ldots, |w_d|)$. Now, we are going to consider how to get solutions for two models $M1$ and $M2$. In fact, the model $M2$ is straightforward from the model $M1$.

11

Model $M1$: We arrive at the following convex optimization problem for the model $M1$ by mod-
13 ifying the basic idea of SVM for crisp linear regression [11,13,16–18].

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{k=1}^{2} \sum_{i=1}^{l} (\xi_{ki} + \xi_{ki}^*)$$

$$\text{subject to} \quad \begin{cases} m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle - m_B \leqslant \varepsilon + \xi_{1i}, \\ \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + m_B - m_{Y_i} \leqslant \varepsilon + \xi_{1i}^*, \\ (m_{Y_i} - \alpha_{Y_i}) - (\langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + m_B - \langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle - \alpha_B) \leqslant \varepsilon + \xi_{2i}, \\ (\langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + m_B - \langle |\mathbf{w}|, \boldsymbol{\alpha}_{X_i} \rangle - \alpha_B) - (m_{Y_i} - \alpha_{Y_i}) \leqslant \varepsilon + \xi_{2i}^*, \\ \xi_{ki}, \xi_{ki}^* \geqslant 0, \ k = 1, 2. \end{cases} \quad (1)$$

Then, constructing a Lagrange function and differentiating it with regard to $m_B, \alpha_B, \mathbf{w}, \xi_{ki}$ and ξ_{ki}^*,
15 $k = 1, 2$, we can derive the corresponding dual optimization problem for model $M1$ as follows:

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\|\mathbf{w}\|^2 - \varepsilon \sum_{k=1}^{2} \sum_{i=1}^{l} (\alpha_{ki} + \alpha_{ki}^*) \\ \quad + \sum_{i=1}^{l} m_{Y_i}(\alpha_{1i} - \alpha_{1i}^*) + \sum_{i=1}^{l} (m_{Y_i} - \alpha_{Y_i})(\alpha_{2i} - \alpha_{2i}^*) \end{cases}$$

$$\text{subject to} \quad \sum_{i=1}^{l} (\alpha_{ki} - \alpha_{ki}^*) = 0, \ \alpha_{ki}, \alpha_{ki}^* \in [0, C], \ k = 1, 2,$$

1 where

$$\mathbf{w} = \sum_{i=1}^{l} (\alpha_{1i} - \alpha_{1i}^*) \mathbf{m}_{\mathbf{X}_i} + \sum_{i=1}^{l} (\alpha_{2i} - \alpha_{2i}^*)(\mathbf{m}_{\mathbf{X}_i} - |\boldsymbol{\alpha}_{X_i}|).$$

3 Solving the above equation under constraints determines the Lagrange multipliers $\alpha_{ki}, \alpha_{ki}^*$, and the optimal regression function is given by

$$f(\mathbf{X}) = \sum_{i=1}^{l} (\alpha_{1i} - \alpha_{1i}^*)\langle \mathbf{m}_{\mathbf{X}_i}, \mathbf{X} \rangle + \sum_{i=1}^{l} (\alpha_{2i} - \alpha_{2i}^*)\langle \mathbf{m}_{\mathbf{X}_i} - |\boldsymbol{\alpha}_{X_i}|, \mathbf{X} \rangle + B. \tag{2}$$

5

Now we need to find m_B and α_B. By Karush–Kuhn–Tucker (KKT) conditions, we can compute
7 m_B as follows:

$$\begin{cases} \alpha_{1i}(\varepsilon + \xi_{1i} - m_{Y_i} + \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + m_B) = 0, \\ \alpha_{1i}^*(\varepsilon + \xi_{1i}^* + m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle - m_B) = 0, \\ (C - \alpha_{1i})\xi_{1i} = 0, \ (C - \alpha_{1i}^*)\xi_{1i}^* = 0 \end{cases}$$

9 and hence

$$\begin{cases} m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle - \varepsilon & \text{for } \alpha_{1i} \in (0, C), \\ m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + \varepsilon & \text{for } \alpha_{1i}^* \in (0, C). \end{cases}$$

11 To find α_B we need to solve the optimization problem given below

$$\operatorname*{minimize}_{\alpha_B \geqslant 0} \sum_{i=1}^{l} \{|m_{Y_i} - \alpha_{Y_i} - (\langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + m_B - \langle \mathbf{w}, \boldsymbol{\alpha}_{X_i} \rangle - \alpha_B)|_\varepsilon\},$$

13 where ε-insensitive loss function $|\xi|_\varepsilon$ is defined by

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leqslant \varepsilon, \\ |\xi| - \varepsilon & \text{otherwise.} \end{cases}$$

15 *Model M2*: Since the solution for the model $M2$ is straightforward from the model $M1$, we can easily get solutions as follows. Here the objective function to be maximized is the same as the one
17 for the model $M1$ but constraints are a little bit different. We can compute $\mathbf{w}$, m_B and α_B in the same way as the one in the model $M1$

$$\text{maximize} \begin{cases} -\frac{1}{2}\|\mathbf{w}\|^2 - \varepsilon \sum_{k=1}^{2} \sum_{i=1}^{l} (\alpha_{ki} + \alpha_{ki}^*) \\ + \sum_{i=1}^{l} m_{Y_i}(\alpha_{1i} - \alpha_{1i}^*) + \sum_{i=1}^{l} (m_{Y_i} - \alpha_{Y_i})(\alpha_{2i} - \alpha_{2i}^*) \end{cases}$$

$$\text{subject to} \quad \sum_{k=1}^{2} \sum_{i=1}^{l} (\alpha_{ki} - \alpha_{ki}^*) = 0 \ \text{and} \ \alpha_{ki}, \alpha_{ki}^* \in [0, C], \ k = 1, 2.$$

19

Table 1
Fuzzy input–fuzzy output data

$Y = (m_Y, \alpha_Y)$	$X = (m_X, \alpha_X)$
$(-1.6, 0.5)$	$(1, 0.5)$
$(-1.8, 0.5)$	$(3, 0.5)$
$(-1.0, 0.5)$	$(4, 0.5)$
$(1.2, 0.5)$	$(5.6, 0.8)$
$(2.2, 1.0)$	$(7.8, 0.8)$
$(6.8, 1.0)$	$(10.2, 0.8)$
$(10.0, 1.0)$	$(11.0, 1.0)$
$(10.0, 1.0)$	$(11.5, 1.0)$
$(10.0, 1.0)$	$(12.7, 1.0)$

Table 2

Model	ε	C	Coefficient	Residual sum
$M1$	0	500	$B = (-2.457, 0.071),\ w = 0.857$	95.314
$M2$	0.01	500	$b = -5.507,\ w = 1.239$	50.338

1   For this model b is exactly same as m_B for the model $H1$ and hence

$$\begin{cases} m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle - \varepsilon & \text{for } \alpha_{1i} \in (0, C), \\ m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m}_{\mathbf{X}_i} \rangle + \varepsilon & \text{for } \alpha_{1i}^* \in (0, C). \end{cases}$$

3   **Example 1.** From Gunn [11], data were constructed using the original x_i, y_i and symmetric fuzzified X_i, Y_i as shown in Table 1. Using the data in Table 1, the obtained results are shown in Table 2.
5 In this example ε and C are heuristically determined. In fact, the values of $C \geqslant 500$ give almost same results. However, results depend highly on the value of ε.

7   *2.2. Case of numerical inputs and fuzzy output*

  Suppose that data pairs $(\mathbf{x}_i, Y_i)$, $i = 1, 2, \ldots, l$ are observed where each $\mathbf{x}_i$ is d-vector of real
9 numbers and each $Y_i \in T(R)$. Let x_{ij} be element of $\mathbf{x}_i$. Then, we assume $x_{ij} \geqslant 0$ by simple translation of all vectors.
11   Let $\mathbf{W} = (W_1, W_2, \ldots, W_d)$, where $W_i = (m_{W_i}, \alpha_{W_i}, \beta_{W_i}), \alpha_{W_i}, \beta_{W_i} \geqslant 0$, $i = 1, \ldots, d$ and let $B = (m_B, \alpha_B, \beta_B), \alpha_B, \beta_B \geqslant 0$. We now consider the following model:

$$M3: f(\mathbf{x}) = B + \langle \mathbf{W}, \mathbf{x} \rangle, \quad B \in T(R), \quad \mathbf{W} \in T(R)^d, \quad \mathbf{x} \in R^d$$

$$= B + W_1 x_1 + W_2 x_2 + \cdots + W_d x_d,$$

13 where $T(R)^d$ is the set of d-vectors of triangular fuzzy numbers.
  We define $\|\mathbf{W}\|^2 = \|\mathbf{m_W}\|^2 + \|\mathbf{m_W} - \boldsymbol{\alpha}_W\|^2 + \|\mathbf{m_W} + \boldsymbol{\beta}_W\|^2$, where $\mathbf{m_W} = (m_{W_1}, \ldots, m_{W_d})$, $\boldsymbol{\alpha}_W = $
15 $(\alpha_{W_1}, \ldots, \alpha_{W_d})$ and $\boldsymbol{\beta}_W = (\beta_{W_1}, \ldots, \beta_{W_d})$. Then, we arrive at the following convex optimization

1   problem for model $M3$ as follows:

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{W}\|^2 + C\sum_{k=1}^{3}\sum_{i=1}^{l}(\xi_{ki} + \xi_{ki}^*)$$

$$\text{subject to} \quad \begin{cases} m_{Y_i} - (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B) \leqslant \varepsilon + \xi_{1i}, \\ (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B) - m_{Y_i} \leqslant \varepsilon + \xi_{1i}^*, \\ (m_{Y_i} - \alpha_{Y_i}) - (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B - \langle\boldsymbol{\alpha}_W, \mathbf{x}_i\rangle - \alpha_B) \leqslant \varepsilon + \xi_{2i}, \\ (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B - \langle\boldsymbol{\alpha}_W, \mathbf{x}_i\rangle - \alpha_B) - (m_{Y_i} - \alpha_{Y_i}) \leqslant \varepsilon + \xi_{2i}^*, \\ (m_{Y_i} + \beta_{Y_i}) - (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B + \langle\boldsymbol{\beta}_W, \mathbf{x}_i\rangle + \beta_B) \leqslant \varepsilon + \xi_{3i}, \\ (\langle\mathbf{m_W}, \mathbf{x}_i\rangle + m_B + \langle\boldsymbol{\beta}_W, \mathbf{x}_i\rangle + \beta_B) - (m_{Y_i} + \beta_{Y_i}) \leqslant \varepsilon + \xi_{3i}^*, \\ \xi_{ki}, \xi_{ki}^* \geqslant 0, \quad k = 1, 2, 3. \end{cases}$$

Then, constructing a Lagrange function and differentiating it with regard to $m_B, \alpha_B, \beta_B, \mathbf{m_W}, \boldsymbol{\alpha}_W, \boldsymbol{\beta}_W, \xi_{ki}$
3   and ξ_{ki}^*, $k = 1, 2$, we can get

$$\mathbf{m_W} = \sum_{i=1}^{l}(\alpha_{1i} - \alpha_{1i}^*)\mathbf{x}_i,$$

$$\boldsymbol{\alpha}_W = \sum_{i=1}^{l}[(\alpha_{1i} - \alpha_{1i}^*) - (\alpha_{2i} - \alpha_{2i}^*)]\mathbf{x}_i,$$

$$\boldsymbol{\beta}_W = \sum_{i=1}^{l}[(\alpha_{3i} - \alpha_{3i}^*) - (\alpha_{1i} - \alpha_{1i}^*)]\mathbf{x}_i$$

and in effect can derive the corresponding dual optimization problem for model $M3$ as follows:

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\sum_{k=1}^{3}\sum_{i,j=1}^{l}(\alpha_{ki} - \alpha_{ki}^*)(\alpha_{kj} - \alpha_{kj}^*)\langle\mathbf{x}_i, \mathbf{x}_j\rangle \\ -\varepsilon\sum_{k=1}^{3}\sum_{i=1}^{l}(\alpha_{ki} + \alpha_{ki}^*) + \sum_{i=1}^{l}m_{Y_i}(\alpha_{1i} - \alpha_{1i}^*) \\ +\sum_{i=1}^{l}(m_{Y_i} - \alpha_{Y_i})(\alpha_{2i} - \alpha_{2i}^*) + \sum_{i=1}^{l}(m_{Y_i} + \beta_{Y_i})(\alpha_{3i} - \alpha_{3i}^*) \end{cases}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{l}(\alpha_{ki} - \alpha_{ki}^*) = 0, \quad k = 1, 2, 3, \\ \alpha_{ki}, \alpha_{ki}^* \in [0, C], \quad k = 1, 2, 3, \; i = 1, \ldots, l. \end{cases}$$

5

1   Let us define $\alpha'_{W_k} = \max\{\alpha_{W_k}, 0\}$ and $\beta'_{W_k} = \max\{\beta_{W_k}, 0\}$ for $k = 1, \ldots, d$. Now we define two vectors $\boldsymbol{\alpha}'_W = (\alpha'_{W_1}, \ldots, \alpha'_{W_d})$ and $\boldsymbol{\beta}'_W = (\beta'_{W_1}, \ldots, \beta'_{W_d})$. Then we have

$$f(\mathbf{x}) = B + \langle \mathbf{W}', \mathbf{x} \rangle$$

$$= B + (\langle \mathbf{m_W}, \mathbf{x} \rangle, \langle \boldsymbol{\alpha}'_W, \mathbf{x} \rangle, \langle \boldsymbol{\beta}'_W, \mathbf{x} \rangle),$$

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

If $\alpha_{W_i} \geqslant 0$ and $\beta_{W_i} \geqslant 0$ for $i = 1, 2, \ldots, d$, then we have

$$f(\mathbf{x}) = B + \left(\sum_{i=1}^{l} (\alpha_{1i} - \alpha_{1i}^*) \langle \mathbf{x}_i, \mathbf{x} \rangle, \sum_{i=1}^{l} [(\alpha_{1i} - \alpha_{1i}^*) - (\alpha_{2i} - \alpha_{2i}^*)] \langle \mathbf{x}_i, \mathbf{x} \rangle, \right.$$

$$\left. \sum_{i=1}^{l} [(\alpha_{3i} - \alpha_{3i}^*) - (\alpha_{1i} - \alpha_{1i}^*)] \langle \mathbf{x}_i, \mathbf{x} \rangle \right).$$

5   Now we need to find m_B, α_B and β_B. By KKT conditions, we can compute m_B as follows:

$$\begin{cases} \alpha_{1i}(\varepsilon + \xi_{1i} - m_{Y_i} + \langle \mathbf{w}, \mathbf{m_{X_i}} \rangle + m_B) = 0, \\ \alpha_{1i}^*(\varepsilon + \xi_{1i}^* + m_{Y_i} - \langle \mathbf{w}, \mathbf{m_{X_i}} \rangle - m_B) = 0, \\ (C - \alpha_{1i})\xi_{1i} = 0, \ (C - \alpha_{1i}^*)\xi_{1i}^* = 0 \end{cases}$$

and hence

$$\begin{cases} m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m_{X_i}} \rangle - \varepsilon & \text{for } \alpha_{1i} \in (0, C), \\ m_B = m_{Y_i} - \langle \mathbf{w}, \mathbf{m_{X_i}} \rangle + \varepsilon & \text{for } \alpha_{1i}^* \in (0, C). \end{cases}$$

7   To find α_B, β_B we need to solve the optimization problem given below

$$\underset{\alpha_B, \beta_B \geqslant 0}{\text{minimize}} \sum_{i=1}^{l} |m_{Y_i} - \alpha_{Y_i} - \langle \mathbf{m_W} - \boldsymbol{\alpha}'_W, \mathbf{x}_i \rangle - m_B + \alpha_B|_\varepsilon$$

$$+ \sum_{i=1}^{l} |m_{Y_i} + \beta_{Y_i} - \langle \mathbf{m_W} + \boldsymbol{\beta}'_W, \mathbf{x}_i \rangle - m_B - \beta_B|_\varepsilon.$$

9   **Example 2.** From Gunn [11], data were constructed using the original x_i, y_i and symmetric fuzzified Y_i as shown in Table 3. Using this data set, the obtained results are shown in Table 4.

11   In this example $\varepsilon = 0.001$ and $C = 500$ are used but it is heuristically determined. In fact, the values of ε near 0 and $C \geqslant 500$ give almost same results. As seen from Table 4, we think this model

13 is not appropriate for this particular data set because residual sum is too big.

Table 3
Real input–fuzzy output data

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

Table 4

Model	ε	C	Coefficient	Residual sum
$M3$	0.001	500	$B = (-5.451, 0, 0), \ W = (1.217, 0.508, 0.508)$	269.514

3. SV fuzzy nonlinear regression

There have been only a few articles for fuzzy nonlinear regression. What researchers in fuzzy nonlinear regression were concerned with was data of the form with crisp inputs and fuzzy output. Some papers, for example [1,2,5], were concerned with the data set with fuzzy inputs and fuzzy output. However, we think those fuzzy nonlinear regression methods look somewhat unrealistic and treat the estimation procedures of some particular models such as linear, polynomial, exponential and logarithmic. In this paper, we treat fuzzy nonlinear regression for data of the form with numerical inputs and fuzzy output, without assuming the underlying model function. In other words, we extend the linear model $M3$ to the nonlinear case Fig. 1.

To do this, we will use the idea of SVM for crisp nonlinear regression [11,13,16–18]. The basic idea is that a nonlinear regression function is achieved by simply preprocessing input patterns $\mathbf{x}_i$ by a map $\mathbf{\Phi} : R^d \to \mathscr{F}$ into some feature space $\mathscr{F}$ and then applying the standard ridge regression learning algorithm. Notice that the only way in which the data appears in algorithm for the model $M3$ is in the form of inner products $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. The algorithm would only depend on the data through inner products in $\mathscr{F}$, i.e. on functions of the form $\langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}_j) \rangle$. Hence it suffices to know and use $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}_j) \rangle$ instead of $\mathbf{\Phi}(\cdot)$ explicitly. The well used kernels for regression problem are given below

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

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

$K(\mathbf{x}, \mathbf{y}) = \tanh(\kappa \langle \mathbf{x}, \mathbf{y} \rangle + \theta)$: Hyperbolic tangent kernel.
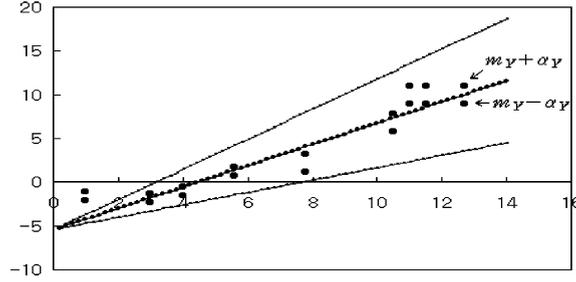
Fig. 1. The SVM fuzzy regression model $M2$ with $\alpha_X = 0.75$.

1   Hence, by replacing $\langle \mathbf{x}, \mathbf{y} \rangle$ with $K(\mathbf{x}, \mathbf{y})$ we obtain the following dual optimization problem:

$$\text{maximize} \begin{cases} -\frac{1}{2} \sum_{k=1}^{3} \sum_{i,j=1}^{l} (\alpha_{ki} - \alpha_{ki}^*)(\alpha_{kj} - \alpha_{kj}^*) K(\mathbf{x}_i, \mathbf{x}_j), \\[2mm] -\varepsilon \sum_{k=1}^{3} \sum_{i=1}^{l} (\alpha_{ki} + \alpha_{ki}^*) + \sum_{i=1}^{l} m_{Y_i}(\alpha_{1i} - \alpha_{1i}^*), \\[2mm] + \sum_{i=1}^{l} (m_{Y_i} - \alpha_{Y_i})(\alpha_{2i} - \alpha_{2i}^*) + \sum_{i=1}^{l} (m_{Y_i} + \beta_{Y_i})(\alpha_{3i} - \alpha_{3i}^*). \end{cases}$$

Here we should notice that the constraints are unchanged

$$\sum_{i=1}^{l} (\alpha_{ki} - \alpha_{ki}^*) = 0, \quad k = 1, 2, 3,$$

$$\alpha_{ki}, \alpha_{ki}^* \in [0, C], \quad k = 1, 2, 3.$$

3   If $\sum_{i=1}^{l}[(\alpha_{1i} - \alpha_{1i}^*) - (\alpha_{2i} - \alpha_{2i}^*)]K(\mathbf{x}_i, \mathbf{x}_j) \geqslant 0$ and $\sum_{i=1}^{l}[(\alpha_{3i} - \alpha_{3i}^*) - (\alpha_{1i} - \alpha_{1i}^*)]K(\mathbf{x}_i, \mathbf{x}_j) \geqslant 0$, $j = 1, \ldots, l$, then we have

$$f(\mathbf{x}) = B + \left(\sum_{i=1}^{l} (\alpha_{1i} - \alpha_{1i}^*) K(\mathbf{x}_i, \mathbf{x}), \sum_{i=1}^{l} [(\alpha_{1i} - \alpha_{1i}^*) - (\alpha_{2i} - \alpha_{2i}^*)] K(\mathbf{x}_i, \mathbf{x}), \right.$$
$$\left. \sum_{i=1}^{l} [(\alpha_{3i} - \alpha_{3i}^*) - (\alpha_{1i} - \alpha_{1i}^*)] K(\mathbf{x}_i, \mathbf{x}) \right).$$

5   Now we need to find m_B, α_B and β_B. By KKT conditions, we can compute m_B as follows:

$$\begin{cases} m_B = m_{Y_i} - \sum_{j=1}^{l} (\alpha_{1j} - \alpha_{1j}^*) K(\mathbf{x}_j, \mathbf{x}_i) - \varepsilon \quad \text{for } \alpha_{1i} \in (0, C), \\[2mm] m_B = m_{Y_i} - \sum_{j=1}^{l} (\alpha_{1j} - \alpha_{1j}^*) K(\mathbf{x}_j, \mathbf{x}_i) + \varepsilon \quad \text{for } \alpha_{1i}^* \in (0, C), \end{cases}$$
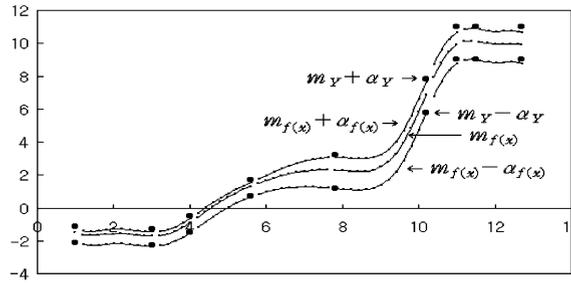
Fig. 2. The SV fuzzy nonlinear regression model.

1 To find α_B, β_B we need to solve the optimization problem given below

$$\underset{\alpha_B, \beta_B \geqslant 0}{\text{minimize}} \sum_{i=1}^{l} \left| m_{Y_i} - \alpha_{Y_i} - \sum_{j=1}^{l} (\alpha_{2j} - \alpha_{2j}^*) K(\mathbf{x}_j, \mathbf{x}_i) - m_B + \alpha_B \right|_\varepsilon$$

$$+ \sum_{i=1}^{l} \left| m_{Y_i} + \beta_{Y_i} - \sum_{j=1}^{l} (\alpha_{3j} - \alpha_{3j}^*) K(\mathbf{x}_j, \mathbf{x}_i) - m_B - \beta_B \right|_\varepsilon .$$

3 **Example 3.** We now apply fuzzy nonlinear regression model to the data in Table 3. According to Gunn [11], the nonlinear regression model is appropriate for the original crisp data of Table 3. When
5 we apply fuzzy nonlinear regression model to this data set, we have the residual sum 0.833 and bias term $B = (3.028, 0.875, 0.515)$. Hence we can recognize fuzzy nonlinear model is more appropriate
7 than the linear model $M3$. For this data set we use Gaussian kernel with $\sigma = 1.0$ and $C = 500$. These parameters are determined in the heuristic way Fig. 2.

9 **4. Conclusion**

In this paper, we have presented a SVM strategy for fuzzy multivariable linear and nonlinear
11 regressions. The experimental results show that the SVM algorithm for fuzzy regression models derives the satisfying solutions and is an attractive approach to modeling fuzzy data. The algorithm
13 combines generalization control with a technique to address the curse of dimensionality, which makes SVM solution not to depend directly on the dimensionality of the input space. The main formulation
15 results in a global quadratic optimization problem with box constraints. However, this is not a computationally expensive way. The examples we considered here are too simple. SVM probably
17 has greatest use when the dimensionality of the input space is high. We thus need to conduct real-life examples with several inputs.
19 There have been some papers treat fuzzy nonlinear regression models. They usually assume the underlying model functions even for data of the form with numerical inputs and fuzzy out-
21 put. The proposed algorithm here is model-free method in the sense that we do not have to assume

1 the underlying model function. This model-free method turned out to be a promising method which has been attempted to treat fuzzy nonlinear regression model with numerical inputs and fuzzy output.
3 The main difference between our SVM approach and the nonlinear approaches by Buckley et al. [2,3] and Celmins [5] is not crisp input–fuzzy output versus fuzzy input–fuzzy output, but model-free
5 versus model-dependent.

Here we use kernel parameter σ and control parameter C determined in a heuristic way. The
7 obvious question that arises is which are the best for a particular problem? Hence we need model selection method to determine these parameters.

References

11   [1] B.E. Boser, I.M. Guyon, V. Vapnik, A training algorithm for optimal margin classifiers, in: Fifth Annu. Workshop on Computational Learning Theory, Pittsburgh, ACM, New York, 1992.
13   [2] J. Buckley, T. Feuring, Linear and non-linear fuzzy regression: evolutionary algorithm solutions, Fuzzy Sets and Systems 112 (2000) 381–394.
15   [3] J. Buckley, T. Feuring, Y. Hayashi, Multivariate non-linear fuzzy regression: an evolutionary algorithm approach, Internat. J. Uncertainty, Fuzziness Knowledge-Based Systems 7 (1999) 83–98.
17   [4] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Knowledge Discovery and Data Mining, 1998.
19   [5] A. Celmins, A practical approach to nonlinear fuzzy regression, SIAM J. Sci. Statist. Comput. 12 (3) (1991) 521–546.
21   [6] C. Cortes, V. Vapnik, Support vector networks, Mach. Learning 20 (1995) 273–297.
     [7] P. Diamond, Least squares fitting of several fuzzy variables, in: J.C. Bezdek (Ed.), Analysis of Fuzzy Information,
23   CRC Press, Tokyo, 1987, pp. 329–331.
     [8] P. Diamond, Fuzzy least squares, Inform. Sci. 46 (1988) 141–157.
25   [9] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge,
27   MA, 1997, pp. 155–161.
   [10] D. Dubois, H. Prade, Theory and Applications, Fuzzy Sets and Systems, Academic Press, New York, 1980.
29   [11] S. Gunn, Support Vector Machines for Classification and Regression, ISIS Tech. Report, University of Southampton, 1998.
31   [12] J. Kacprzyk, M. Fedrizzi, Fuzzy Regression Analysis, Physica-Verlag, Heidelberg, 1992.
   [13] A.J. Smola, B. Scholkopf, A tutorial on support vector regression, NeuroCOLT2 Tech. Report, NeuroCOLT, 1998.
33   [14] H. Tanaka, S. Uejima, K. Asia, Fuzzy linear regression model. Internat. Congress on Applied Systems Research and Cybernetics, Aculpoco, Mexico, 1980.
35   [15] H. Tanaka, S. Uejima, K. Asia, Linear regression analysis with Fuzzy model, IEEE Trans. Systems Man. Cybernet. 12 (6) (1982) 903–907.
37   [16] V. Vapnik, The Nature of Statistical Learning Theory, Springer, Berlin, 1995.
   [17] V. Vapnik, Statistical Learning Theory, Springer, Berlin, 1998.
39   [18] V. Vapnik, S. Golowich, A. Smola, Support vector method for function approximation, regression estimation, and signal processing, Adv. Neural Inform. Process. Systems 9 (1996) 281–287.