Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 2, pp. 507~513

Least-Squares Support Vector Machine for Regression Model with Crisp Inputs-Gaussian Fuzzy Output

Changha Hwang[1]

Abstract

Least-squares support vector machine (LS-SVM) has been very successful in pattern recognition and function estimation problems for crisp data. In this paper, we propose LS-SVM approach to evaluating fuzzy regression model with multiple crisp inputs and a Gaussian fuzzy output. The proposed algorithm here is model-free method in the sense that we do not need assume the underlying model function. Experimental result is then presented which indicate the performance of this algorithm.

Keywords : Fuzzy, least squares, regression, support vector machine

1. Introduction

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.(1982) in a linear system, is based on the extension principle. Similar to traditional least-squares, Diamond(1988) defined a distance on a triangular fuzzy number space to measure the best fit for the regression model to observed data, and then derived regression parameters based on the distance. Xu(1991) presented a distance on a fuzzy number space by the integral of distance of every level set. Xu and Li(2001) discussed the problem of multidimensional least-squares fitting. Using Xu's distance on a fuzzy number space, Hong et al.(2004) proposed a ridge estimation method of fuzzy regression models with multiple crisp inputs and a Gaussian fuzzy output. Yang and Ko(1996) derived the distance for two Gaussian fuzzy numbers, which is different from Xu's. In this paper, using Yang and Ko's

---

1) Professor, Dept. of Statistical Information, Catholic University of Daegu.
E-mail : chhwang@cu.ac.kr

distance, we propose least-squares support vector machine(LS-SVM) approach to evaluating fuzzy regression models with multiple crisp inputs and a Gaussian fuzzy output. The details of LS-SVM were illustrated in Suykens et al.(2002). Another direction of SVM approach to fuzzy regression was illustrated in Hong and Hwang(2003).

2. Distance of Gaussian Fuzzy Numbers

In this section, we review linear operations and distance of Gaussian fuzzy number, which are needed to illustrate LS-SVM for fuzzy regression models. These are taken from Xu and Li(2001) and Yang and Ko(1996). We first illustrate Gaussian fuzzy number. Let R be the real number set.

Definition 1. A fuzzy number $\widetilde{A}$ is said to be a Gaussian fuzzy number if its membership function can be expressed as

$$\widetilde{A}(x) = \exp\left[-\left(\frac{x-a}{\sigma}\right)^2\right], \quad x \in R \ (\sigma > 0)$$

and we write $\widetilde{A} = (a, \sigma)$. Denote by N the set of all Gaussian fuzzy numbers.

We next illustrate Zadeh's extension principle for the linear operations of Gaussian fuzzy number.

Theorem 1. Let $\widetilde{A} = (a, \sigma)$ and $\widetilde{B} = (b, \tau)$ be Gaussian fuzzy numbers. Then

(1) $t\widetilde{A} = (ta, t\sigma) \ (t > 0)$,

(2) $\widetilde{A} + \widetilde{B} = (a + b, \sigma + \tau)$.

Theorem 2. Let $\widetilde{A} = (a, \sigma)$ and $\widetilde{B} = (b, \tau)$ be Gaussian fuzzy numbers. Then the distance $d^2(\widetilde{A}, \widetilde{B})$ for any two Gaussian fuzzy numbers is defined as follows:

$$\begin{aligned} d^2(\widetilde{A}, \widetilde{B}) &= (a-b)^2 + \left(\left(a - \frac{\sqrt{\pi}}{2}\sigma\right) - \left(b - \frac{\sqrt{\pi}}{2}\tau\right)\right)^2 + \left(\left(a + \frac{\sqrt{\pi}}{2}\sigma\right) - \left(b + \frac{\sqrt{\pi}}{2}\tau\right)\right)^2 \\ &= 3(a-b)^2 + \frac{\pi}{2}(\sigma - \tau)^2. \end{aligned}$$

Note that the distance $d(\widetilde{A}, \widetilde{B})$ here is similar to the squared Hausdorff-like distance proposed by Albrecht(1992).

3. LS-SVM for Fuzzy Regression

In this section, we will modify the underlying idea of LS-SVM for the purpose of deriving the convex optimization problems for fuzzy linear and nonlinear regression models with multiple crisp inputs and a Gaussian fuzzy output. The basic idea of LS-SVM gives computational efficiency in finding solutions of fuzzy regression models particularly for multivariate case.

Suppose that we are given training data $\{(\boldsymbol{x}_i, \widetilde{Y}_i),\ i=1,\cdots,m\} \subset R^n \times N$, where $\boldsymbol{x}_i \in R^n$ and $\widetilde{Y}_i = (y_i, s_i) \in N$. Xu and Li(2001) considered the following model

$$\widetilde{Y} = \widetilde{A}_0 + \widetilde{A}_1 x_1 + \cdots + \widetilde{A}_n x_n, \quad \widetilde{A}_j = (a_j, \sigma_j),\ j=0,1,\cdots,n \tag{1}$$

and considered the least-squares optimization problem as follows:

$$\min\ M(\widetilde{A}_0, \widetilde{A}_1, \cdots, \widetilde{A}_n) = \sum_{i=1}^{m} d^2(\widetilde{A}_0 + \widetilde{A}_1 x_{i1} + \cdots + \widetilde{A}_n x_{in}, \widetilde{Y}_i). \tag{2}$$

Therefore, according to Theorem 2, the least-squares problem (2) can be rewritten as

$$\min\ M(\widetilde{A}_0, \widetilde{A}_1, \cdots, \widetilde{A}_n) = 3\sum_{i=1}^{m} (\boldsymbol{a}^t \boldsymbol{x}_i + a_0 - y_i)^2 + \frac{\pi}{2}\sum_{i=1}^{m} (\boldsymbol{\sigma}^t \boldsymbol{x}_i + \sigma_0 - s_i)^2, \tag{3}$$

where $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{in})^t$, $\boldsymbol{a} = (a_1, \cdots, a_n)^t$ and $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_n)^t$. Here, the superscript t denotes the transpose of matrix. Throughout this paper, we assume that $x_{ij} > 0$ $(i=1,\cdots,m,\ j=1,\cdots,n)$ by simple translation of all data.

We now modify this idea for the purpose of deriving LS-SVM algorithm for fuzzy linear and nonlinear regression model. We define $\widetilde{\boldsymbol{A}} = (\widetilde{A}_1, \cdots, \widetilde{A}_n)^t$, and then we have $\| \widetilde{\boldsymbol{A}} \|^2 = 3\|\boldsymbol{a}\|^2 + \frac{\pi}{2}\|\boldsymbol{\sigma}\|^2$. Hence, we arrive at the following convex optimization problem for model (1) as follows:

$$\min\ \frac{1}{2}\|\widetilde{\boldsymbol{A}}\|^2 + \frac{3C}{2}\sum_{i=1}^{m}\xi_{1i}^2 + \frac{\pi C}{4}\sum_{i=1}^{m}\xi_{2i}^2,$$

under the equality constraints

$$y_i = \boldsymbol{a}^t\boldsymbol{x}_i + a_0 + \xi_{1i},\quad s_i = \boldsymbol{\sigma}^t\boldsymbol{x}_i + \sigma_0 + \xi_{2i},\quad i=1,\cdots,m.$$

Introducing Lagrange multipliers α_{1i} and α_{2i}, $i=1,\cdots,m$, we construct a Lagrange function as follows:

$$L = \frac{1}{2}\|\widetilde{\boldsymbol{A}}\|^2 + \frac{3C}{2}\sum_{i=1}^{m}\xi_{1i}^2 + \frac{\pi C}{4}\sum_{i=1}^{m}\xi_{2i}^2 - \sum_{i=1}^{m}\alpha_{1i}(\boldsymbol{a}^t\boldsymbol{x}_i + a_0 + \xi_{1i} - y_i)$$
$$- \sum_{i=1}^{m}\alpha_{2i}(\boldsymbol{\sigma}^t\boldsymbol{x}_i + \sigma_0 + \xi_{2i} - s_i).$$

Then, the conditions for optimality are given by

$$\frac{\partial L}{\partial \boldsymbol{a}} = 0 \rightarrow \boldsymbol{a} = \frac{1}{3}\sum_{i=1}^{m}\alpha_{1i}\boldsymbol{x}_i$$
$$\frac{\partial L}{\partial a_0} = 0 \rightarrow \sum_{i=1}^{m}\alpha_{1i} = 0$$

$$\frac{\partial L}{\partial \sigma} = 0 \rightarrow \sigma = \frac{2}{\pi} \sum_{i=1}^{m} \alpha_{2i} x_i$$
$$\frac{\partial L}{\partial \sigma_0} = 0 \rightarrow \sum_{i=1}^{m} \alpha_{2i} = 0$$
$$\frac{\partial L}{\partial \xi_{1i}} = 0 \rightarrow \xi_{1i} = \frac{1}{3C} \alpha_{1i}, \ i = 1, \cdots, m$$
$$\frac{\partial L}{\partial \xi_{2i}} = 0 \rightarrow \xi_{2i} = \frac{2}{\pi C} \alpha_{2i}, \ i = 1, \cdots, m$$
$$\frac{\partial L}{\partial \alpha_{1i}} = 0 \rightarrow a^t x_i + a_0 + \xi_{1i} = y_i, \ i = 1, \cdots, m$$
$$\frac{\partial L}{\partial \alpha_{2i}} = 0 \rightarrow \sigma^t x_i + \sigma_0 + \xi_{2i} = s_i, \ i = 1, \cdots, m$$

with solutions

$$a = \frac{1}{3} \sum_{i=1}^{m} \alpha_{1i} x_i, \quad \sigma = \frac{2}{\pi} \sum_{i=1}^{m} \alpha_{2i} x_i$$

and

$$\begin{bmatrix} 0 & 0 & \mathbf{1}^t & \mathbf{0}^t \\ 0 & 0 & \mathbf{0}^t & \mathbf{1}^t \\ 1 & 0 & \frac{1}{3}\Omega + \frac{1}{3C} I & O \\ 0 & 1 & O & \frac{2}{\pi}\Omega + \frac{2}{\pi C} I \end{bmatrix} \begin{bmatrix} a_0 \\ \sigma_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ y \\ s \end{bmatrix}$$

with $\alpha_1 = (\alpha_{11}, \cdots, \alpha_{1m})^t$, $\alpha_2 = (\alpha_{21}, \cdots, \alpha_{2m})^t$, $\mathbf{0} = (0, \cdots, 0)^t$, $\mathbf{1} = (1, \cdots, 1)^t$, $m \times m$ zero matrix O, $m \times m$ identity matrix I, $m \times m$ matrix Ω of $\Omega_{ij} = x_i^t x_j$, $y = (y_1, \cdots, y_m)^t$ and $s = (s_1, \cdots, s_m)^t$.

Hence, the prediction $\tilde{Y} = (\hat{y}, \hat{s})$ given by the LS-SVM procedure on the new unlabeled example x is

$$(\frac{1}{3} \sum_{i=1}^{m} \alpha_{1i} x_i^t x + a_0, \ \frac{2}{\pi} \sum_{i=1}^{m} \alpha_{2i} x_i^t x + \sigma_0).$$

Similar to Xu and Li(2001), the goodness of fit of observed value (y_i, s_i) and the estimated value $(\hat{y}_i, \hat{s}_i)$ is defined by

$$\exp\left[- \left(\frac{\frac{1}{3} \sum_{i=1}^{m} \alpha_{1i} x_i^t x + a_0 - y_i}{\frac{2}{\pi} \sum_{i=1}^{m} \alpha_{2i} x_i^t x + \sigma_0 + s_i} \right) \right]. \tag{4}$$

See for details Xu and Li(2001).

Notice that the condition $\text{Rank}(X) = n+1$ should hold in Xu and Li(2001). Here, X denotes the design matrix consisting of m input vectors. If $\text{Rank}(X)$

$< n+1$, or in other situations where numerical stability problems occur, we can use LS-SVM.

Next, we will study LS-SVM to be used in estimating fuzzy nonlinear regression model. In contrast to fuzzy linear regression, there have been only a few articles on fuzzy nonlinear regression. In this paper we treat fuzzy nonlinear regression, without assuming the underlying model function. This could be achieved by simply preprocessing input patterns x_i by a map $\Phi: R^n \to E$ into some feature space E and then applying LS-SVM linear regression algorithm. The details are illustrated in Suykens et al.(2002).

First notice that the only way in which the data appears in the training problem is in the form of dot products $x_i^t x_j$. The algorithm would only depend on the data through dot products in E, i.e. on functions of the form $\Phi(x_i)^t \Phi(x_j)$. Hence it suffices to know and use $K(x_i, x_j) = \Phi(x_i)^t \Phi(x_j)$ instead of $\Phi(\cdot)$ explicitly. The only difference between LS-SVMs for linear and nonlinear function estimations is the use of mapping function Φ. The well used kernels for regression problem are given below.

$$K(x, y) = (x^t y + 1)^p, \quad K(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}}.$$

Here, p and σ^2 are kernel parameters. In final, the fuzzy nonlinear LS-SVM regression solution is given by

$$\left(\frac{1}{3} \sum_{i=1}^{m} \alpha_{1i} K(x_i, x) + a_0, \ \frac{2}{\pi} \sum_{i=1}^{m} \alpha_{2i} K(x_i, x) + \sigma_0\right).$$

Similar to the linear case (4), we can define the goodness of fit for fuzzy nonlinear regression.

4. Numerical Example and Conclusions

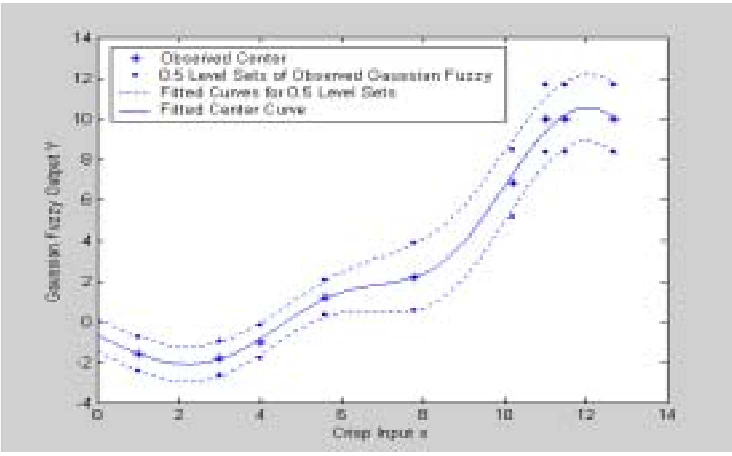
In this section, one example is considered to verify the effectiveness of the proposed LS-SVM procedure for Xu and Li's fuzzy model. We consider nonlinear regression analysis only, since the linear case is analogous. As an illustration of this algorithm, we consider a toy problem involving one crisp input. The data is taken from Gunn(1998), except s_i. The values of s_i are assumed by the author. The data are given in Table 1.

According to Gunn(1998), the nonlinear regression model is appropriate for the original crisp data. Hence we apply fuzzy nonlinear regression model to this data set. Here, we use Gaussian kernel for estimating nonlinear regression model. We have used leave-one-out(LOO) cross-validation based on the sum of squares in the optimization problem (3) to determine an optimal combination of C and σ, which are $C = 140$ and $\sigma = 2.6$.

The goodness of fit of observed value (y_i, s_i) and the estimated value $(\hat{y}_i, \hat{s}_i)$ is reported in Table 1. The goodness of fit of every (y_i, s_i) and $(\hat{y}_i, \hat{s}_i)$ are all greater than 0.9. This implies that the fitted result of our algorithm very good.

<**Table 1**> Data, Estimated values and Goodness of Fit

x_i	(y_i, s_i)	$(\hat{y}_i, \hat{s}_i)$	Goodness of Fit
1	(−1.6, 1.0)	(−1.5752, 1.0035)	0.9998
3	(−1.8, 1.0)	(−1.8617, 1.0253)	0.9991
4	(−1.0, 1.0)	(−0.8625, 0.9401)	0.9950
5.6	(1.2, 1.0)	(1.1128, 1.0655)	0.9982
7.8	(2.2, 2.0)	(2.1866, 1.9483)	1.0000
10.2	(6.8, 2.0)	(7.2221, 2.0502)	0.9892
11.0	(10.0, 2.0)	(9.3373, 1.9868)	0.9728
11.5	(10.0, 2.0)	(10.1871, 1.9755)	0.9978
12.7	(10.0, 2.0)	(10.0533, 2.0048)	0.9998



<**Fig. 1**> Fuzzy nonlinear regression model

Fig. 1 illustrates results for fuzzy nonlinear model. The asterisks represent the observed center values. The dots represent 0.5 level sets of observed Gaussian fuzzy numbers. The solid curve explains the fitted regression curve for center. The two dotted curves explain the fitted curves for 0.5 level sets of Gaussian fuzzy numbers. As seen from Figure 1, the proposed algorithm performs well for fuzzy nonlinear regression model.

From the example we can realize that the proposed algorithms derive the satisfying solutions and are an attractive approach to modelling fuzzy data. The main formulation results in solving a simple matrix inversion problem. Hence, this is not a computationally expensive way.

References

1. Albrecht, M. (1992). Approximation of functional relationships to fuzzy observations, *Fuzzy Sets and Systems*, 49, 301-305.
2. Diamond, P. (1988). Fuzzy least squares, *Information Sciences*, 46, 141-157.
3. Gunn, S. (1998). Support vector machines for classification and regression, ISIS Technical Report, University of Southampton.
4. Hong, D., Hwang, C. (2003). Support vector fuzzy regression machines, *Fuzzy Sets and Systems*, 138, 271-281.
5. Hong, D., Hwang, C., Ahn, C. (2004). Ridge estimation for regression models with crisp inputs and Gaussian fuzzy output, *Fuzzy Sets and Systems*, 142, 307-319.
6. Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J. (2002). Least Squares Support Vector Machines, World Scientific, Singapore.
7. Tanaka, H., Uejima, S., Asia, K. (1982). Linear regression analysis with fuzzy model, *IEEE Trans. Man. Cybernet,* 12, 903-907.
8. Xu, R. (1991). A linear regression model in fuzzy environment, *Adv. Modelling Simulation*, 27, 31-40.
9. Xu, R., Li, C. (2001). Multidimensional least-squares fitting with a fuzzy model, A linear regression model in fuzzy environment, *Fuzzy Sets and Systems*, 119, 215-223.
10. Yang, M., Ko, C. (1996). On a class of fuzzy c-numbers clustering procedures for fuzzy data, *Fuzzy Sets and Systems*, 84, 49-60.

[received date : Dec. 2003 , accepted date : May. 2004]