

A Fuzzy Model of Support Vector Regression Machine

Pei-Yi Hao and Jung-Hsien Chiang

Abstract

Fuzziness must be considered in systems where available information is uncertain. A model of such a vague phenomenon might be represented as a fuzzy system equation which can be described by the fuzzy functions defined by Zadeh's extension principle. In this paper, we incorporate the concept of fuzzy set theory into the support vector machine (SVM). This integration preserves the benefits of SVM regression model and fuzzy regression model, where the SVM learning theory characterizes properties of learning machines which enable them to generalize well and the fuzzy set theory provides an effective means of capturing the approximate, inexact nature of real world.

Keywords: Support Vector Machines (SVMs), Support Vector Regression, Fuzzy Regression.

1. Introduction

In modeling some systems where available information is uncertain, we must deal with a fuzzy structure of the system considered. This structure is represented as a fuzzy function whose parameters are given by fuzzy sets. The fuzzy functions are defined by Zadeh's extension principle [5], [13], [14]. Considering a fuzzy function as a model of fuzzy structure of fuzzy system, a fuzzy regression analysis is formulated [10], [11]. The fuzzy parameters of the fuzzy model obtained represent a possibility distribution which corresponds to the fuzziness of the system. This fuzzy regression model might be very useful for finding a fuzzy structure in an evaluation system.

The Support Vector Machines (SVMs), developed at AT&T Bell Laboratories by Vapnik and co-works [3], [12], have been very successful in pattern classification and function estimation problems for crisp data. It is based on the idea of structural risk minimization, which shows that the generalization error is bounded by the

sum of the training error and a term depending on the Vapnik–Chervonenkis (VC) dimension. By minimizing this bound, high generalization performance can be achieved. A comprehensive tutorial on SVM classifier has been published by Burges [1]. In regression and time series prediction applications, excellent performances were also obtained [4]. Lin et al. [8] and Huang et al. [7] first introduced the use of fuzzy set theory for SVM classification problems. Whereas Chiang et al. [2] applied the SVM theory for the fuzzy rules based modeling.

In this paper, we incorporated the concept of fuzzy set theory into the SVM regression model. The parameters to be identified in SVM regression model, such as the components of weight vector and bias term, are set to be fuzzy numbers. Besides, the desired outputs in training samples are also fuzzy number. Incorporating the concept of fuzzy set theory into the SVM regression preserves the benefits of SVM regression and fuzzy regression, where the VC theory characterizes properties of learning machines which enable them to generalize well in the unseen data. Finally, the fuzzy SVM regression model might be very useful for finding a fuzzy structure in an evaluation system where available information is inexact.

2. SVM Regression Model

Suppose we are given a training data set $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\} \subset \mathfrak{X} \times R$, where \mathfrak{X} denotes the space of input patterns—for instance, R^k . In ε -SVM regression [4,12], the goal is to find a function $f(\mathbf{x})$ that has at most ε deviation from the actually obtained targets y_i for all the training data. In other words, we do not care about error as long as they are less than ε , but will not accept any deviation larger than ε . An ε -insensitive loss function

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon, \\ |\xi| - \varepsilon & \text{otherwise,} \end{cases} \quad (1)$$

is used so that the error is penalized only if it is outside the ε -tube. Figure 1 depicts this situation graphically. To make the SVM regression nonlinear, this could be achieved by simply mapping the training patterns \mathbf{x}_i by a nonlinear transform $\Phi: \mathfrak{X} \rightarrow F$ into some high dimensional feature space F . A simple example of the

Corresponding Author: J.-H. Chiang is with the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan, R.O.C.

E-mail: jchiang@mail.ncku.edu.tw

Manuscript received 20 Jun. 2006; revised 11 Oct. 2006; accepted 12 Mar. 2007.

nonlinear transform Φ is the polynomial transform function:

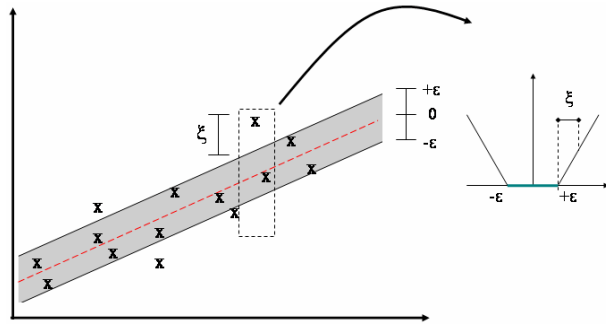


Fig. 1. The epsilon insensitive loss setting corresponds for a linear SV regression machine.

$$\begin{aligned} \mathbf{x} &= (x_1, x_2) \rightarrow \\ \Phi(\mathbf{x}) &= (\sqrt{3}x_1, \sqrt{3}x_2, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{6}x_1x_2, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, x_1^3, x_2^3) \end{aligned} \quad (2)$$

where the degree of the polynomial transform function is 3, and the dimension in input space \mathfrak{X} and feature space F is 2 and 9, respectively. A best fitting function

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \quad (3)$$

is estimated in feature space F , where “.” denotes the dot product in the feature space F . To avoid overfitting in the very high-dimensional feature space, one should add a capacity control term, which in the SVM case results to be $\|\mathbf{w}\|^2$. Formally, the SVM regression model can be written as a convex optimization problem by requiring:

$$\begin{aligned} \text{minimize}_{\mathbf{w}, b, \xi_i, \xi_i^*} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{subject to} \quad & y_i - (\mathbf{w}'\Phi(\mathbf{x}_i) + b) \leq \varepsilon - \xi_i \\ & (\mathbf{w}'\Phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon - \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \quad \forall i. \end{aligned} \quad (4)$$

The constant $C > 0$ determines the trade off between the complexity of $f(\mathbf{x})$ and the amount up to which deviations larger than ε are tolerated. What makes SVM regression attractive is we can estimate a linear function in the feature space, while it is a nonlinear function estimated in the original space, as shown in Fig. 2. Besides, the regression task was achieved by solving a convex programming with linear constraints; in other words, it has a unique solution.

However, the size of ε -tube is a pre-defined constant. The selection of a parameter ε may seriously affect the modeling performance. Besides, the ε -insensitive zone in the SVM regression model has a tube (or slab) shape. Namely, all training data points are equally treated during the training of SVM regression model, and are penalized only if they are outside the ε -tube. In many real-world applications, the effects of the training points are different. We would require that the precise training

points must be regressed correctly, and would allow more errors on imprecise training points.

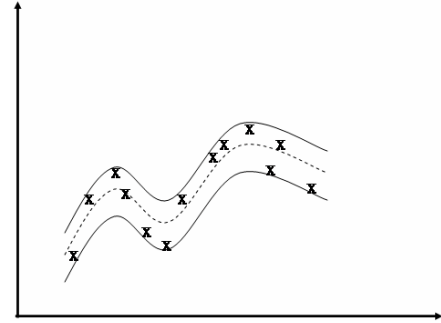


Fig. 2. The epsilon insensitive loss setting corresponds for a nonlinear SV regression machine.

3. Fuzzy SVM Regression

In many real-world applications, available information is often uncertain, imprecise and incomplete and thus, usually is represented by fuzzy sets or a generalization of interval data. For handling those fuzzy data, fuzzy regression analysis is an important tool and has been successfully applied in different applications. In this section, we applied the fuzzy set theory for the SVM regression model. The fuzzified parameters within SVM regression will make it more elastic.

First, we deal with fuzzy desired output in the regression task. The given output data, denoted by $\tilde{Y}_i = (y_i, e_i)$, are symmetric triangular fuzzy numbers, where y_i is a center and e_i is a width. The membership function of \tilde{Y}_i is given by

$$\mu_{\tilde{Y}_i}(y) = 1 - \frac{|y - y_i|}{e_i} \quad (5)$$

Then, for handling those fuzzified training data, the components in weight vector and bias term using in the SVM regression model are also set to be fuzzy numbers. Give fuzzy weight vector $\mathbf{W} = \{\mathbf{w}, \mathbf{c}\}$ and fuzzy bias term $B = \{b, d\}$, $\mathbf{W} = \{\mathbf{w}, \mathbf{c}\}$ is the fuzzy weight vector, where each components within it are fuzzy numbers. It was denoted in the vector form of $\mathbf{w} = [w_1, \dots, w_n]^t$, and $\mathbf{c} = [c_1, \dots, c_n]^t$, which means “approximation \mathbf{w} ”, described by the center \mathbf{w} and the width \mathbf{c} . Similarly, $B = \{b, d\}$ is the fuzzy bias term, which means “approximation b ”, described by the center b and the width d . The fuzzy parameters studied in this work are restricted to a class of “triangular” membership functions. The fuzzy function

$$Y = W_1\Phi(\mathbf{x}_1) + \dots + W_n\Phi(\mathbf{x}_n) + B = \mathbf{W} \cdot \Phi(\mathbf{x}) + B, \quad (6)$$

is defined by the following membership function [10]:

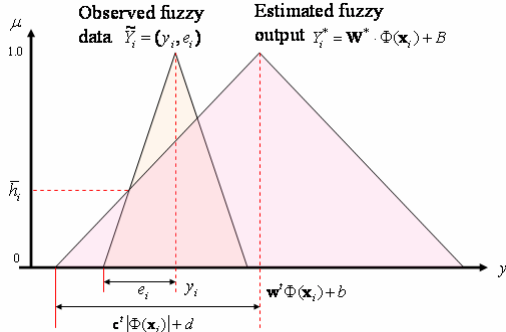


Fig. 3. Degree of fitting of Y_i^* to given fuzzy data \tilde{Y}_i .

$$\mu_Y(y) = \begin{cases} 1 - \frac{|y - (\mathbf{w}^t \Phi(\mathbf{x}) + b)|}{\mathbf{c}^t |\Phi(\mathbf{x})| + d} & \mathbf{x} \neq 0, \\ 1 & \mathbf{x} = 0, y = 0, \\ 0 & \mathbf{x} = 0, y \neq 0, \end{cases} \quad (7)$$

where $|\Phi(\mathbf{x})| = [|\Phi(\mathbf{x})_1|, \dots, |\Phi(\mathbf{x})_n|]^T$, n is the dimension of feature space, and $\mu_Y(y) = 0$ when $\mathbf{c}^t |\Phi(\mathbf{x})| + d \leq |y - (\mathbf{w}^t \Phi(\mathbf{x}) + b)|$. To formulate a fuzzy regression model, the followings are assumed to hold.

- (1) The data in the feature space can be represented by a fuzzy linear model: $Y_i^* = \mathbf{W}^* \cdot \Phi(\mathbf{x}_i) + B^*$. Given \mathbf{x}_i , Y_i^* can be obtained as

$$\mu_{Y_i^*}(y) = 1 - \frac{|y - (\mathbf{w}^t \Phi(\mathbf{x}_i) + b)|}{\mathbf{c}^t |\Phi(\mathbf{x}_i)| + d}.$$

- (2) The degree of the fitting of the estimated fuzzy linear model $Y_i^* = \mathbf{W}^* \cdot \Phi(\mathbf{x}_i) + B^*$ to the given data $\tilde{Y}_i = (y_i, e_i)$ is measured by the following index \bar{h}_i [10] which maximizes h subject to $\tilde{Y}_i^h \subset Y_i^{*h}$, where

$$\tilde{Y}_i^h = \{y \mid \mu_{\tilde{Y}_i}(y) \geq h\}$$

$$Y_i^{*h} = \{y \mid \mu_{Y_i^*}(y) \geq h\}$$

which are h -level sets. This index \bar{h}_i is illustrated in Fig. 3. The key is that the observed fuzzy data, adjusted for the h -certain factor, is contained within the estimated fuzzy output, adjusted for the h -certain factor. The degree of fitting of the linear model to all data $\tilde{Y}_1, \dots, \tilde{Y}_N$ is defined by $\min_j \lceil \bar{h}_j \rceil$.

- (3) The vagueness of the fuzzy linear model is defined by $\sum_j c_j + d$.
- (4) The capacity control term [9] of the fuzzy regression model is defined by the Euclidean norm $\|\mathbf{w}\|^2$.

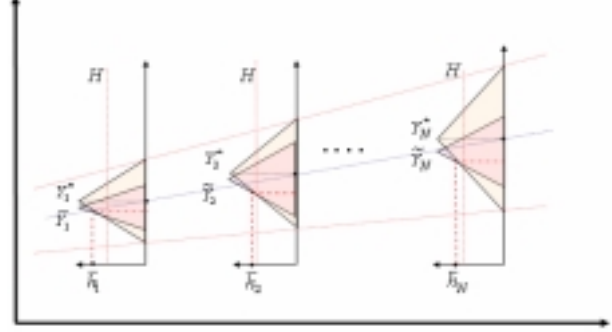


Fig. 4. Explanation of fuzzy linear regression model with $Y_i^* = \mathbf{W}^* \cdot \Phi(\mathbf{x}_i) + B^*$

The regression problem is explained as obtaining the fuzzy weight vector $\mathbf{W}^* = \{\mathbf{w}, \mathbf{c}\}$ and the fuzzy bias term $B^* = \{b, d\}$, such that the fitting degree between estimated output $\mu_{Y_i^*}(y)$ and desired output $\mu_{\tilde{Y}_i}(y)$ is more than a given constant H for all i , where H is chosen as the degree of the fitting of the fuzzy linear model by the decision-maker. Figure 4 depicts this situation graphically. The value \bar{h}_i can be obtained as

$$\bar{h}_i = 1 - \frac{|y_i - (\mathbf{w}^t \Phi(\mathbf{x}_i) + b)|}{(\mathbf{c}^t |\Phi(\mathbf{x}_i)| + d) - e_i}, \quad (8)$$

which is equal to the early work by Tanaka [10]. Our regression task here is therefore to minimize

$$J = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_j c_j + d \right), \quad (9)$$

subject to $\bar{h}_i \geq H$ for all $i=1, \dots, N$, where $\|\mathbf{w}\|^2$ is the term which characterizes the model complexity, minimize of $\|\mathbf{w}\|^2$ can be understood in the context of regularization operators [9]. $\sum c_j + d$ is the term which characterizes the vagueness of the model. The more vagueness in the fuzzy linear regression model means the more inexactness in the regression result. C is a trade off parameter chosen by the decision-maker. The value of H determines the low bound for the degree of the fitting of the fuzzy linear model, and \bar{h}_i is the degree of fitting of the estimated fuzzy linear model $Y_i^* = \mathbf{W}^* \cdot \Phi(\mathbf{x}) + B^*$ to the given fuzzy desired output data $\tilde{Y}_i = (y_i, e_i)$.

More specifically, according to Eq. (8), our problem is to find out the fuzzy weight vector $\mathbf{W}^* = \{\mathbf{w}, \mathbf{c}\}$ and fuzzy bias term $B^* = \{b, d\}$, which are the solution of the following quadratic programming problem:

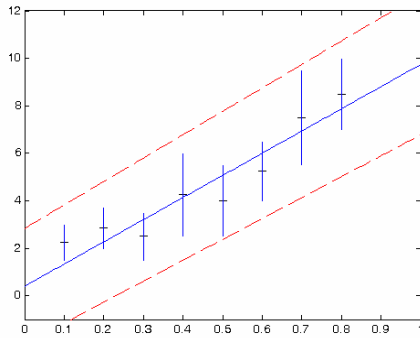


Fig.5. The result of Tanaka's fuzzy linear regression model (RMSE=0.7190).

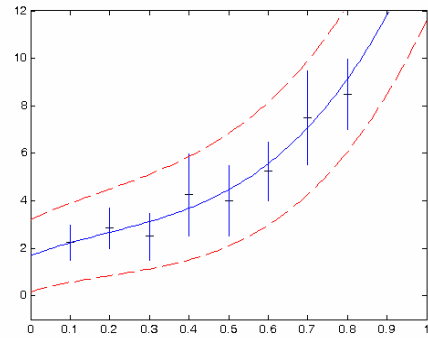


Fig.6. The result of proposed fuzzy SVM regression model with polynomial kernel function of degree 3 (RMSE=0.4525).

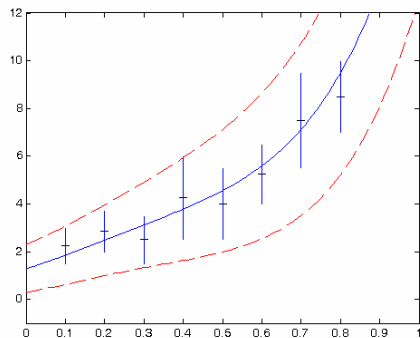


Fig.7. The result of proposed fuzzy SVM regression model with polynomial kernel function of degree 5 (RMSE=0.5519).

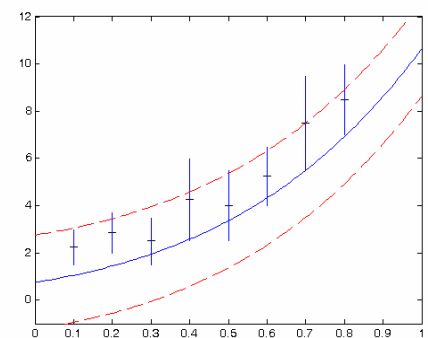


Fig.8. The result of SVM regression model with polynomial kernel function of degree 5 (RMSE=1.3409).

$$\min_{\mathbf{w}, \mathbf{c}, b, d} J = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_j c_j + d \right)$$

subject to

$$\begin{aligned} (\mathbf{w}^t \Phi(\mathbf{x}_i) + b) + (1-H)(\mathbf{c}^t |\Phi(\mathbf{x}_i)| + d) &\geq y_i + (1-H)e_i \\ -(\mathbf{w}^t \Phi(\mathbf{x}_i) + b) + (1-H)(\mathbf{c}^t |\Phi(\mathbf{x}_i)| + d) &\geq -y_i + (1-H)e_i \end{aligned}$$

for $i=1, \dots, N$, $d \geq 0$, and $c_j \geq 0$, for $j=1, \dots, n$.

Compared Fig. 4 with Fig. 1, we can understand the difference between SVM regression model and fuzzy SVM regression model. The SVM regression model seeks a linear function in the feature space that has at most ϵ deviation from the actually obtained targets y_i for all the training data, whereas fuzzy SVM regression model seeks a fuzzy linear function with fuzzy parameters in the feature space that has at last H fitting degree from the fuzzy desired targets \tilde{Y}_i for all the training data.

5. Experiments

We present two types of experiments to give an illustrative point of view of the proposed fuzzy support vector regression machine.

First, we present a numerical example that can be visualized. We then apply the proposed method to the house price problem.

A. Numerical Example

In this section, we use one data set, crisp inputs-fuzzy outputs, shown in Table 1. This data is taken from Tanaka and Lee [11]. The proposed fuzzy support vector regression approaches are applied to this data set from an illustrative point of view.

Table 1. Crisp input-fuzzy output data.

No.(i)	1	2	3	4
Crisp Input x	0.1	0.2	0.3	0.4
$\tilde{Y}=(y,e)$	(2.25,0.75)	(2.875,0.875)	(2.5,1.0)	(4.25,1.75)
No.(i)	5	6	7	8
Crisp Input x	0.5	0.6	0.7	0.8
$\tilde{Y}=(y,e)$	(4.0,1.5)	(5.25,1.25)	(7.5,2.0)	(8.5,1.5)

Figure 5 illustrates the result of the Tanaka's fuzzy linear regression model [10]. The resulted Root Mean-Square-Error (RMSE) is 0.7190. Figures 6 and 7 illustrate the result of the proposed fuzzy SVM regression model with polynomial kernel function of degree 3 and 5, respectively. In this example, we set the parameter

$H=0.5$ and $C=100$. The resulted RMSEs are 0.4525 and 0.5519, respectively. As in Figs. 6 and 7, the solid curve explains $\mathbf{w}'\Phi(x)+b$, whereas the two dash curve explain $\mathbf{w}'\Phi(x)+b+\mathbf{c}'/\Phi(x)/+d$ and $\mathbf{w}'\Phi(x)+b-\mathbf{c}'/\Phi(x)/-d$, respectively. The result of the proposed regression model explains “approximate $\mathbf{w}'\Phi(x)+b$ ”, described by center $\mathbf{w}'\Phi(x)+b$ and width $\mathbf{c}'/\Phi(x)/+d$. The width explains the vagueness of the resulted regression model. Figure 8 illustrates the result of the SVM regression model with polynomial kernel function of degree 5.

In this SVM regression model, we set the epsilon insensitive value $\varepsilon=\max(e_i)=2.0$. Namely, at each point x_i , we allow an error of ε , and the error is penalized only if it is outside the ε -tube. The resulted RMSE in SVM regression model is 1.3409. As in Fig. 8, the solid curve explains $f(x)=\mathbf{w}'\Phi(x)+b$, whereas the two dash curve explains $f(x)+\varepsilon$ and $f(x)-\varepsilon$, respectively. Compared with the SVM regression approach, the proposed fuzzy support vector regression model works fairly well.

Here we should note that the resulted fuzzy nonlinear SVM regression model is more vague around $x=[0.5\ 0.8]$ since the given fuzzy desired-output data are more inexact in this range. Noted the size of ε -tube in original SVM regression model is a constant predefined by the decision-maker. The size of tube is assumed to be independent of x_i . It is improper since in many real applications the effects of the training points are different. We would require that the precise training points must be regressed correctly, and would allow more errors on imprecise training points.

B. The House Price Example

We now apply the proposed fuzzy SVM regression approach to the house price model. The house price data shown in Table 2 is taken from Tanaka [10,11].

Table 2. Input-output data concerning house prices.

NO.	$\tilde{Y} = (y, e)$	x_1	x_2	x_3
1	(606, 100)	38.09	36.43	5
2	(710, 50)	62.10	26.50	6
3	(808, 100)	63.76	44.71	7
4	(826, 150)	74.52	38.09	8
5	(865, 250)	75.38	41.40	7
6	(852, 200)	52.99	26.49	4
7	(917, 200)	62.93	26.49	5
8	(1031, 250)	72.04	33.12	6
9	(1092, 600)	76.12	43.06	7
10	(1203, 100)	90.26	42.64	7
11	(1394, 350)	85.70	31.33	6
12	(1420, 250)	95.27	27.64	6
13	(1601, 300)	105.98	27.64	6
14	(1632, 500)	79.25	66.81	6
15	(1699, 650)	120.50	32.25	6

y_i : i th house price (100,000 yen). x_1 : first floor space (m^2).
 x_2 : second floor space (m^2). x_3 : number of rooms.

The vagueness e_i (1,000 yen) is assigned by the authors. From these data, we decide the fuzzy SVM regression system

$$Y = W_1x_1^2 + W_2x_2^2 + W_3x_3^2 + W_4x_1x_2 + W_5x_1x_3 + W_6x_2x_3 + W_7x_1 + W_8x_2 + W_9x_3 + B.$$

Namely, we use the polynomial transform function of degree 2 to map the input data into the feature space. The result of fuzzy weight vector and fuzzy bias term are given in Table 3, where we set $H = 0.5$ and $C = 70$ in this experiment. The estimated values Y_i^* and the fitting degree to the desired fuzzy output \tilde{Y}_i are shown in Table 4. For the example of NO. 1, the estimated value is obtained as $Y_1^*=(710.59,358.3)$ while the given data is $\tilde{Y}_1=(606,100)$. Note that the solution satisfies the relation $\tilde{Y}_i^h \subset Y_i^{*h}$ with $H=0.5$. The following results are reached.

- (1) The degree of fitting of the estimated fuzzy model Y_i^* to the given fuzzy desired output data \tilde{Y}_i are larger than $H=0.5$ for all examples, as illustrated in Table 4.
- (2) The fact that $w_3, w_5, w_6,$ and w_9 are negative, depends on the correlations between variable x_3 . In the case of a fixed floor space, the larger the number of rooms, the lower the price, since the small rooms diminish the space.
- (3) The larger house price we estimated in this regression model, the larger vagueness in house price we obtained. This is nature in real world experience.

Table 3. Fuzzy parameters (\mathbf{W} and B) with $H=0.5$.

Fuzzy Parameter	W_1	W_2	W_3	W_4	W_5
Center(W_i)	0.11	0.06	-0.04	0.04	-0.30
Width(C_i)	0.13	0.05	0.0	0.11	0.02
Fuzzy Parameter	W_6	W_7	W_8	W_9	B
Center(W_i)	-0.22	0.01	-0.01	-0.01	534.39
Width(C_i)	0.01	0.0	0.0	0.0	0.0

Table 4. Estimation y_i^* obtained from fuzzy SVM regression

NO.	Estimated			NO.	Estimated		
	House Price	Vagueness	Degree of Fitting		House Price	Vagueness	Degree of Fitting
1	710.59	358.30	0.59	9	1127.98	985.35	0.91
2	871.99	416.39	0.56	10	1366.33	1374.24	0.87
3	962.55	704.42	0.74	11	1248.52	1041.12	0.79
4	1025.07	865.92	0.72	12	1398.45	1242.43	0.98
5	1104.45	938.78	0.65	13	1616.36	1564.14	0.99
6	831.60	240.80	0.50	14	1446.43	1501.50	0.81
7	916.95	431.23	1.00	15	1990.09	2149.11	0.81
8	1041.68	724.82	0.98				

6. Conclusions

The difference between SVM regression model and fuzzy SVM regression model is the SVM regression model seeks a linear function that has at most ε deviation from the actually obtained targets y_i for all the training data, whereas fuzzy SVM regression model seeks a fuzzy linear function with fuzzy parameters that has at last H fitting degree from the fuzzy desired targets \tilde{Y}_i for all the training data.

Another important feature in SVM algorithm is the prime optimization problem can be reformulated as the Wolf dual problem by using the Lagrange multiplier method. The benefit of transforming into dual problem is, by using the kernel trick ($K(\mathbf{x}, \mathbf{y}) \equiv \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$), this makes SVM possible to deal with the feature space with vast high dimension (possibly infinite dimension). In our approach, we do not reformulate the prime fuzzy SVM regression problem as the dual problem because the definitions of operations on fuzzy number using here. This limits the dimensionality of feature space in the proposed fuzzy SVM regression model to be finite. Hence, we only use the polynomial regression models. However, almost all fuzzy regression approaches suggested polynomials as regression models since any function can be represented by the polynomial approximation. How to transform the prime fuzzy SVM regression problem into the dual problem will be our further work.

7. References

- [1] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 955-974, 1998.
- [2] J.-H. Chiang and P.-Y. Hao, "Support vector learning mechanism for fuzzy rule-based modeling: a new approach," *IEEE Trans. on Fuzzy Systems*, vol. 12, no. 1, pp. 1-12, 2004.
- [3] C. Cortes, and V. N. Vapnik, "Support vector network," *Machine learning*, vol. 20, pp. 1-25, 1995.
- [4] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. N. Vapnik, "Support vector regression machines," *In Advances in Neural Information Processing Systems*, vol. 9, pp. 155-161, 1996.
- [5] D. Dubois and H. Prade, "Operations on fuzzy number," *Int. J. Syst. Sci.*, vol. 9, pp. 613-626, 1978.
- [6] P.-Y. Hao and J.-H. Chiang, "A fuzzy model of support vector regression," *IEEE Int. Conf. On Fuzzy Systems*, vol. 1, pp. 738-742, 2003.
- [7] H.-P. Huang and Y.-H. Liu, "Fuzzy support vector machines for pattern recognition and data mining," *International Journal of Fuzzy Systems*, vol. 4, no. 3, pp. 826-835, Sep. 2002.
- [8] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464-471, Mar. 2002.
- [9] A. Smola, B. Schölkopf, and K.-R. Müller, "The connection between regularization operations and support vector kernels," *Neural Networks*, vol. 11, pp. 637-649, 1998.
- [10] H. Tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," *IEEE. Trans. On Syst., Man, and Cyber.*, vol. 12, no. 6, pp. 903-907, 1982.
- [11] H. Tanaka and H. Lee, "Interval regression analysis by quadratic programming approach," *IEEE Trans. on Fuzzy Systems*, vol. 6, no. 4, pp. 473-481, 1998.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
- [13] R. R. Yager, "On solving fuzzy mathematical relationships," *Information and Control*, vol. 41, pp. 29-55, 1979.
- [14] L. A. Zadeh, "The concept of linguistic variable and its application to approximate reasoning," *Inform.Sci.*, vol. 8, pp. 199-249, 1975.