Fuzzy Cost Support Vector Regression

Hadi Sadoghi Yazdi, Tahereh Royani, Mehri Sadoghi Yazdi, Sohrab Effati

Abstract—In this paper, a new version of support vector regression (SVR) is presented namely Fuzzy Cost SVR (FCSVR). Individual property of the FCSVR is operation over fuzzy data whereas fuzzy cost (fuzzy margin and fuzzy penalty) are maximized. This idea admits to have uncertainty in the penalty and margin terms jointly. Robustness against noise is shown in the experimental results as a property of the proposed method and superiority relative conventional SVR.

Keywords—Support vector regression; Fuzzy input; Fuzzy cost.

I. INTRODUCTION

THE standard support vector machine works over crisp training samples. Chun-fu Lin [1, and 2] proposed fuzzy support vector machine (FSVM) by considering the noise in the training samples. They used the membership function to express the membership value of a sample to positive or negative class, but with crisp training data. So it remains a conventional support vector machine from view point of fuzzy theory. Importance degree of training data are modeled in the FSVM by insertion of membership value μ_i in penalty term of

cost function to form of $\frac{1}{2}\|W\|^2 + C(\sum_{i=1}^{l} \mu_i \xi_i)$. It is noted

that the error term ξ_i is scaled by μ_i. The fuzzy membership values are used to weight the soft penalty term in the cost function of SVM. The weighted soft penalty term reflects the relative fidelity of the training samples during training. Important sample with large membership value will have more emphasis in the FSVM training procedure and more effect over determination of hyperplanes.

Hong into [3] presents support vector fuzzy regression machines. This paper introduces the use of SVM for multivariate fuzzy linear and nonlinear regression models. Presented model in [3] for regression includes fuzzy input and output $(\tilde{x}, \tilde{y})$ to form of

$$\tilde{y} = w^T \tilde{x} + \tilde{b} .\qquad (1)$$

Hadi Sadoghi Yazdi is assistant professor in Engineering Department, Tarbiat Moallem University of Sabzevar, Sabzevar, Iran, and Computer Department, Ferdowsi Mashad University, Mashad, Iran (sadoghi@sttu.ac.ir)
Tahereh Royani is Msc student in Engineering Department, Tarbiat Moallem University of Sabzevar, Sabzevar, Iran.
Mehri Sadoghi Yazdi is MSc student in Department of Computer Engineering, faculty of Electrical and Computer Engineering, Shahid Beheshti University of Tehran (me.sadooghi@mail.sbu.ac.ir).
Sohrab Effati is assistant professor in Mathematics Department, Tarbiat Moallem University of Sabzevar, Sabzevar, Iran and Mathematics Department, Ferdowsi Mashad University, Mashad, Iran

Then a SVM model is used for calculation of crisp w (weights). This model includes conventional fuzzy regression with new constraints. Upper and lower bound of fuzzy input and output are used for generation of constraints. But effect of fuzzy variables (input and output) over cost of SVR has not been considered. Assuredly, uncertainty in input data infects over margin and penalty maximization in the SVR which has not been studied in the previous works.

In [4] Ji studied support vector machine with fuzzy chance constrains to following form

$$Minimize \quad \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{l} \xi_i$$

$$subject\ to \qquad\qquad\qquad (2)$$

$$Pos\{\ y_i(W^T \tilde{X}_i + b) + \xi_i \geq 1\ \} \geq \lambda$$

$$\xi_i \geq 0, \qquad i = 1,2,...,l.$$

They showed that $Pos\{\tilde{a} \leq 0\} \geq \lambda$ with triangular fuzzy number $\tilde{a} = (r_1, r_2, r_3)$ and for any given level $\lambda(0 < \lambda \leq 1)$ is equivalent to: $(1-\lambda)r_1 + r_2 \leq 0$. Thereupon, constraints of (2) are simplified.

In our previous work [5], we apply probabilistic constraints for reducing of noisy samples in maximization of margin. A Constraint is to form of $Pr(d_i(w^T x_i + b) \geq u_i) \geq \delta_i$ where

u_i is independent random variable with known distribution functions and $0 \leq \delta_i \leq 1$ is value of effect of ith samples in fixation of the optimal hyperplane.

Liu in [6] presented total margin-based adaptive fuzzy support vector machines (TAF-SVM). TAF-SVM is a type of FSVM [1, and 2] which also corrects the skew of the optimal separating hyperplane due to the very imbalanced data sets by using different cost algorithm. This work is performed with dividing training data into two categories with different importance and result in dual problem is different boundary for Lagrange multipliers.

In [1], linear and quadratic functions are presented for μ_i in the FSVM which two main targets are followed, increasing margin and decreasing misclassification error. In [7] authors present two new methods for calculation of membership function of μ_i based on geometry distribution of the training samples. Those samples are near to optimal hyperplane, have similar geometry property. The main idea of FSVM [1] is that if the input is detected as an outlier or noisy sample, membership function decreases so total error term decrease. In [8] new method for μ_i of FSVM is presented which follows in the same idea that one input is assigned a low membership of the class if it is detected as an outlier. However, method of

[8] treats each input as an input of the opposite class with higher membership and it makes full use the data and achieves better generalization ability. Also in two different works [9, and 10], authors try to determine membership function in multi-category data classification.

After studying of related works, we categorize those to following form,

I) Standard FSVM and its variants, which change membership function μ_i.

II) SVMs with special constraints for better operation against noisy samples.

III) SVM as a method for finding optimal parameters of regression model.

Main idea in this work is presentation of full fuzzy support vector machine. We cannot believe fuzzy input or fuzzy penalty exist alone. If we assume that input signal is a fuzzy number then with this assumption fuzzyfication permeates into output part of SVM include margin and penalty terms. In this paper a new fuzzy cost and fuzzy input signal is considered and this work is organized as follows.

The SVM and SVR are discussed in Section 2 with details. Section 3 devotes to explain the proposed method namely fuzzy cost SVR (FCSVR). Experimental results are discussed in Section 4 and final section include to conclusions and future work.

II. SUPPORT VECTOR MACHINE AND REGRESSION

We first discuss Support Vector Machine and Regression, prior to introducing our approach. The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The mapping function can be either a classification function, i.e., the category of the input data, or a regression function. Initially developed for solving classification problems, support vector techniques can be successfully applied to regression. The general regression learning problem is set as follow:

Suppose we are given training data $\{(X_1, y_1), (X_2, y_2) \ldots, (X_l, y_l)\} \subset X \times R$, where X denotes the space of the input patterns (e.g. $X = R^D$). In ε-SV regression, our goal is to find a function f(x) that has at most ε deviation from the actually obtained targets y_i for all the training data. The regressor must not only fit the given data well, but also make minimal errors in predicting the values at any other arbitrary point in R^D. Nonlinear regression is accomplished by fitting a linear regressor in a higher dimensional feature space. A nonlinear transformation ϕ is used to transform data points from the input space of dimension D into a feature space having a higher dimension L. The nonlinear mapping is denoted by $\phi: R^D \to R^L$.

This problem can be written as a convex optimization problem:

$$Minimize \quad \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$subject\,to \quad y_i - W^T\phi(X_i) - b \le \varepsilon + \xi_i$$
$$- y_i + W^T\phi(X_i) + b \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0 \tag{3}$$

where C > 0 is a constant, ξ_i, ξ_i^* are slack variables for soft margin SVM, that allow to accept some deviation larger than ε that is precision. It turns out that in most cases the optimization problem (3) can be solved more easily in its dual formulation.

$$Maximize \quad -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(X_i, X_j) -$$
$$\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) \tag{4}$$

$$subject\,to \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C]$$

where α_i, α_i^* are Lagrange coefficients and matrix K is termed as a kernel matrix and its elements are given by $K(X_i, X_j) = \phi(X_i)^T\phi(X_j)$, $i, j = 1, 2, \ldots, M$.

By solving (4) we can find Lagrange coefficients and by replacing them, we have

$$W = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\phi(X_i),$$

Thus we can find hyperplane function as

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(X_i, X_j) + b \tag{5}$$

III. THE PROPOSED FUZZY COST SUPPORT VECTOR REGRESSION (FCSVR)

In this section, we discuss the proposed algorithm for support vector regression, termed as fuzzy cost support vector regression (FCSVR). Consider the fuzzy sample set $S = \{(\widetilde{X}_1, y_1), (\widetilde{X}_2, y_2), \ldots (\widetilde{X}_l, y_l)\}$, where $\widetilde{X}_i = (\widetilde{x}_1, \widetilde{x}_2, \ldots \widetilde{x}_d)$ is a fuzzy input vector and y_i is desired output. Fuzzy input can be having different form of membership functions. Here we will consider the following linear membership function related to each fuzzy sample:

$$\mu_i(x) = \begin{cases} 1 & x \le X_i \\ \dfrac{X_i + d_i - x}{d_i} & X_i \le x \le X_i + d_i \\ 0 & x \ge X_i + d_i \end{cases} \tag{6}$$

where d_i is tolerance of ith input vector and $d_i \in (0,1]$, $x \in R$, $i = 1,\dots,l$.

The support vector machine for fuzzy linear examples is to solve the below fuzzy quadratic programming:

$$Minimize \quad \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$subject\ to \quad y_i - W^T \widetilde{X}_i - b \le \varepsilon + \xi_i \tag{7}$$
$$- y_i + W^T \widetilde{X}_i + b \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0.$$

In (7), fuzzy input lead to fuzzy cost $Z = \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$. For entering of fuzzy concept in Z, we use following algorithm includes finding upper and lower cost function (Z) and its fuzzification.

Step I) Boundary calculation of cost function
Range of fuzzy samples is $[X_i, X_i + d_i]$ so; Z is obtained in bounds with solution of two classical convex quadratic programming (QP) problems to following form ((8) for lower bound and (9) for upper bound).

$$Minimize \quad Z = \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$\tag{8}$$

$$subject\ to \quad y_i - W^T X_i - b \le \varepsilon + \xi_i$$
$$- y_i + W^T X_i + b \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0$$

and

$$Minimize \quad Z = \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$subject\ to \quad y_i - W^T(X_i + d_i) - b \le \varepsilon + \xi_i \tag{9}$$
$$- y_i + W^T(X_i + d_i) + b \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0.$$

By solving (8) and (9), obtained values of Z are Z_1 and Z_2. In the next step, cost function of Z is fuzzified for studying of all states of input signal in $[X_i, X_i + d_i]$.

Step II) Fuzzyfication of cost function
Lower and upper bound of Z are,

$$Min\{Z_1, Z_2\} = Z_l \tag{10}$$

$$Max\{Z_1, Z_2\} = Z_u \tag{11}$$

Where Z_u is upper bound and Z_l is lower bound of the object function of (8) and (9), respectively, and other optimum values are varying between two values where inputs are varying between $[X_i, X_i + d_i]$. Now we can consider following linear membership function to determine optimal grade for Z.

$$\mu_g(x) = \begin{cases} 1, & Z \le Z_l \\ \dfrac{Z_u - Z}{Z_u - Z_l}, & Z_l \le Z \le Z_u \\ 0, & Z \ge Z_u. \end{cases} \tag{12}$$

And each of rows of classic model shall now be represented by a fuzzy set.

Step III) Finding of decision space
The membership function of the fuzzy set "decision" of fuzzy model is as following form:

$$\mu_i(x) = \begin{cases} 0 & y_i - W^T(X_i + d_i) - b \ge \varepsilon + \xi_i \\ \dfrac{\varepsilon + \xi_i - (y_i - W^T(X_i + d_i) - b)}{W^T d_i} & y_i - W^T(X_i + d_i) - b \le \varepsilon + \xi_i \le y_i - W^T X_i - b \\ 1 & y_i - W^T X_i - b \le \varepsilon + \xi_i \end{cases} \tag{13}$$

and

$$\mu_i^*(x) = \begin{cases} 0 & -y_i + W^T X_i + b \ge \varepsilon + \xi_i^* \\ \dfrac{\varepsilon + \xi_i^* - (-y_i + W^T X_i + b)}{W^T d_i} & -y_i + W^T X_i + b \le \varepsilon + \xi_i^* \le -y_i + W^T(X_i + d_i) + b \\ 1 & -y_i - W^T(X_i + d_i) + b \le \varepsilon + \xi_i^* \end{cases} \tag{14}$$

Where $W^T d_i \ne 0$.

Maximizing the minimum of $\mu(x)$ is optimal solution of this model.

$$Maximum$$
$$Min\{\mu_g(x), \mu_{c1}(x), \dots, \mu_{cl}(x), \mu_{c1}^*(x), \dots, \mu_{cl}^*(x)\} \tag{15}$$

By using α-cut method, we arrive to following constraint programming

$$\text{Maximize} \quad \alpha$$
$$\text{subject to} \quad \mu_g(x) \geq \alpha$$
$$\mu_{ci}(x) \geq \alpha \qquad (16)$$
$$\mu_{ci}^*(x) \geq \alpha$$
$$0 \leq \alpha \leq 1$$

And with replacing from above and supposing that $W^T d_i$ is against zero we have

$$\text{Maximize} \quad \alpha$$
$$\text{subject to} \quad -\frac{1}{2}\|W\|^2 - C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right) \geq \alpha(Z_u - Z_l) - Z_u$$
$$\varepsilon + \xi_i - \left(y_i - W^T(X_i + d_i) - b\right) \geq \alpha W^T d_i \qquad (17)$$
$$\varepsilon + \xi_i^* - \left(-y_i + W^T X_i + b\right) \geq \alpha W^T d_i$$

With solving this problem, we find optimized *W, b* and maximum α.

IV. EXPERIMENTAL RESULTS

This section demonstrates the effectiveness of our proposed models for linear function approximation. The experimental results pertaining to our proposed models are compared to conventional support vector regression model. In this work, we study effect of measurement noise over the proposed method in estimation of desired function. We use Matlab for implementation and testing our method. Results are obtained from average of 400 times execution of program. Before doing experiments, some definitions are mentioned.

Triangular or trapezoidal form of fuzzy numbers is used for simulation of uncertain data in the operation of regression. They fall in duration $[X_i, X_i + d_i]$. If $\widetilde{X}$ is a fuzzy number then alpha-cut of $\widetilde{X}$ is showed by $\widetilde{X}_\alpha = \{x : \mu_{\widetilde{X}} \geq \alpha\}$ that is a closed interval and it is denoted to $\widetilde{X}_\alpha = [X_\alpha^L, X_\alpha^U]$ where $\alpha \in [0,1]$.

A LR-type fuzzy number $\widetilde{X}$ with its membership function $\mu_{\widetilde{X}}(x)$:

$$\mu_{\widetilde{X}}(x) = \begin{cases} L\left(\dfrac{m_1 - x}{\alpha}\right) & \text{for } x \leq m_1 \\ 1 & \text{for } m_1 \leq x \leq m_2 \\ R\left(\dfrac{x - m_2}{\beta}\right) & \text{for } x \geq m_2 \end{cases}$$

This is called an LR-type TFN (Trapezoidal Fuzzy Number) where m_1, m_2 are boundaries which in Fig 1.a are 2, 6 respectively. α, β are slopes of right and left side of trapezoid. We show two kinds of TFN in the following figure.
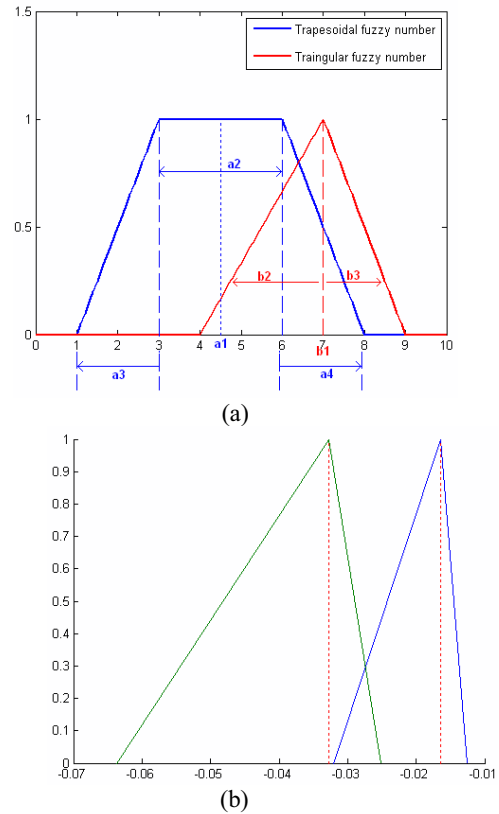

(a)

(b)

Fig. 1: a) Two common samples of TFN, b) LR-type triangular fuzzy number

In general case, fuzzy number $\widetilde{X}$ is a number in duration $[X_i, X_i + d_i]$ with defined uncertainty degree. Noise may affect over parameters of fuzzy numbers or effect of noise can be modeled to following form over LR-type fuzzy numbers,

$$\mu_{\widetilde{X}_n}(x) = \begin{cases} L\left(\dfrac{\hat{m}_1 - x}{\hat{\alpha}}\right) & \text{for } x \leq \hat{m}_1 \\ 1 & \text{for } \hat{m}_1 \leq x \leq \hat{m}_2 \\ R\left(\dfrac{x - \hat{m}_2}{\hat{\beta}}\right) & \text{for } x \geq \hat{m}_2 \end{cases}$$

Where $\hat{m}_1$, $\hat{m}_2$, $\hat{\alpha}$, $\hat{\beta}$ are noisy parameters of LR-type fuzzy number which has been corrupted with uniform noise. Of course accurate study of noise effects and method of contamination is a new works in field of fuzzy numbers.

Signal to Noise Ratio (SNR) is defined $\dfrac{20 \log D_s}{D_n}$ where D_s is main value of parameters and D_n is domain of noise. Also error is defined to form of $\dfrac{1}{N}\sum_{i=0}^{N}(\hat{y}_i - y_i)^2$ where $\hat{y}_i$ is obtained output using SVR or FCSVR method and y_i is desired output and N is number of training samples.

Example: with given samples are shown in the following figure, we want to estimate linear equation to form of $y = wx + b$. Of course for checking results, we know given data are generated from $y = 3.78x + 4$ and noise is added to y and we model noise to form of added noise into parameters of fuzzy numbers.
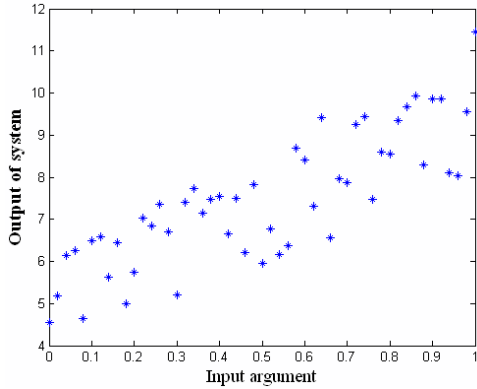

Fig. 2: Noisy captured data in signal to noise ratio equal 13.9dB.

Obtained results using standard SVR and FCSVR are shown in Table 1. Optimum value of input tolerance (d_i , mentioned in (17)) is obtained using exhaustive search and shown in second column of table. Maximum value of membership degree α is in the final column of table. Error indicates superiority of the proposed FCSVR relative standard SVR. $\hat{w}_{SVR}$, $\hat{b}_{SVR}$ are estimated parameters by SVR and $\hat{w}_{FCSVR}$, $\hat{b}_{FCSVR}$ are estimated parameters by FCSVR. Also error of SVR (e_{SVR}) and error of FCSVR (e_{FCSVR}) are shown in Fig 3.

TABLE I

RESULT OF ESTIMATION OF $y = wx + b$ FROM CAPTURED NOISY SAMPLES (Y_N) (AS SHOWN IN FIG 2) IN DIFFERENT SNR OVER 400 RUNS

| SNR*	d_i	$\hat{w}_{SVR}$	$\hat{w}_{FCSVR}$	$\hat{b}_{SVR}$	$\hat{b}_{FCSVR}$	e_{SVR}	e_{FCSVR}	α
26.02	0.08	3.82	3.68	4.10	4.01	0.04	0.0005	0.18
20	0.14	3.92	3.75	4.20	4.00	0.17	0.0006	0.19
16.47	0.2	4.00	3.83	4.29	3.96	0.35	0.0024	0.20
13.97	0.28	4.11	3.96	4.39	3.86	0.66	0.0112	0.23
12.76	0.3	4.10	3.97	4.46	3.87	1.03	0.02	0.22
12.04	0.34	4.19	4.05	4.50	3.82	1.49	0.0409	0.22
11.37	0.4	4.21	4.10	4.55	3.72	1.54	0.0438	0.24
10.45	0.42	4.33	4.18	4.58	3.80	2.38	0.0784	0.21
9.11	0.48	4.15	4.07	4.70	3.67	2.90	0.0819	0.23

*SNR per db
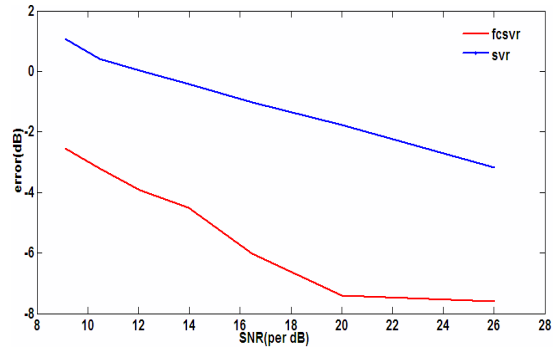
Fig. 3: Comparison of the proposed FCSVR and standard SVR

Fig 4 demonstrates optimum tolerance (d_i) in different SNRs. In low noise condition or low SNR, for having low error needs to decrease of d_i . It means decreasing of certainty degree (d_i) must be performed if it is sensed signal has been contaminated with noise. So, if SNR decreases for having lower error, d_i must increase. But main problem understands of level of noise existing in signal. In the future work we follow it for completion regression system.

Founded maximum membership (α) in (17) shows relation of α with SNR. With increasing noise level, α increases but 10-15% for 50% decreasing in SNR but for medium to high value of SNR. Increasing in α means fuzzy values are selected in narrower range. In the other words, according (17) constraints $\mu_g(x) \geq \alpha, \mu_{ci}(x) \geq \alpha, \mu_{ci}^*(x) \geq \alpha$ are satisfied in higher certainty. For $\mu_g(x) \geq \alpha$ with higher α means, optimum Z in (9) goes towards Z_l or cost function is spotted with higher degree of certainty. As we know, Z includes margin of SVR and penalty term, so decreasing uncertainty in margin is concluded in medium up to high level of SNR and simultaneous, penalty term have higher certainty. From $\mu_{ci}(x) \geq \alpha, \mu_{ci}^*(x) \geq \alpha$ we find in this condition, constraints move toward standard SVR.

Abstractly, in high value of SNR, regression model moves toward SVR with high value of uncertainty because of uncertainty in modeling of input data. This lemma is correct only in medium and high range of SNR (more than 13dB) according to Fig 5. But we cannot present any subject for low value of SNR now.
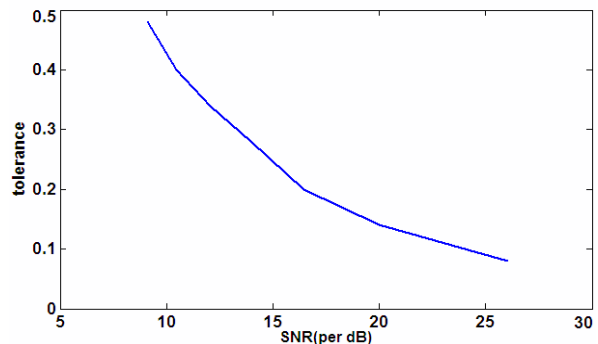

Fig. 4: Input tolerance (di) versus SNR
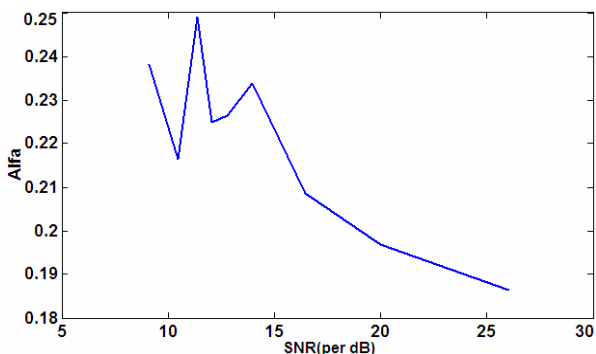
Fig. 5: Obtained α versus SNR

Fig 6 shows function estimation using SVR and FCSVR in two different (medium and high value) SNR. Robustness of FCSVR against noise is noticeable relative to SVR.
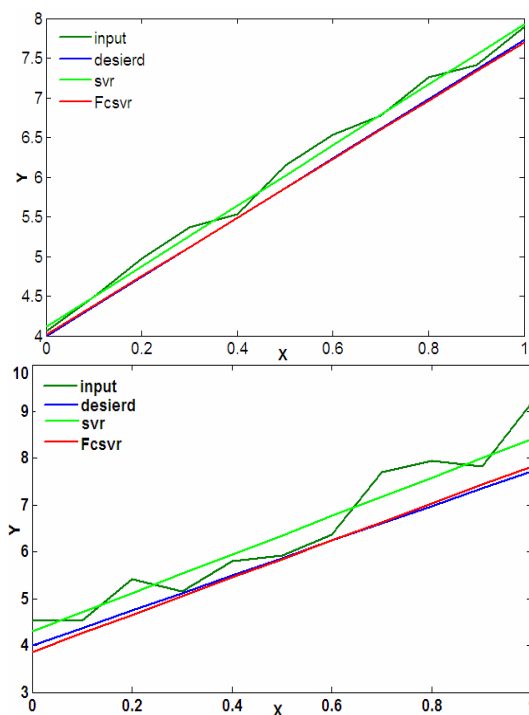

Fig. 6: Estimation of Y=3.73X+4 with two different SNR, left figure) SNR=30dB- right figure) SNR=14dB

V. CONCLUSION AND FUTURE WORKS

Noisy samples are caused decreasing performance in the support vector regression method. Fuzzy margin with fuzzy penalty concept was introduced in this paper. This idea could help into decreasing of noise effect. Several experiments were performed and compared with standard SVR. Results indicate to superiority of the proposed method relative conventional SVR.

REFERENCES

[1]   C-f. Lin, S-d. Wang, "Fuzzy Support Vector Machine," IEEE Trans. on Neural Networks, 2002, vol. 13, no. 2, pp. 464–471, Mar. 2002.
[2]   C-f. Lin, S-d. Wang "Training Algorithms for Fuzzy Support Vector Machines with Noisy Data," Pattern Recognition Letters, vol.25, pp. 1647–1656, 2004.
[3]   D.H. Hong, C. Hwang, "Support vector fuzzy regression machines," Fuzzy Sets and Systems 138, pp.271-281, 2003.
[4]   A-B. Ji, J-H. Pang, S-H. Li, J-P. Sun, "Support Vectro Machine for Classification Based on Fuzzy Training Data," Proceedings of the Fifth Int. Conf. on Machine Learning and Cybernetics, Dalian, pp.1609-1614, 13-16 August 2006.
[5]   H.Sadoghi Yazdi, S. Effati, Z. Saberi, "The Probabilistic Constraints in the Support Vector Machine," Applied Mathematics and Computation, vol.194, no.2, p.467-479, Dec 2007 .
[6]   Y-H. Liu, and Y-T. Chen, "Face Recognition Using Total Margin-Based Adaptive Fuzzy Support Vector Machines," IEEE Trans. on Neural Networks, vol. 18, no.1, pp.178-192, Jan. 2007.
[7]   L. Chu, and C. Wu, "A Fuzzy Support Vector Machine Based on Geometric Model," Proceedings of the fifth World Congress on Intelligent Control and Automation, Hangzhou, P.R. China, pp.1843-1846,June 15-19,2004.
[8]   Y. Wang, S. Wang, and K. K. Lai, "A New Fuzzy Support Vector Machine to Evaluate Credit Risk," IEEE Trans. on Fuzzy Systems, vol.13 no.6, pp.820-831, Dec. 2005.
[9]   Jayadeva, R. Khemchandani, S. andra, "Fuzzy Linear Proximal Support Vector Machines for Multi-Category Data Classification," Neurocomputing, vol.67, pp.426–435, 2005.
[10] T-Y. Wang, H-M. Chiang, "Fuzzy Support Vector Machine for Multi-Class Text Categorization," Information Processing and Management, vol.43, pp.914–929, 2007.