2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Fuzzy c-Regression Models Combined with Support Vector Regression

Tatsuya Higuchi and Sadaaki Miyamoto

***Abstract*— Fuzzy c-regression models (FCRM) give us multiple clusters and regression models of each cluster simultaneously, while support vector regression models (SVRM) involve kernel methods which enable us to analyze non-linear structure of the data. We combine these two concepts and propose the united fuzzy c-support vector regression models (FC-SVRM). In case that c is unknown, we introduce sequential regression models (SRM) into SVRM, and propose support vector sequential regression models (SVSRM). We show numerical examples to compare results from these methods.**

I. Introduction

Two main branches in data mining researches are studies of classification and regression, It is well-known that support vector machines [13] are playing the central role in classification. Support vector regression [3] has also been actively studied, but there are much rooms for more investigations.

The fuzzy c-regression [7] is another important issue in regression models, where two or more regression models should be identified for a set of independent variables.

In this paper the combination of these two is considered. We thus formulate fuzzy c-regression models based on support vector regressions. Moreover sequential extraction of regression models is studied using the present framework that is effective when we do not know c, the number of models.

This formulation includes linear and nonlinear regressions; the latter is realized by using positive definite kernel functions.

The rest of this paper is organized as follows. Section 2 gives preliminaries, which is followed by Section 3 where the proposed method of regression models using support vectors is described. Section 4 handles a method of sequentially extracting regression models. Section 5 shows numerical examples for comparing results from these methods. Section 6 finally concludes the paper.

II. Preliminaries

We begin with giving notations. A column vector is denoted by $\boldsymbol{v}$, and hence a row vector is written as $\boldsymbol{v}^T$ using transpose. A data set $(\boldsymbol{x}_k^T, y_k)^T, k = 1,\dots,n$ is assumed to be given, where $\boldsymbol{x}_k = (x_k^1,\dots,x_k^p)^T \in \boldsymbol{R}^p$ are data of independent variables, while $y_k \in \boldsymbol{R}$ are those of a dependent variable.

The membership of x_k to cluster i is denoted by $u_{ik}, i = 1,\dots,c, k = 1,\dots,n$, where the number of clusters is c. $A = (a_{\ell m})$ means a matrix A with (ℓ, m) components $a_{\ell m}$.

Tatsuya Higuchi is with the Graduate School of Systems and Information Engineering, University of Tsukuba, 305-8573, Japan (email: s1011208@u.tsukuba.ac.jp), and Sadaaki Miyamoto is with the Department of Risk Engineering, University of Tsukuba, 305-8573, Japan (email: miyamoto@risk.tsukuba.ac.jp)

A. Fuzzy c-regression model

Fuzzy c-Regression Models (abbreviated FCRM) [7] inspired by fuzzy c-means clustering [2] finds c different regressions. Let us consider linear regression models with parameters $A = (\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_c)$, $\boldsymbol{\alpha}_i = (\alpha_i^1,\dots,\alpha_i^{p+1})^T$ and a variable $\boldsymbol{z} = (\boldsymbol{x}^T, 1)^T$ for simplification of description [8], [9]:

$$f(\boldsymbol{x};\boldsymbol{\alpha}_i) = \sum_{j=1}^{p} \alpha_i^j x^j + \alpha_i^{p+1} = \boldsymbol{\alpha}_i^T \boldsymbol{z} \quad (1)$$

FCRM thus uses

$$d_2(\boldsymbol{x}_k;\boldsymbol{\alpha}_i) = (y_k - f(\boldsymbol{x}_k;\boldsymbol{\alpha}_i))^2 \quad (2)$$

and minimizes

$$J_{\text{fcrm}}(U;A) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d_2(\boldsymbol{x}_k;\boldsymbol{\alpha}_i) \quad (3)$$

with fuzzifying parameter $m(>1)$. We note $m = 2$ is a standard choice, and it is adopted in this paper.

$U = (u_{ik})$ satisfies the next constraint

$$M_{\text{f}} = \{(u_{ik}) : u_{ik} \in [0,1], \sum_{i=1}^{c} u_{ik} = 1 \text{ for all } k\}$$

The objective function (3) should be alternatively minimized by U and A until convergence.

The minimizing solution for U with fixed A is given by

$$u_{ik} = \left[\sum_{j=1}^{c}\left(\frac{d_2(\boldsymbol{x}_k;\boldsymbol{\alpha}_i)}{d_2(\boldsymbol{x}_k;\boldsymbol{\alpha}_j)}\right)^{\frac{1}{m-1}}\right]^{-1}, \quad (4)$$

while optimum solution for A with fixed U is

$$\boldsymbol{\alpha}_i = \left(\sum_{k=1}^{n}(u_{ik})^m \boldsymbol{z}_k\boldsymbol{z}_k^T\right)^{-1}\sum_{k=1}^{n}(u_{ik})^m y_k \boldsymbol{z}_k. \quad (5)$$

B. Kernel function for regression

The method of positive definite kernels [13], [10], [1] for FCRM is called Fuzzy c-Kernel Regression Models (abbreviated FC-KRM) here. A similar method is studied [15], but the formularization is different from the following one. Assume that a kernel function is denoted by $g(\boldsymbol{x}_j, \boldsymbol{x}_\ell), j = 1,\dots,n, \ell = 1,\dots,n$, and associated parameters for regression is $A = (\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_c)$, $\boldsymbol{\alpha}_i = (\alpha_i^1,\dots,\alpha_i^n)^T$. Then the regression model is given by

$$f(\boldsymbol{x};\boldsymbol{\alpha}_i) = \sum_{k=1}^{n}\alpha_i^k g(\boldsymbol{x}_k, \boldsymbol{x}) \quad (6)$$

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

Let us introduce

$$\boldsymbol{y} = (y_1, \dots, y_n)^T, V_i = \begin{pmatrix} u_{i1} & 0 & \dots & 0 \\ 0 & u_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{in} \end{pmatrix}$$

Then the objective function to be minimized is

$$J_{\text{fckrm}}(U, A) = \sum_{i=1}^{c} (\boldsymbol{y} - G\boldsymbol{\alpha}_i)^T (V_i)^m (\boldsymbol{y} - G\boldsymbol{\alpha}_i) + \frac{1}{2} \sum_{i=1}^{c} \lambda_i \boldsymbol{\alpha}_i^T G \boldsymbol{\alpha}_i \quad (7)$$

where $g_{j\ell} = g(\boldsymbol{x}_j, \boldsymbol{x}_\ell)$ and $G = (g_{j\ell})$ with regularizing parameter $\lambda_i > 0$. The optimum solution for U is given by (4), while the minimizing solution for A with fixed U is as follows:

$$\boldsymbol{\alpha}_i = \left((V_i)^m G + \frac{\lambda_i}{2} I_n \right)^{-1} (V_i)^m \boldsymbol{y}. \quad (8)$$

Note also that the regression model is given by (6). The alternative optimization for U and A is repeated until convergence.

III. Fuzzy c-Support Vector Regression Models

We propose a combination of the support vector regression model (SVRM) [6] and FCRM, which we call Fuzzy c-Support Vector Regression Models (abbreviated as FC-SVRM). FC-SVRM uses the following instead of (2):

$$d_\epsilon(z_{ik}) = \begin{cases} z_{ik} - \epsilon & (\epsilon \le z_{ik}) \\ 0 & (-\epsilon \le z_{ik} < \epsilon) \\ -z_{ik} - \epsilon & (z_{ik} < -\epsilon) \end{cases}$$

where $z_{ik} = y_k - f(\boldsymbol{x}_k; \boldsymbol{\alpha}_i)$.

We introduce slack variables $\Xi = (\xi_{ik})$:

$$\xi_{ik} \ge y_k - f(\boldsymbol{x}_k; \boldsymbol{\alpha}_i) - \epsilon, \quad (9)$$

$$\xi_{ik} \ge 0, \quad (10)$$

$$\xi_{ik} \ge -(y_k - f(\boldsymbol{x}_k; \boldsymbol{\alpha}_i)) - \epsilon. \quad (11)$$

The objective function for FC-SVRM is then as follows:

$$J_{\text{fcsvrm}}(U, \Xi, A) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m \xi_{ik} + \frac{1}{2} \sum_{i=1}^{c} \lambda_i \boldsymbol{\alpha}_i^T G \boldsymbol{\alpha}_i. \quad (12)$$

We minimize (12) under the constraints (9)-(11). Langrange multipliers $\boldsymbol{\gamma}_i^+ = (\gamma_i^{+1}, \dots, \gamma_i^{+n}), \boldsymbol{\gamma}_i^- = (\gamma_i^{-1}, \dots, \gamma_i^{-n})$ corresponding (9),(11) respectively are introduced and we define that $\Gamma^+ = (\boldsymbol{\gamma}_1^+, \dots, \boldsymbol{\gamma}_c^+), \Gamma^- = (\boldsymbol{\gamma}_1^-, \dots, \boldsymbol{\gamma}_c^-)$. Then the Lagrangian function for the dual problem is:

$$W_{\text{fcsvrm}}(\Gamma^+, \Gamma^-) = -\frac{1}{2} \sum_{i=1}^{c} \frac{1}{\lambda_i} \sum_{k,\ell=1}^{n} (\gamma_i^{+k} - \gamma_i^{-k})(\gamma_i^{+\ell} - \gamma_i^{-\ell}) g_{k\ell} - \sum_{k=1}^{n} \sum_{i=1}^{c} \gamma_i^{+k}(-y_k + \epsilon) - \sum_{k=1}^{n} \sum_{i=1}^{c} \gamma_i^{-k}(y_k + \epsilon) \quad (13)$$

where the constraints for (13) are given by

$$0 \le \gamma_i^{+k} + \gamma_i^{-k} \le (u_{ik})^m. \quad (14)$$

The maximizing solution $\gamma_i^{+k}, \gamma_i^{-k}$ for (13) with constraints (14) should be used for regression (6) with fixed U. Thus we have

$$f(\boldsymbol{x}; \boldsymbol{\gamma}_i^+, \boldsymbol{\gamma}_i^-) = \frac{1}{\lambda_i} \sum_{\boldsymbol{x}_k \in SV} (\gamma_i^{+k} - \gamma_i^{-k}) g(\boldsymbol{x}_k, \boldsymbol{x})$$

where SV implies the set of $\boldsymbol{x}_k$ such that either γ_i^{+k} or γ_i^{-k} is non-zero. On the other hand, the solution of U with fixed $\boldsymbol{\gamma}_i^+, \boldsymbol{\gamma}_i^-$ is given by

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\xi_{ik}}{\xi_{jk}} \right)^{\frac{1}{m-1}} \right]^{-1} \quad (15)$$

FC-SVRM algorithm repeats the alternate optimization of $\boldsymbol{\gamma}_i^+, \boldsymbol{\gamma}_i^-$ and U until convergence. Each optimization realizes monotonic decrease of J_{fcsvrm}. For model selection, we use leave-one-out error [14].

IV. Sequential Extraction of Regression Models Using Support Vector Regression

A method of sequentially extracting regression models has been studied [11], [12], which is useful when the number of clusters (models) c is unknown. We propose Support Vector Sequential Regression Models (abbreviated as SVSRM) based on the noise cluster model [4], [5] in this section. A sequential regression model of (6) with $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$ is assumed. The objective function for SVSRM is

$$J_{\text{svsrm}}(U, \boldsymbol{\xi}; \boldsymbol{\alpha}) = \sum_{k=1}^{n} u_{1k} \xi_k + \sum_{k=1}^{n} u_{0k} \delta + \frac{\lambda}{2} \boldsymbol{\alpha}^T G \boldsymbol{\alpha}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ and a noise parameter $\delta (> 0)$ is used. Note that $U = (u_{ik})$ is crisp with the following constraint:

$$M_{\text{c}} = \{(u_{ik}) : u_{ik} \in \{0, 1\}, \sum_{i=0}^{1} u_{ik} = 1 \text{ for all } k\}$$

The Langrangian function for the dual problem is given by the next $W_{\rm svsrm}$:

$$
\begin{aligned}
&W_{\rm svsrm}(\boldsymbol{\gamma}^+, \boldsymbol{\gamma}^-) = \\
&-\frac{1}{2\lambda}\sum_{k,\ell=1}^{n}(\gamma^{+k}-\gamma^{-k})(\gamma^{-\ell}-\gamma^{-\ell})g_{k\ell} \\
&-\sum_{k=1}^{n}\gamma^{-k}(y_k+\epsilon)-\sum_{k=1}^{n}\gamma^{+k}(-y_k+\epsilon) \\
&+\sum_{k=1}^{n}u_{0k}\delta
\end{aligned} \quad (16)
$$

where $\boldsymbol{\gamma}^+ = (\gamma^{+1}, \dots, \gamma^{+n}), \boldsymbol{\gamma}^- = (\gamma^{-1}, \dots, \gamma^{-n})$; these γ^{+k}, γ^{-k} should satisfy the following constraints:

$$0 \le \gamma^{+k} + \gamma^{-k} \le u_{1k} \quad (17)$$

The optimum solution for γ^{+k}, γ^{-k} (with fixed U) is derived by maximizing (16) under the constraints (17), while the optimum solution for U with fixed γ^{+k}, γ^{-k} is as follows:

$$
\begin{aligned}
(u_{1k}, u_{0k}) = (1, 0) \Leftrightarrow \xi_k \le \delta \\
(u_{1k}, u_{0k}) = (0, 1) \Leftrightarrow \xi_k > \delta
\end{aligned} \quad (18)
$$

Note that we are handling crisp solutions for U.

Assume that the data set is X, while $X^{(t)}$ implies the subset of data for $t(\ge 1)$-th iteration, and the extracted cluster for t is $C^{(t)}$. Then the sequential extraction algorithm SVSRM is as follows:

SVSRM: Support Vector Sequential Regression Models.

SVSRM1. Let $X^{(0)} = X$ and set initial values for γ^{+k}, γ^{-k}. Put $t = 1$.

SVSRM2. For $X^{(t)}$, calculate (18) for all u_{1k}, and u_{0k}.

SVSRM3. For $X^{(t)}$, maximize (16) for γ^{+k}, γ^{-k} under the constraint (17).

SVSRM4. Test convergence criterion.
If not convergent, go to **SVSRM2**.
If convergent, Put $C^{(t)} = \{x_k : u_{1k} = 1\}$

SVSRM5. Let $X^{(t+1)} = X^{(t)} - C^{(t)}$.
If the number of objects in $X^{(t+1)}$ is sufficiently small, then stop.
Else put $t = t + 1$ and go to **SVSRM2**.

End of SVSRM.

At every extraction we change the parameters and calculate leave-one-out error [14]. We choose one having the minimum error.

V. Numerical Examples

Three examples are shown whereby we compare the results of the proposed methods with other existing methods.

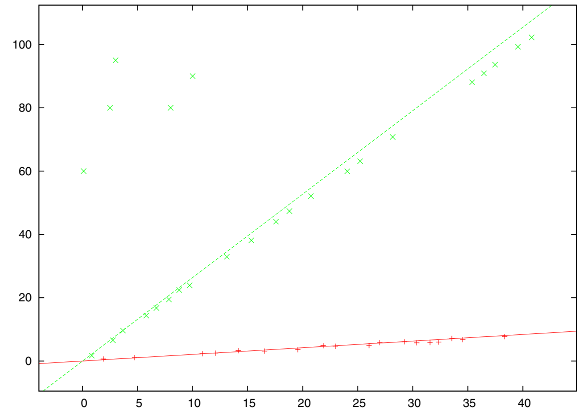

Fig. 1. An artificial data set with noises and the models derived from FC-KRM ($m = 2.0$, $c = 2$)

In the shown Figures 1-9, different clusters (i.e., regression models) are shown with different colors: Cluster 1 is given in red, cluster 2 is given in green, and cluster 3 is given in blue. The distinction with the colors is not important when c-regression models are considered, while the distinction is essential in sequential models, where the method first extracts cluster 1, secondly cluster 2, and so on. Objects in the respective clusters are shown by red +, green $\times$, and blue $*$. Non-support vectors are shown by ♦. Outliers (noises) after SVSRM stopped are shown with ♢.

A. An artificial data set for linear regression models

The methods of FC-SVRM and SVSRM are compared with the existing FC-KRM. Figures 1-3 show a set of linear regression models with noises. Comparing Fig. 1 and Fig. 2, we observe that the result of FC-SVRM is superior to that of FC-KRM, which shows the robustness of FC-SVRM. Figure 3 shows that SVSRM successfully extracts the two regression models.

B. A real data set

Figures 4-6 show the results from a real data set describing relations of GDP and energy comsumptions in Asia countries in different years [1], which is called GDP-Energy in the figure captions. The two proposed methods and FCRM give similar results, but SVSRM has more errors near origin as seen from Fig. 6.

C. Nonlinear models

The nonlinear models in Figs. 7-9 have been obtained by using the Gaussian kernel. The artificial data has two models and outliers (alias noises). The result in Fig. 7 from FC-KRM and that in Fig. 8 from FC-SVRM are both affected by outliers, while the result in Fig. 9 from SVSRM separates outliers from the models and thus better models have been obtained.

[1]This data set has been provided by Prof. Uchiyama at the University of Tsukuba.

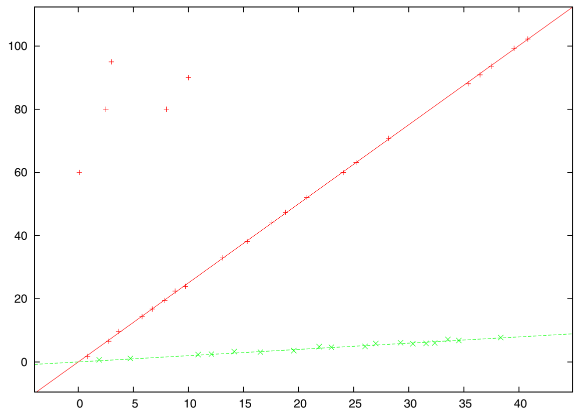

Fig. 2. An artificial data set with noises and the models derived from FC-SVRM ($m = 2.0$, $c = 2$)

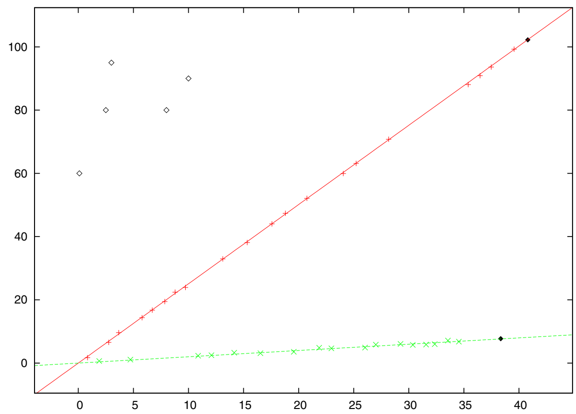

Fig. 3. An artificial data set with noises and the models derived from SVSRM ($\delta = 0.7$)

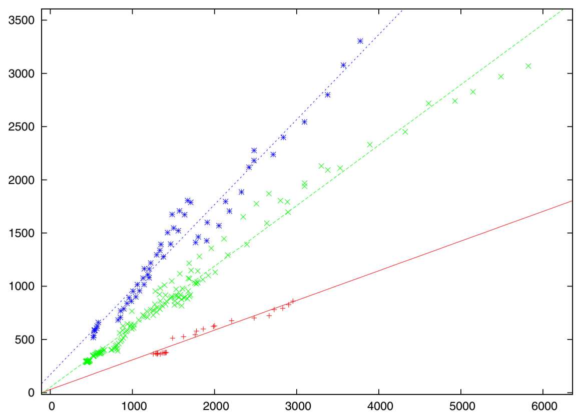

Fig. 4. GDP-Energy data and the models derived from FCRM ($m = 2.0$, $c = 3$)

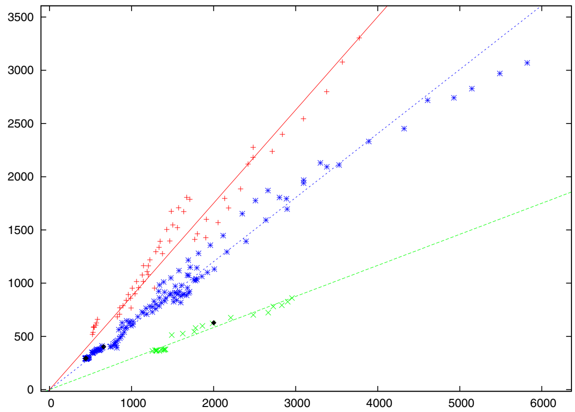

Fig. 5. GDP-Energy data and the models derived from FC-SVRM ($m = 2.0$, $c = 3$)

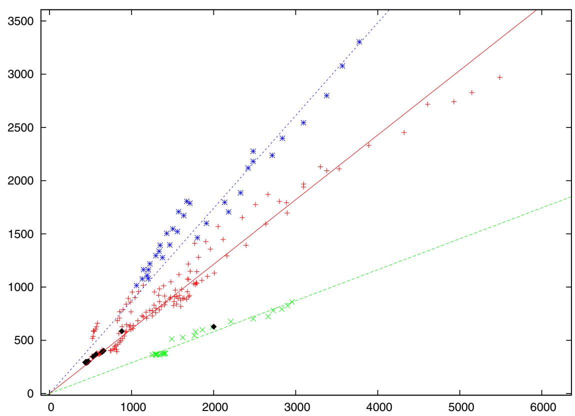

Fig. 6. GDP-Energy data and the models derived from SVSRM ($\delta = 350$)

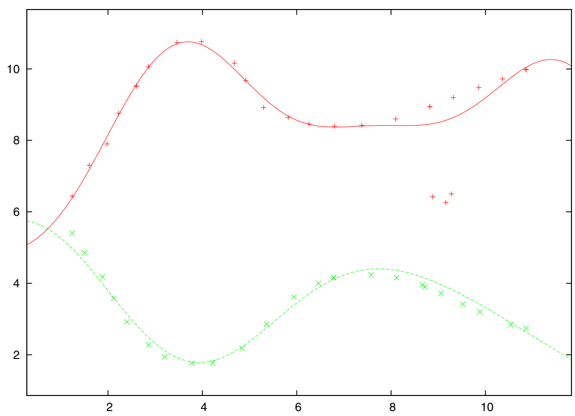

Fig. 7. An artificial data set with noises and the nonlinear models derived from FC-KRM with Gaussian kernel ($m = 2.0$, $c = 2$)

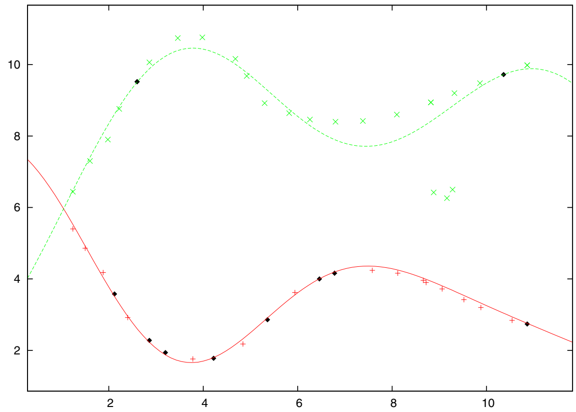

Fig. 8. An artificial data set with noises and the nonlinear models derived from FC-SVRM with Gaussian kernel ($m = 2.0$, $c = 2$)

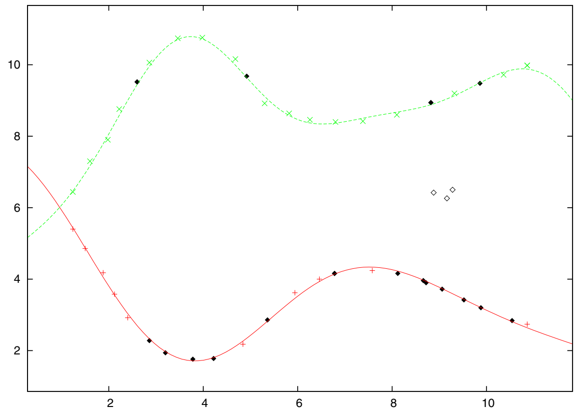

Fig. 9. An artificial data set with noises and the nonlinear models derived from SVSRM with Gaussian kernel ($\delta = 0.9$)

VI. Conclusion

We have proposed two methods for deriving multiple regression models using the support vector regression. One is fuzzy c-regression, while the second is sequential extraction of crisp regression models. Fuzzy version of the sequential regression model can be formulated without difficulty. We, however, did not yet find an advantage of the fuzzy model in the sequential cases.

There are many research possibilities for further investigation. First, a main drawback of the proposed model is the amount of computation, as quadratic optimization is needed, while existing methods have explicit formulas for optimum solutions. Second issue is the determination of appropriate values of the parameters. In spite of these problems, the present formulation opens a new perspective in the studies of both linear and nonlinear regression models. A specific problem as a future study is to introduce fuzziness into SVSRM which we are trying from now.

Acknowledgment

The authors would like to thank Prof. Yohji Uchiyama for proving us data of GDP and energy consumption. This work has partly been supported by the Grant-in-Aid for Scientific Research (KAKENHI) No. 23500269 and No. 26330270, JSPS, Japan.

References

[1] S. Akaho, *Kernel Multivariate Analysis*, Iwanami-shoten, Tokyo, 2008 (in Japanese).
[2] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, 1981.
[3] N. Cristianini, J. Shawe-Taylor: *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, 2000.
[4] R. N. Davé, Characterization and detection of noize in clustering, *Pattern Recognition Letters*, Vol.12, pp.657-664, 1991.
[5] R. N. Davé, R. Krishnapuram, Robust clustering methods: a unified view, *IEEE Transactions on Fuzzy Systems*, Vol.5, No.2, pp.270-293, 1997.
[6] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik: Support vector regression machines. In M. C. Mozer, M. I. Jordan, and T. Petsche, eds., *Advances in Neural Information Processing Systems 9*, pp. 155-161, MIT Press, Cambridge, MA, 1997.
[7] R. J. Hathaway, J. C. Bezdek: Switching regression models and fuzzy clustering, *IEEE Trans. on Fuzzy Systems*, Vol. 1, No. 3, pp. 195-204, 1993.
[8] S. Miyamoto, *Introduction to Cluster Analysis*, Morikita-shuppan, Tokyo, 1999 (in Japanese).
[9] S. Miyamoto, H. Ichihashi, K. Honda, *Algorithms for Fuzzy Clustering*, Springer, Berlin, 2008.
[10] B. Schölkopf, A. J. Smola, *Learning with Kernels*, the MIT Press, 2002.
[11] H. J. Tang: *Sequential Extraction of Clusters in Hard and Fuzzy Regression Models*, Master's Thesis, Graduate School of Systems and Information Engineering, 2012.
[12] H. J. Tang, S. Miyamoto: Semi-supervised sequential kernel regression models with pairwise constraints, V. Torra et al., eds., *Modeling Decisions for Artificial Intelligence*, 2013. LNAI 8234, Springer, Berlin, pp. 166-178, 2013.
[13] V. N. Vapnik: *Statistical Learning Theory*. Wiley, New York, 1998.
[14] A. Elisseeff, M. Pontil: Leave-one-out error and stability of learning algorithms with applications, *NATO science series sub series iii computer and systems sciences*, 190, 2003.
[15] Z. Sun: Fuzzy probability c-regression estimation based on least squares support vector machine, *Neural Information Processing*, Springer, Berlin, 2006.