Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 2, pp. 609 ~ 617

Switching Regression Analysis via Fuzzy LS-SVM[1]

Changha Hwang[2]

Abstract

A new fuzzy c-regression algorithm for switching regression analysis is presented, which combines fuzzy c-means clustering and least squares support vector machine. This algorithm can detect outliers in switching regression models while yielding the simultaneous estimates of the associated parameters together with a fuzzy c-partitions of data. It can be employed for the model-free nonlinear regression which does not assume the underlying form of the regression function. We illustrate the new approach with some numerical examples that show how it can be used to fit switching regression models to almost all types of mixed data.

Keywords : Fuzzy c-means clustering, Fuzzy c-regression, Least squares support vector machine, Outlier, Switching regression

1. Motivation and Fuzzy c-Regression

Switching regression models were developed as a way of allowing data to arise from a combination of two or more distinct data generation processes. Thus, instead of assuming that a single regression model can account for the data $\boldsymbol{m}_i = (\boldsymbol{x}_i, y_i)$, a switching regression model is specified by

$$y = f_k(\boldsymbol{x}\,;\,\boldsymbol{\beta}_k) + \varepsilon_k,\ \ 1 \le k \le c \tag{1}$$

In a statistical framework the optimal estimate of $\boldsymbol{\beta}_i$ depends on assumptions made about the distribution of random vectors ε_i. Commonly, the ε_i are assumed to be independently generated from some probability density function such as the Gaussian distribution with mean 0 and unknown variance. A switching regression

---

1) The present research was conducted by the research fund of Dankook university in 2005.
2) Division of Information and Computer Science, Dankook University, Seoul, 140-714, Korea
E-mail : chwang@dankook.ac.kr

model is discussed in varying detail in the literature such as Hamermesh(1970), Quandit(1972), Quandit and Ramsey(1978), Hathaway Bezdek(1993), Gallet(1999), Leski(2004) and so on. Various applications of switching regression models are often found in the economics and computer science. To solve the switching regression problem, mixture density estimation technique and fuzzy c-regression model(FCRM) are usually employed. In this paper we focus on FCRM switching regression analysis.

Conventional FCRM yields simultaneous estimates of parameters of c regression models together with a fuzzy c-partitioning of the data. Each regression model takes the functional form

$$y_i = f_k(\mathbf{x}_i, \boldsymbol{\beta}_k),\ k=1,\cdots,c,\ i=1,\cdots,n, \tag{2}$$

where $\mathbf{m}_j = (\mathbf{x}_j, y_j)$ denotes the i-th data point and the function f_k is parameterized by $\boldsymbol{\beta}_k \in R^{d_k}$. The membership degree $U_{ik} \in U$ is interpreted as a weight representing the extent to which the value predicted by the model $f_k(\mathbf{x}_i, \boldsymbol{\beta}_k)$ matches y_k. This prediction error is defined by

$$E_{ik} = (y_i - f_k(\mathbf{x}_i, \boldsymbol{\beta}_k))^2 \tag{3}$$

but other measures can be applied as well, provided they fulfill the minimizer property stated by Hathaway and Bezdek(1993). The family of objective functions for fuzzy c-regression models is defined by

$$E_m(U, \{\boldsymbol{\beta}_k\}) = \sum_{k=1}^{c} \sum_{i=1}^{n} U_{ik}^m E_{ik}, \tag{4}$$

where $m \in (1, \infty)$ denotes a weighting exponent which determines the fuzziness of the resulting clusters. One possible approach to the minimization of the objective function (4) is the group coordinate minimization method that results in the following algorithm:

- **Initialization** Given a set of data $\{\mathbf{x}_i, y_i\}_{i=1}^n$ specify c, the structure of the regression models (3) and the error measure (4). Choose a weighting exponent $m \in (1, \infty)$ and a termination tolerance $\varepsilon > 0$. Initialize the partition matrix U randomly.

- **Repeat** for $r = 1, 2, \cdots$

Step 1 Calculate values for the model parameters $\boldsymbol{\beta}_k$ that minimize the cost function

$$E_m(U, \{\beta_k\}).$$

Step 2 Update the partition matrix

$$U_{ik}^{(r)} = \frac{1}{\sum_{l=1}^{c}\left(\frac{E_{ik}}{E_{lk}}\right)^{1/(m-1)}}, \quad 1 \le k \le c, \quad 1 \le i \le n.$$

until $\| U^{(r)} - U^{(r-1)} \| < \varepsilon$.

A specific situation arises when the regression functions f_k are linear in the parameters β_k. In this case, the parameters can be obtained as a solution of a set of weighted least squares problem where the membership degrees of the fuzzy partition matrix U serve as the weights. The optimal parameters $\beta_k^{(r)}$ are then computed as follows,

$$\beta_k^{(r)} = (X' D_k X) X' D_k y,$$

where $X = \{1_n, (x_1, \cdots, x_n)'\}$ and D_k is the diagonal matrix with $U_{ik}^{(r)}$ as its i-th diagonal element.

In this paper, we propose a new FCRM for switching regression analysis, which is basically model-free method based on weighted least squares support vector machine(LS-SVM) and thus can be applied to almost all types of mixed data.

2. Weighted LS-SVM

We now slightly modify LS-SVM proposed by Suykens and Vanderwalle(1999). Let the data set be denoted by $\{x_i, y_i\}_{i=1}^n$, with each the input $x_i \in R^d$ and output $y_i \in R$. We can assume the functional form of unknown regression function f for given input vector x by

$$f(x) = w' \phi(x) + b,$$

where b is a bias term and w is an appropriate weight vector. Here the feature mapping function $\phi(\cdot): R^d \to R^{d_f}$ maps the input vector space to the higher dimensional feature space where the dimension d_f is greater or equal to d of input vector space. The objective function is defined with a LS-SVM regularization parameter γ as

$$L(w, b, e) = \frac{1}{2} w' w + \frac{\gamma}{2} \sum_{i=1}^{n} \theta_i e_i^2$$

subject to equality constraints

$$y_i - w' \phi(x_i) - b = e_i, \quad i = 1, \cdots, n,$$

where $e = (e_1, \cdots, e_n)'$ is a vector of measures of the prediction error and θ_i is a weight assigned to the i-th data point. To find minimizers of the objective function, we can construct the Lagrangian function as follows,

$$L(w, b, e: \alpha) = \frac{1}{2} w' w + \frac{\gamma}{2} \sum_{i=1}^{n} \theta_i e_i^2 - \sum_{i=1}^{n} \alpha_i (w' \phi(x_i) + b - e_i - y_i)$$

where α_i's are the Lagrange multipliers.

By employing KKT(Karush-Kuhn-Tucker) conditions for the Lagrangian function we can easily obtain the equations for optimal solutions by partial differentiating the Lagrangian function over $\{w, b, e_k, \alpha_k\}$, which leads to the solution of a linear equation

$$\begin{bmatrix} 0 & 1' \\ 1 & K + \gamma^{-1} \Theta^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}.$$

Here $K = \{K_{kl}\}$ with $K_{kl} = \phi(x_k)' \phi(x_l) = K(x_k, x_l)$, $k, l = 1, \cdots, n$ and Θ is diagonal matrix with θ_i as its i-th element. $K(\cdot, \cdot)$ is an appropriate kernel function. Several choices of the kernel function $K(\cdot, \cdot)$ are possible. By solving the linear system, Lagrange multipliers α_i, $i = 1, \cdots, n$, and bias b can be obtained. With these estimates we can get the predicted value $f(x)$ for the given x as follows

$$f(x) = \sum_{i=1}^{n} K(x, x_i) \alpha_i + b.$$

The linear regression model can be regarded as a special case of nonlinear regression model. By using an identity feature mapping function ϕ in nonlinear regression model, that is, $K(x_k, x_l) = x_k' x_l$, it reduces to linear regression model.

3. FCRM using Fuzzy LS-SVM

By adding a weight representing the membership degree of being considered as outlier to each data point on LS-SVM, we can obtain a new objective function of FCRM based on weighted LS-SVM as follows:

$$L(w_k, b_k, e_{ik} \mid U, V) = \frac{1}{2}\sum_{k=1}^{c} w_k' w_k + \frac{\gamma}{2}\sum_{i=1}^{n}\sum_{k=1}^{c}\frac{U_{ik}^m}{V_i}e_{ik}^2$$

with constraint

$$y_i - w_k' \phi(x_i) - b_k = e_{ik},\ i=1,\cdots,n,\ k=1,\cdots,c,$$

where w_k is an appropriate weight vector for the k-th regression model and γ is a regularization parameter. $V_i \in V$ is a weight representing the membership degree of the i-th data point being considered as outlier. The value of V_i can be interpreted as the membership degree of the i-th data point not belonging to either regression model. We define $V_i \in V$ as

$$V_i = v\left(\sum_{k=1}^{c} U_{ik}^m e_{ik}^2\right)^{\frac{1}{2}},\ i=1,\cdots,n,$$

where the value of v is determined so that $\sum_{i=1}^{n}\frac{1}{V_i} = n$ (number of samples), the value of U_{ik} represents the membership degree of the i-th data point belonging to the k-th regression model and is inversely proportional to the value of e_{ik}^2.. We call this LS-SVM based on these weights as fuzzy LS-SVM.

Then $V_i \in V$, $U_{ik} \in U$ and $f_k(x_i)$ are simultaneously obtained as follows:

■ **Initialization** Given a set of data $\{x_i, y_i\}_{i=1}^n$ specify c, the structure of the regression models and the error measure. Choose a weighting exponent $m \in (1, \infty)$ and a termination tolerance $\varepsilon > 0$. Initialize the partition matrix U and V randomly.

■ **Repeat** for $r = 1, 2, \cdots$

Step 1 Calculate values of the k-th regression model $f_k(x) = K\alpha_k + b_k$ for $k=1,\cdots,c$ with $\{\alpha_k, b_k\}$ from the minimization of the objective function $L(w_k, b_k, e_{ik} \mid U^{(r)}, V^{(r)})$.

Step 2 Update $V^{(r-1)}$ to $V^{(r)}$ with $U_{ik} = U_{ik}^{(r)}$. Update $U^{(r-1)}$ to $U^{(r)}$ to satisfy, with $e^2_{ik} = e^2_{ik}(y_i, f_k(\mathbf{x}_i))$,

if $e^2_{ik} > 0$ for $k = 1, \cdots, c$,

$$U_{ik} = \frac{1}{\sum_{l=1}^{c} \left(\frac{e^2_{ik}}{e^2_{lk}} \right)^{1/(m-1)}}$$

otherwise

$$U_{ik} = 0 \text{ if } e^2_{ik} > 0, \text{ and}$$

$$U_{ik} \in (0, 1) \text{ with } \sum_{k=1}^{c} U_{ik} = 1.$$

until $\| U^{(r)} - U^{(r-1)} \| < \varepsilon$.

4. Numerical Studies and Discussion

We present two examples to illustrates how well the new algorithm detects outliers and fits the mixed data to the nonlinear regression models.

Example 1. We consider the c=2 switching regression models given by,

$$y = f_1(x, \beta) + \varepsilon_1 = -2x + \varepsilon_1,$$
$$y = f_2(x, \beta) + \varepsilon_2 = \sin(2x) + \varepsilon_2,$$

where $\varepsilon_k \sim i.i.d.\ N(0, 0.1^2)$, $k = 1, 2$. We generate 50 data points to illustrate the performance of algorithm and add 4 ouliers to the models.

<Figure 1> Experimental Results of Example 1

The identity kernel function $K(x_i, x_j) = x_i x_j$ for the linear regression and the radial basis kernel function $K(x_i, x_j) = \exp(-\| x_i - x_j \|^2/2.5)$ for the nonlinear regression are used to estimate the regression functions of the c=2 switching regression models. The regularization parameter γ is set to 100 in this example. The left plot of Figure 1 shows two true regression functions and data set consisting of 54 data points used in the example. The right plot shows how each data point is labeled after precise iteration of algorithm. For the data set we obtain the partition matrix U and the outlier detecting vector V, according to their values we partition the data points belonging to each model and identify the ouliers. We can recognize the proposed FCRM works quite well.

Example 2. We now consider the c=2 switching regression models given by,

$$y = f_1(x, \beta) + \varepsilon_1 = x + \cos(2\pi x) + \varepsilon_1,$$
$$y = f_2(x, \beta) + \varepsilon_2 = \sin(2x + 1) + \varepsilon_2,$$

where $\varepsilon_1 \sim i.i.d.\ N(0, 0.05^2)$, $\varepsilon_2 \sim i.i.d.\ N(0, 0.1^2)$. As in Example 1, we generate 50 data points to illustrate the performance of algorithm and add 4 ouliers to the models. The radial basis kernel function $K(x_i, x_j) = \exp(-\| x_i - x_j \|^2/1.5)$ are used to estimate the regression functions of the c=2 switching regression models. The regularization parameter γ is set to 100 in this example.

<Figure 2> Experimental Results of Example 2

The left plot of Figure 2 shows two true regression functions and data set consisting of 54 data points. The right plot shows the estimated regression functions of models and the partitioned data points fitted to each model.

We now illustrate further how well the proposed FCRM identifies the ouliers in two examples. We generate 200 data sets, each of which consists of 50 normal data points and 4 outliers as before. The outliers can be identified by the 4 largest values of V in each example. Table 1 shows the average value of normal data points and outliers identified by the values of V and number of correct identifications of each outlier by the largest 4 values of V among 200 data sets used in each example, respectively. From Table 1 we can see that the outlier information of the data set is easily investigated through the outlier weight vector. If we want to select m data points from the data set as the outlier candidates, the first m data points corresponding to the first m largest data point in the outlier weight vector V.

<Table 1> Identification results of outliers

	$\overline{V}_{\text{in}}$	$\overline{V}_{\text{out}}$	no. of correct identifications			
			outlier1	outlier2	outlier3	outlier4
Example 1	28.0695	210.3426	200	196	200	199
Example 2	199.5798	1087.6172	149	190	199	166

To conclude, we can say the proposed FCRM fits data quite well. This FCRM can detect quite well outliers in switching regression models while yielding the simultaneous estimates of the associated parameters together with a fuzzy c -partitions of data. We can recognize that it can be employed for the model-free nonlinear regression which does not assume the underlying form of the regression function.

References

1. Caudill, S. and Acharyal, R.(1998). Maximum-likelihood estimation of a mixture of normal regressions: Starting values and singularities, *Communications in Statistics- Simulation and Computation*, **27**, 3, 667-674.
2. Gallet, C. A.(1999). Gradual switching regression estimates of alcohol demand elasticities, *Applied Economics Letters*, **6**, 377-379.
3. Hamermesh, D. S.(1970). Wage bargains, threshold effects, and the Phillips curve, *Quarterly Journal of Economics*, **84**, 501-517.
4. Hathaway, R. and Bezdek, J.(1993). Switching regression models and fuzzy clustering, *IEEE Transactions on Fuzzy Systems*, **1**, 195-204.
5. Leski, J. M.(2004). Epsilon-insensitive fuzzy c-regression models:

introduction to epsilon-insensitive fuzzy modeling, *IEEE Trans Syst Man Cybern B*, 34, 4-15.
6. Quandit, R.(1972). A new approach to estimating switching regressions, *Journal of American Statistical Associations*, 67, 306-310.
7. Quandit, R. and Ramsey, J.(1978). Estimating mixtures of normal distributions and switching regressions, *Journal of American Statistical Associations*, 73, 730-752.
8. Suykens, J. and Vanderwalle, J.(1999). Least squares support vector machine classifier, *Neural Processing Letters*, 9, 293-300.

[received date : Apr. 2006, accepted date : May. 2006]