Nonlinear regression modeling and detecting change point via the relevance vector machine

Tateishi, Shohei
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

出版情報：MI Preprint Series. 2010-5, 2010-12-09. Springer
バージョン：

MI Preprint Series

Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

Nonlinear regression modeling and detecting change point via the relevance vector machine

Shohei Tateishi
& Sadanori Konishi

MI 2010-5

(Received January 19, 2010)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Nonlinear regression modeling and detecting change points via the relevance vector machine

Shohei Tateishi[1] and Sadanori Konishi[2]

[1] *Graduate School of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

[2] *Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

s-tateishi@math.kyushu-u.ac.jp   konishi@math.kyushu-u.ac.jp

Abstract

We consider the problem of constructing nonlinear regression models in the case that the structure of data has discontinuities at some unknown points. We propose two-stage procedure in which the change points are detected by RVM at the first stage, and the smooth curve are effectively estimated along with the technique of regularization method at the second. In order to select tuning parameters in the regularization method, we derive a model selection and evaluation criterion from information-theoretic viewpoints. Simulation results and real data analyses demonstrate that our methodology performs well in various situations.

Key Words and Phrases: Basis expansion, Change point, Information criterion, Relevance vector machine, Nonlinear regression, Regularization.

1 Introduction

Nonlinear regression model based on basis expansions is a useful tool to analyze data with complex structure. The essential idea behind basis expansions is to express a regression function as a linear combination of known functions, called basis functions (Konishi and Kitagawa, 2008; Hastie *et al.*, 2009). In constructing the model, the basis functions are chosen according to the structure of data. For example, B-splines (Eilers and Marx, 1996; de Boor, 2001; Imoto and Konishi, 2003) and radial basis functions (Bishop, 1995; Kawano and Konishi, 2007; Ando *et al.*, 2008), In particular, Gaussian basis functions have been widely used to construct nonlinear regression models. In applying these models, it is assumed that the structure of data is smooth.

However, the underlying true structure which is generating data cannot be smooth at some points where jump discontinuity may occur. Thus, the application of a usual nonlinear regression model described above will lead difficulty of obtaining effective information from the data in which the mean structure is suddenly changed.

Roughly speaking, the approaches for the change point problems can be classified whether one change point exists or more than one. As examples of the former style, kernel-based estimation methods have been proposed by Muller (1992) and local polynomial methods have been used by Loader (1996). As examples of the latter style, Qiu (2003) and Gijbels *et al.* (2007) proposed a jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. Although Qiu (2003) and Gijbels *et al.* (2007) are free from assumption of knowing the number of jumps, they leads very rough result functions even in continuity regions.

In order to overcome this difficulty, we propose the method of appropriately estimating a nonlinear structure with the change points by applying RVM (Tipping, 2001) and regularization method. We present a two-stage procedure to fit discontinuous regression curve.

In the first stage, RVM is applied to the regression model with discontinuous basis functions, and the candidates for the change points are detected. When using RVM, most coefficients in the model are estimated exactly zero so that we can narrow down candidates for change points. In the second stage, the regularization method is applyed to nonlinear regression model with normal Gaussian basis functions in order to get the smooth curve expect for change points. The regularization or shrinkage method has been widely used to overcome unstability and ill-posed problems arising in a maximum likelihood or a least squares procedure, and it has been proved successful in several fields, including image processing and machine learning (see, e.g., Hastie *et al.*, 2009; Bishop, 2006). It imposes a penalty with respect to parameters of objective functions that are utilized in optimization problems, and various kinds of penalties have been proposed (Frank and Friedman, 1993; Tibshirani, 1996; Fan and Li, 2001; Candes and Tao, 2007). One of the most commonly used penalty methods is ridge regression (Hoerl and Kennard, 1970), which imposes an L_2 norm penalty on regression coefficients. The ridge regression achieves good prediction performance through a bias-variance trade-off.

It is a crucial issue to determine the tuning parameters, including the number of basis functions, a smoothing parameter and a hyperparameter associated with Gaussian basis functions. To choose these parameters, we derive model selection criterion from information-theoretic viewpoint. The proposed nonlinear modeling procedure is investigated through the numerical examples.

This paper is organized as follows. Section 2 describes the framework of nonlinear regression model based on basis expansions. In Section 3 we present a method of detecting change points by using RVM. Section 4 provides the discontinuous nonlinear regression model. In section 5 we introduce regularization method imposing L_2 norm penalty. Section 6 provides a model selection criterion for evaluating statistical models estimated by the regularization method. In Section 7 we investigate the performance of our nonlinear regression modeling techniques through Monte Carlo simulations and real data analyses. Some concluding remarks are presented in Section 8.

2 Nonlinear regression model with basis expansions

Suppose that we have n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, 2, \cdots, n\}$, where y_α are random response variables and x_α are explanatory variables. We consider the regression model

$$y_\alpha = u(x_\alpha) + \epsilon_\alpha, \quad \alpha = 1, 2, \cdots, n, \tag{1}$$

where $u(\cdot)$ is an unknown smooth function and ϵ_α are independently, normally distributed with mean zero and variance σ^2. It is assumed that the function $u(\cdot)$ can be expressed as a linear combination of basis functions $b_j(x)$ $(j = 1, 2, \cdots, m)$ in the form

$$u(x; \boldsymbol{w}) = w_0 + \sum_{j=1}^{m} w_j b_j(x) = \boldsymbol{w}^T \boldsymbol{b}(x) \ , \tag{2}$$

where $\boldsymbol{b}(x) = (1, b_1(x), \cdots, b_m(x))^T$ is a vector of basis functions and $\boldsymbol{w} = (w_0,\ w_1, \cdots,\ w_m)^T$ is an unknown coefficient parameter vector. A variety of basis functions are used according to the structure of data.

One of the many basis functions is Gaussian basis function given by

$$b_j(x) = \exp\left\{-\frac{(x-c_j)^2}{2h_j^2}\right\}, \quad j = 1, 2, \cdots, m, \tag{3}$$

where c_j is the center of the basis function, h_j^2 is a parameter that determines the dispersion and $||\cdot||$ is the Euclidian norm. However, basis functions (3) often yield inadequate results because of the lack of overlapping among basis functions. In order to overcome this problem, Ando *et al.* (2008) proposed the use of Gaussian basis functions with a hyperparameter, i.e. functions of the form

$$b_j(x) = \exp\left\{-\frac{(x-c_j)^2}{2\nu h_j^2}\right\}, \quad j = 1, 2, \cdots, m, \tag{4}$$

where ν is a hyperparameter that adjusts the dispersion of basis functions. Ando *et al.* (2008) showed that nonlinear models with these basis functions were effective in capturing the information from the data.

However, the models with these basis functions will lead to smooth curve estimates, even though change points are present. Therefore, they will be oversmoothed and change points will not be visible in resulting curve. In order to overcome this problem, we use discontinuous basis functions.

3 Detecting change points and estimation

For n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, \cdots, n\}$, the nonlinear regression model based on basis functions $\phi_j(x)$ $(j = 1, \cdots, n)$ is expressed as

$$y_\alpha = \boldsymbol{w}_c^T \boldsymbol{\phi}(x_\alpha) + \epsilon_\alpha, \ \alpha = 1, \cdots, n, \tag{5}$$

where $\boldsymbol{\phi}(x_\alpha) = (1, \phi_1(x_\alpha), \cdots, \phi_n(x_\alpha))^T$, $\boldsymbol{w}_c = (w_{0c},\ w_{1c}, \cdots,\ w_{nc})^T$ and ϵ_α are error terms. If the error terms ϵ_α are independently and normally distributed with mean 0 and variance β^{-1} $(\beta > 0)$, the nonlinear regression model (5) has a probability density function

$$f(y_\alpha|\boldsymbol{w}_c, \beta) = \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp\left[-\frac{\{y_\alpha - \boldsymbol{w}_c^T\boldsymbol{\phi}(x_\alpha)\}^2}{2\beta^{-1}}\right], \ \alpha = 1, \cdots, n. \tag{6}$$

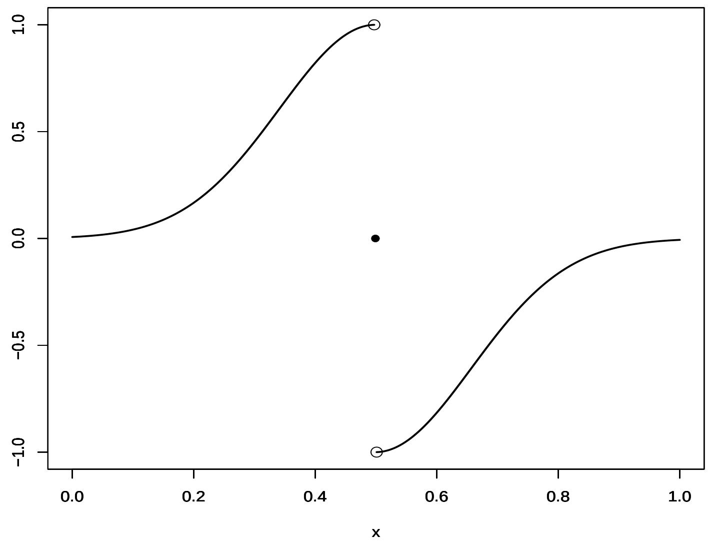

Fig. 1: The basis function $\phi(x)$. It is discontinuous at the center of the basis function.

For a explanatory variable x, we use discontinuous Gaussian basis functions given by

$$\phi_j(x) = \begin{cases} \exp\left(-\dfrac{||x-x_j||^2}{h_c^2}\right), & (x < x_j) \\ 0, & (x = x_j) \\ -\exp\left(-\dfrac{||x-x_j||^2}{h_c^2}\right), & (x > x_j) \end{cases} ,j = 1, 2, \cdots, n, \tag{7}$$

and Figure 1 shows this basis function $\phi(x)$. The discontinuous Gaussian basis function $\phi(x)$ flips at the center of the basis function, and then the point whose absolute value of coefficient is large can be considered to be the candidate for change point. Because, it means the points behind and before the center greatly stop away from each other.

Next we suppose that the parameter vector $\boldsymbol{w}$ has a Gaussian prior density

$$\pi(\boldsymbol{w}_c|\boldsymbol{\alpha}) = (2\pi)^{-\frac{n}{2}}|A|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{w}_c^T A \boldsymbol{w}_c\right), \tag{8}$$

where $\boldsymbol{\alpha} = (\alpha_0, \cdots, \alpha_n)^T$ is an $(n+1)$ hyperparameter vector and $A = \mathrm{diag}(\alpha_0, \cdots, \alpha_n)$. Using Bayes' theorem, we see that the posterior distribution for the weights $\boldsymbol{w}$ has Gaussian density

$$\pi(\boldsymbol{w}_c|\boldsymbol{y}, \boldsymbol{\alpha}, \beta) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{w}_c - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{w}_c - \boldsymbol{\mu})\right\},$$

where the posterior covariance matrix and mean are respectively

$$\Sigma = (\beta\Phi^T\Phi + A)^{-1}, \quad \boldsymbol{\mu} = \beta\Sigma\Phi^T\boldsymbol{y}, \tag{9}$$

where $\Phi = (\boldsymbol{\phi}(x_1)^T, \cdots, \boldsymbol{\phi}(x_n)^T)^T$.

The values of hyperparameters $\boldsymbol{\alpha}, \beta$ are determined by using maximization of marginal likelihood function

$$p(\boldsymbol{y}|\boldsymbol{\alpha}, \beta) = \int f(\boldsymbol{y}|\boldsymbol{w}_c, \beta)p(\boldsymbol{w}_c|\boldsymbol{\alpha})d\boldsymbol{w}_c, \tag{10}$$

where $f(\boldsymbol{y}|\boldsymbol{w}_c, \beta) = \prod_{\alpha=1}^n f(y_\alpha|\boldsymbol{w}_c, \beta)$. Setting the derivatives of marginal likelihood to zero, we obtain estimators of $\boldsymbol{\alpha}, \beta$ given by

$$\hat{\alpha}_j = \frac{\gamma_j}{\mu_j^2}, \quad \hat{\beta}^{-1} = \frac{||\boldsymbol{y} - \Phi\boldsymbol{\mu}||^2}{n - \sum_k \gamma_k}, \quad j = 0, \cdots, n, \ k = 0, \cdots, n. \tag{11}$$

where $\gamma_j = 1 - \alpha_j\Sigma_{jj}$, μ_j is $(j+1)$-th element of $\boldsymbol{\mu}$ and Σ_{jj} is $(j+1)$-th diagonal element of Σ. Because these estimators depend on each other, we need re-estimation of (9) and (11). As mentioned above, the technique for estimation by the sequential computation based on the maximizing marginal likelihood using ARD prior (Neal, 1996) is known as relevance vector machine (RVM; Tipping, 2001). Using RVM, most coefficients are estimated to be exactly zero. It can be thought that the point corresponding to the coefficient estimated to be 0 except for intercept is a candidate for the change point. So, we can narrow down candidates for change points, and we set up the vector of discontinuous basis functions those have non-zero coefficients given by

$$\boldsymbol{\phi}_{\hat{T}}(x_\alpha) = (\phi_{\tau_1}(x), \cdots, \phi_{\tau_{n_t}}(x))^T, \tag{12}$$

where $\hat{T} = \{\tau_1, \cdots, \tau_{n_t}\}$ is a set of candidates for change points, n_t is the number of them, and ϕ_{τ_k} $(k = 1, \cdots, n_t)$ is a discontinuous basis function (7) whose center is τ_k.

4 Discontinuous nonlinear regression model

Although the discontinuous basis functions help to detecting change points, the smooth curve cannot be gained by using only such basis functions. Therefore, we assume the nonlinear model involving continuous basis functions as below.

For n independent observations $\{(y_\alpha, x_\alpha); \alpha = 1, \cdots, n\}$, the nonlinear regression model based on basis functions $b_j(x)$ $(j = 1, \cdots, n)$ given in Section 2 is expressed as

$$y_\alpha = \boldsymbol{w}^T \boldsymbol{b}(x_\alpha) + \epsilon_\alpha, \ \alpha = 1, \cdots, n, \tag{13}$$

where $\boldsymbol{b}(x_\alpha) = (1, \psi_1(x_\alpha), \cdots, \psi_m(x_\alpha), \boldsymbol{\phi}_{\hat{T}}(x_\alpha)^T)^T$, $\boldsymbol{w} = (w_0, w_1, \cdots, w_{m+n_t})^T$ and ϵ_α are error terms. If the error terms ϵ_α are independently and normally distributed with mean 0 and variance σ^2, the nonlinear regression model (13) has a probability density function

$$f(y_\alpha|\boldsymbol{w}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y_\alpha - \boldsymbol{w}^T \boldsymbol{b}(x_\alpha)\}^2}{2\sigma^2}\right], \ \alpha = 1, \cdots, n. \tag{14}$$

For smooth parts in estimated curve except for the change points, we use Gaussian basis functions (4) as basis function $\psi(x)$.

Unknown parameters in the regression model (13) include the coefficient parameters w_j $(j = 1, \cdots, m)$, the centers c_j and dispersion parameters h_j^2. In order to avoid local minimum and identification problems (Moody and Darken, 1989), the centers c_j and dispersion h_j^2 are determined by using the k-means clustering algorithm. The data set of observations of the explanatory variables $\{x_1, \cdots, x_n\}$ is divided into m clusters $\{C_1, \cdots, C_m\}$; centers c_j and dispersions h_j^2 are determined by

$$\hat{c}_j = \frac{1}{n_j} \sum_{x_\alpha \in C_j} x_\alpha, \quad \hat{h}_j^2 = \frac{1}{n_j} \sum_{x_\alpha \in C_j} ||x_\alpha - c_j||^2, \tag{15}$$

where n_j is the number of observations included in the the j-th cluster C_j. Replacing c_j and h_j^2 in equation (3) by $\hat{c}_j$ and $\hat{h}_j^2$ respectively, we obtain a set of m basis functions

$$\psi_j(x; \hat{c}_j, \hat{h}_j^2) = \exp\left(-\frac{||x - \hat{c}_j||^2}{2\nu \hat{h}_j^2}\right), \quad j = 1, 2, \cdots, m. \tag{16}$$

And then, the coefficient parameters w_j $(j = 0, 1, \cdots, m)$ are estimated by the maximum penalized likelihood method.

5 Estimation based on regularization

The maximum likelihood estimates of the coefficient vectors $\boldsymbol{w}$ and σ^2 are respectively given by

$$\hat{\boldsymbol{w}} = (B^T B)^{-1} B^T \boldsymbol{y}, \qquad \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - B\hat{\boldsymbol{w}})^T (\boldsymbol{y} - B\hat{\boldsymbol{w}}),$$

where $B = (\boldsymbol{b}(x_1)^T, \cdots, \boldsymbol{b}(x_n)^T)^T$ and $\boldsymbol{y} = (y_1, \cdots, y_n)^T$. However, when fitting a nonlinear model to data with a complex structure the maximum likelihood method often yields unstable estimates and leads to overfitting. We therefore estimate $\boldsymbol{w}$ and σ^2 by the method of regularization. Instead of using the log-likelihood function, we consider maximizing the penalized log-likelihood function

$$l_\lambda(\boldsymbol{\theta}) = \sum_{\alpha=1}^{n} \log f(y_\alpha|\boldsymbol{w}, \sigma^2) - \frac{n\lambda}{2}\boldsymbol{w}^T K \boldsymbol{w}, \tag{17}$$

where $\boldsymbol{\theta} = (\boldsymbol{w}^T, \sigma^2)^T$, $\lambda\,(>0)$ is a smoothing parameter that controls the smoothness of the fitted model and K is a known $(m + n_t + 1)$-th square matrix (Konishi and Kitagawa, 2008). The typical form of K is given by $K = I_{m+n_t+1}$ for the identity matrix or $K = D_2^T D_2$ for a second-order difference matrix. Then, the maximum penalized likelihood estimates of $\boldsymbol{w}$ and σ^2 are respectively given by

$$\hat{\boldsymbol{w}} = (B^T B + n\lambda\hat{\sigma}^2 K)^{-1} B^T \boldsymbol{y}, \quad \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - B\hat{\boldsymbol{w}})^T(\boldsymbol{y} - B\hat{\boldsymbol{w}}). \tag{18}$$

Note that these estimators depend on each other. Therefore, we provide an initial value for the variance $\sigma^2_{x(0)}$ first, then $\hat{\boldsymbol{w}}$ and $\hat{\sigma}^2_x$ are updated until convergence. The ridge estimators continuously shrink the coefficients as λ increases.

6 Model selection criteria

The statistical model estimated by the regularization method depends upon the number of basis functions m, the value of the smoothing parameter λ and the value of the hyperparameter ν in the Gaussian basis functions. It is a crucial issue to determine these values appropriately.

Konishi and Kitagawa (1996) introduced evaluation criteria of statistical models that can be applied to the evaluation of statistical models estimated by various types of estimation procedures such as the robust and penalized likelihood procedures. By using the result, the model selection criterion for evaluating the statistical model constructed by Gaussian basis functions is given by

$$\text{GIC} = n\{\log(2\pi) + 1\} + n\log\hat{\sigma}^2 + 2\text{tr}\{R^{-1}Q\}, \tag{19}$$

where R and Q are $(m+n_t+2)$-th square matrices and are, respectively, given by

$$R = \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} B'B + n\lambda\hat{\sigma}^2 K & \frac{1}{\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{\hat{\sigma}^2}\mathbf{1}_n'\Lambda B & \frac{n}{2\hat{\sigma}^2} \end{bmatrix}, \tag{20}$$

$$Q = \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} \frac{1}{\hat{\sigma}^2}B'\Lambda^2 B - \lambda K\hat{\boldsymbol{w}}\mathbf{1}_n'\Lambda B & \frac{1}{2\hat{\sigma}^4}B'\Lambda^3\mathbf{1}_n - \frac{1}{2\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^4}\mathbf{1}_n'\Lambda^3 B - \frac{1}{2\hat{\sigma}^2}\mathbf{1}_n'\Lambda B & \frac{1}{4\hat{\sigma}^6}\mathbf{1}_n'\Lambda^4\mathbf{1}_n - \frac{n}{4\hat{\sigma}^2} \end{bmatrix} \tag{21}$$

with $\mathbf{1}_n = (1, \cdots, 1)^T$ and $\Lambda = \mathrm{diag}(y_1 - \hat{\boldsymbol{w}}'\boldsymbol{b}(x_1), \cdots, y_n - \hat{\boldsymbol{w}}'\boldsymbol{b}(x_n))$. We select the optimal value of the number of basis functions, a regularization parameter and a hyperparameter that minimize GIC.

7 Numerical examples

In this section, Monte Carlo simulations and real data analysis were conducted to investigate the effectiveness of our proposed nonlinear regression modeling. In all experiments, we use an identity matrix as K in (17) and we fixed the value of h_c in (7) by sufficiently large. In addition, the model selection criterion GIC was used for choosing the number of basis functions m, a regularization parameter λ, hyperparameter ν, and combination of appropriate change points.

7.1 Simulation study

For the first simulation study, repeated random samples $\{(x_\alpha, y_\alpha); \alpha = 1, \cdots, n\}$ with $n = 100$ were generated from a true regression model $y_\alpha = u(x_\alpha) + \epsilon_\alpha$. The design points x_α are points that divides equally $[0, 1]$ and the errors ϵ_α are independently, normally distributed with mean 0 and standard deviation $\eta = 0.2$. We considered the following true regression model:

$$u(x) = \sin(2\pi x) + I(x \geq x_{50}), \tag{22}$$

where I is an indecator function, that is, $I(x \geq a) = 0\ (x < a), I(x \geq a) = 1\ (x \geq a)$ and $x_{50} \fallingdotseq 0.495$.

In Figure 2 the left panel shows the true curve (22) and the right panel shows the estimated curve obtained by our proposed nonlinear regression modeling procedure. In

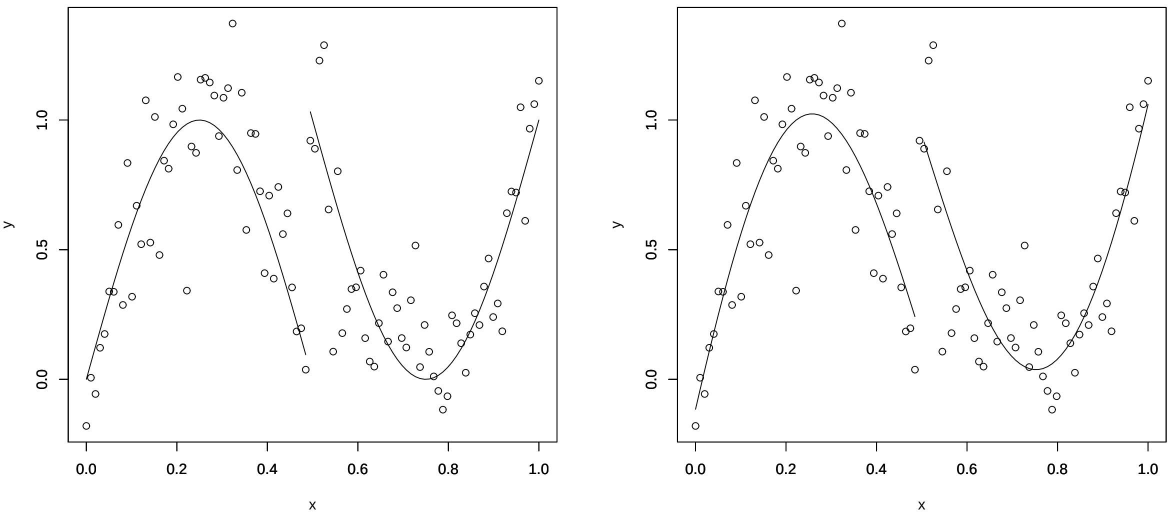

Fig. 2: The true curve generating data (left) and estimated curve using our proposed procedure (right).

this study, we made the points that corresponded to two high ranks of absolute values of coefficients estimated by RVM the candidates for the change points.

We performed 100 repetitions, and then it was 88 times that the point of about true change point x_{50} selected as a change point. In these 88 cases, the mean of selected point as jump point was 0.491 and the standard deviation was 7.74×10^{-3}. We observe that our modeling procedure captures the true structure effectively.

7.2 Real data analysis

7.2.1 Nile data

The data consists of the 100 measurements of annual flow of the Nile river at Ashwan from 1871 to 1970 (Cobb, 1978). Cobb (1978) and Muller (1992) sugget that a change occurs in the year 1898 and the same point was selected as change point by GIC. We made the points that corresponded to two high ranks of absolute values of coefficients estimated by RVM the candidates for the change points.

Figure 3 shows the Nile data and estimated curve obtained by our proposed method and estimated curve is smooth except for the change point. We observed that our modeling procedure captures change points and nonlinear structure of the data.

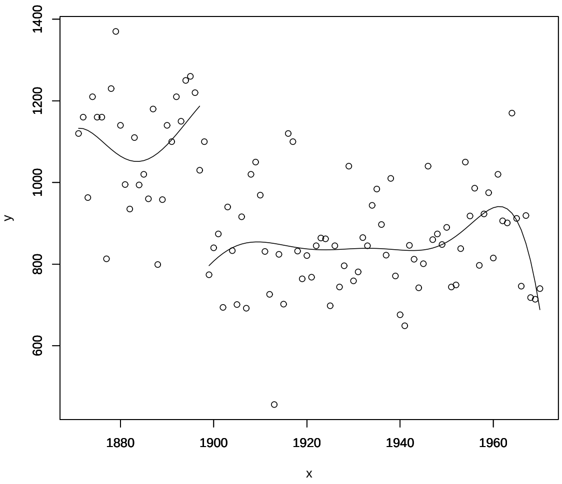

Fig. 3: Nile data

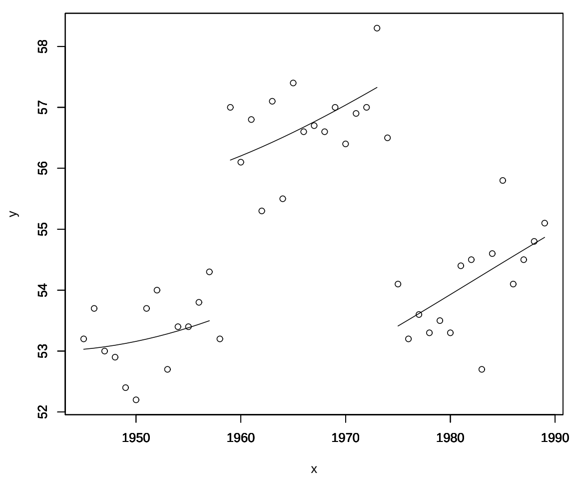

Fig. 4: Penny thickness data

7.2.2 Penny thickness data

The data consists of 90 measurements in mils ($\fallingdotseq 0.025\ mm$) of the thickness of 90 US Lincoln pennies (Scott, 1992). There are two measurements each year, from 1945 through 1989, and we use the mean of each year, that is, $n = 45$ like Gijbels *et al.* (2007). Speckman (1994) found that there were changes in thickness around the years 1958 and 1974 using their jump detection procedure. We made the points that corresponded to three high ranks of absolute values of coefficients estimated by RVM the candidates for the change points.

Figure 4 shows the Penny thickness data and the result curve estimated by our proposed method. We observed that our modeling procedure captures the structure of the data.

8 Concluding remarks

We have proposed a discontinuous nonlinear regression modeling procedure along with the technique of RVM and regularization method. The proposed methods assume unknown number of jump points, and we have used the discontinuous basis functions to detect multiple change points. Furthermore, we have used the normal Gaussian basis functions to get smoothness excluding change points. In order to choose optimal values of adjusted parameters and combination of change points, we presented the model selection criterion from information-theoretic approaches. The normal Gaussian basis function regression model has been widely used to draw information from data with complex structure. However, using only normal Gauusian basis funciton will lead to over smooth curve estimates. The simulation results reported here demonstrate the effectiveness of the proposed modeling strategy in terms of prediction accuracy.

References

[1] Ando, T., Konishi, S. and Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks. *Journal of Statistical Planning and Inference* **138**, 3616–3633.

[2] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press.

[3] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[4] Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n (with discussion). *Annals of Statistics* **35**, 2313–2351.

[5] Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*. **65**, 243–251.

[6] de Boor, C. (2001). *A Practical Guide to Splines*. Springer.

[7] Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.*, **11**, 89–121.

[8] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

[9] Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–148.

[10] Gijbels, I., Lambert, A. and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics* **59**, 235–272.

[11] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning* (2nd edition). Springer–Verlag, New York.

[12] Imoto, S. and Konishi, S. (2003). Selection of smoothing parameter in B-spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics* **55**, 671–687.

[13] Kawano, S. and Konishi, S. (2007). Nonlinear regression modeling via regularized Gaussian basis functions. *Bull. Inform. Cybern.*, **39**, 83–96.

[14] Konishi, S., and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, 875–890.

[15] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer.

[16] Moody, J. and Darken, C. J. (1989). Fast learning in networks of locally-turned processing units. *Neural Computation.* **1**, 281–294.

[17] Neal, R. M. (1996). *Bayesian Learning for Neural Networks.* Springer.

[18] Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics* **15**, 437–453.

[19] Scott, D. W. (1992). Multivariate density estimation. Theory, practice and visualization. New York: Wiley.

[20] Speckman, P. L. (1997). Detection of change-points in nonparametric regression. *Unpublished manuscript.*

[21] Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research.* **1**, 211–244.

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine