

関連ベクターマシンに基づく非線形回帰モデルと 変化点探索

九州大学大学院数理学府 立石 正平
九州大学大学院数理学研究院 小西 貞則

1 はじめに

回帰分析において、複雑な非線形構造を内包する現象を解明するために、スプライン関数やカーネル関数に基づく非線形回帰モデルが用いられる。しかし、ある点を境に平均構造が突然変化するデータに対しては、非線形回帰モデルの当てはめが困難となり、データから有効な情報を得ることが難しくなる。このようなデータに対しては、変化点を適切に捉えることが重要となる(井元・小西(2000))。

本報告では、Tipping (2001) による関連ベクターマシン (RVM: Relevance Vector Machine) を適用することで、変化点をもつ非線形構造を適切に推定する方法を提案する。基底関数展開に基づく非線形回帰モデルを仮定する場合、調整パラメータの選択が本質的となる。この問題に対して、モデル評価基準 (Konishi and Kitagawa, 2008) を導出し、調整パラメータの選択を行う。最後に、提案手法の有効性を数値例を通して検証する。

2 基底関数展開に基づく非線形回帰モデル

目的変数 Y と説明変数 X に関して、 n 組のデータ $\{(y_i, x_i); i = 1, 2, \dots, n\}$ が観測されたとする。このとき、説明変数と目的変数の間に次の関係が成り立つと仮定する。

$$y_i = u(x_i; \mathbf{w}) + \epsilon_i, \quad (i = 1, 2, \dots, n). \quad (1)$$

ここで、 ϵ は誤差項、 $u(x; \mathbf{w})$ は、未知のパラメータベクトル \mathbf{w} に依存する回帰関数である。

次に、以下のように回帰関数として基底関数と呼ばれる既知の非線形関数 $\mathbf{b}(x)$ に基づくモデルを考える。

$$u(x; \mathbf{w}) = \mathbf{w}'\mathbf{b}(x). \quad (2)$$

ただし、 $\mathbf{b}(x) = (1, b_1(x), b_2(x), \dots, b_m(x))'$ は $(m+1)$ 次元基底関数ベクトル、 $\mathbf{w} = (w_0, w_1, \dots, w_m)'$ は $(m+1)$ 次元係数パラメータベクトルとする。誤差項 ϵ_i は互いに独立に平均 0、分散 σ^2 の正規分布に従うとすると、基底関数に基づく回帰モデルは次のように表される。

$$f(y_i | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{(y_i - \mathbf{w}'\mathbf{b}(x_i))^2}{2\sigma^2} \right\}. \quad (3)$$

3 変化点探索と曲線推定

本報告では、変化点の探索と曲線推定の2段階の推定を行う。まず、第1段階では、変化点を捉えるための基底関数に基づく非線形回帰モデルを仮定する。次に、第2段階では、第1段階で用いた基底関数に加え、さらに、ガウス型基底関数を融合した新たな非線形回帰モデルを仮定する。

3.1 関連ベクターマシンに基づく変化点探索

n 組のデータ $\{(y_i, x_i); i = 1, 2, \dots, n\}$ が観測されたとき、観測値ベクトル $\mathbf{y} = (y_1, \dots, y_n)'$ の分布は (3) 式より次の分布に従う。

$$\mathbf{y}|\mathbf{w}, \beta \sim N(\Phi\mathbf{w}, \sigma^2 I_n) \quad (4)$$

3.1 節では、 $\beta^{-1} = \sigma^2$ とし、 β はハイパーパラメータとして扱う。 Φ は基底関数からなる $n \times (n+1)$ 行列で、

$$\Phi = \begin{bmatrix} 1 & \phi_1(x_1) & \cdots & \phi_n(x_1) \\ 1 & \phi_1(x_2) & \cdots & \phi_n(x_2) \\ \vdots & & & \\ 1 & \phi_1(x_n) & \cdots & \phi_n(x_n) \end{bmatrix} \quad (5)$$

とする。ここで、データ数 n と同じ数の基底関数を設定することが重要となる。変化点探索における基底関数 $\phi(x)$ として、次の関数を用いる。

$$\phi_j(x) = \begin{cases} \exp\left(\frac{\|x - x_j\|^2}{h^2}\right), & (x < x_j) \\ 0, & (x = x_j) \\ -\exp\left(\frac{\|x - x_j\|^2}{h^2}\right), & (x > x_j) \end{cases}, j = 1, 2, \dots, n. \quad (6)$$

図1には基底関数 $\phi(x)$ の例を与える。 $\phi(x)$ は、その中心の点の前後で値が大きく変化する関数であるので、係数パラメータの推定値の絶対値が大きくなる点に変化点である可能性があると考えられる。

次に係数パラメータ \mathbf{w} に次の事前分布を設定する。

$$\mathbf{w}|\boldsymbol{\alpha} \sim N(\mathbf{0}, A^{-1}). \quad (7)$$

ただし、 $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_n)'$ 、 $A = \text{diag}(\alpha_0, \dots, \alpha_n)$ である。このような事前分布は、ARD(関連自動決定)事前分布と呼ばれる。このとき、ベイズの定理より、 \mathbf{w} の事後分布は

$$\mathbf{w}|\mathbf{y}, \boldsymbol{\alpha}, \beta \sim N(\boldsymbol{\mu}, \Sigma) \quad (8)$$

となる。ただし、

$$\Sigma = (\beta\Phi'\Phi + A)^{-1}, \quad (9)$$

$$\boldsymbol{\mu} = \beta\Sigma\Phi'\mathbf{y} \quad (10)$$

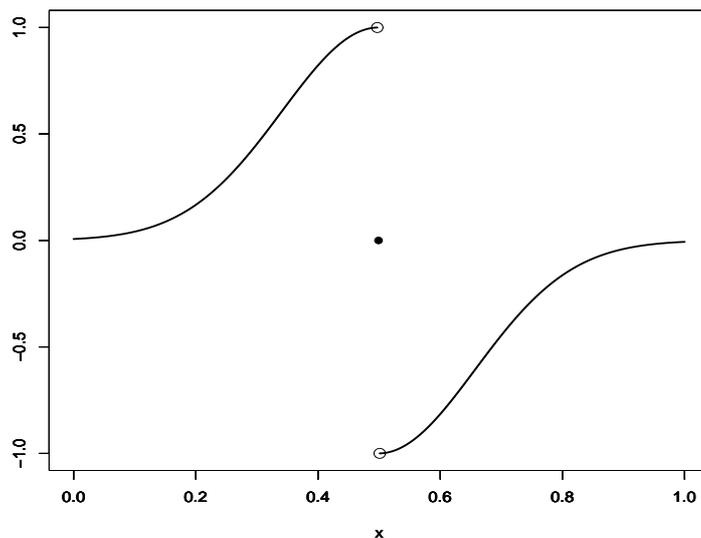


図 1: 基底関数 $\phi(x)$ の構成例 .

とする . 係数パラメータベクトル w の推定は , (10) 式の事後平均ベクトルで行う .

ここで , ハイパーパラメータ $\alpha \beta$ の推定について考える . 観測値ベクトル y の周辺尤度関数は

$$p(\mathbf{y}|\alpha, \beta) = \int f(\mathbf{y}|\mathbf{w}, \beta)\pi(\mathbf{w}|\alpha)d\mathbf{w} \quad (11)$$

であり , 自然共役の関係性から ,

$$\mathbf{y}|\alpha, \beta \sim N(\mathbf{0}, \beta^{-1}I_n + \Phi A^{-1}\Phi') \quad (12)$$

となる . この周辺尤度を最大にするハイパーパラメータの推定量 $\hat{\alpha} \hat{\beta}$ は次で与えられる .

$$\hat{\alpha}_j = \frac{\gamma_j}{\mu_j^2}, \quad \hat{\beta}^{-1} = \frac{\|\mathbf{y} - \Phi\boldsymbol{\mu}\|^2}{n - \sum_{j=0}^{n+1} \gamma_j} \quad (13)$$

ただし , $\gamma_j = 1 - \alpha_j \Sigma_{jj}$ である . $\hat{\alpha}_j$ と $\hat{\beta}$ は互いに依存しているため , 上式で繰り返し更新を行う .

以上のように , 周辺尤度最大化に基づいて係数パラメータの推定とハイパーパラメータの推定を逐次計算によって行う手法は関連ベクターマシン (RVM: Relevance Vector Machine) と呼ばれ , Tipping (2001) によって提案された . RVM は , 推定に不要な基底関数にかかる係数パラメータを真に 0 に推定する . 従って , 切片項を除く n 個の係数パラメータの中で , 0 と推定されなかった基底関数に対応する点の変化点の候補と考えることができ , 変化点の可能性がある点を $\{\tau_1, \tau_2, \dots, \tau_t\}$ と絞り込むことができる . ここで , 変化点候補の組み合わせの集合族 $\{\emptyset, \tau_1, \dots, (\tau_1, \tau_2), \dots, (\tau_1, \tau_2, \dots, \tau_t)\}$ の元の 1 つを $T = \{t_1, \dots, t_{n_T}\}$ ($n_T < n$) とする .

3.2 正則化法に基づく曲線推定

3.1節では，(6)式の基底関数 $\phi(x)$ を用いた非線形回帰モデルを仮定し，RVM を適用することで，変化点探索を行った．次に，(3)式の基底関数ベクトル $\mathbf{b}(x)$ を

$$\mathbf{b}(x) = (1, \psi_1(x), \dots, \psi_m(x), \tilde{\phi}_T(x)')' \quad (14)$$

とした基底関数展開に基づく非線形回帰モデルを仮定する．ここで， $\tilde{\phi}_T(x) = (\phi_{t_1}(x), \dots, \phi_{t_{n_T}}(x))'$ とする． $\psi(x)$ は次で与えられるガウス型基底関数である．

$$\psi_k(x) = \exp\left(-\frac{\|x - c_k\|^2}{2\nu s_k^2}\right), \quad k = 1, 2, \dots, m. \quad (15)$$

ただし， $m (< n)$ はガウス型基底関数の個数， c_k は関数の中心， s_k^2 は関数の広がりを表す量である．また， $\nu (> 0)$ は $\psi_k(x)$ の広がり度を調整するパラメータである (Ando *et al.*, 2008)．

ここで， c_k と s_k^2 の決定法を述べる． n 個のデータ $\{x_1, \dots, x_n\}$ をクラスタリング手法の一つである K-means 法によって m 個のクラスタ C_1, \dots, C_m に分割し，各クラスタ C_k に含まれる n_k 個のデータに基づいて c_k と s_k^2 は，

$$c_k = \frac{1}{n_k} \sum_{x_i \in C_k} x_i, \quad s_k^2 = \frac{1}{n_k} \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (16)$$

で決定する．

複雑な現象を近似するために非線形モデルを当てはめるとき，モデルのパラメータを最尤法によって推定すると，しばしばデータに過度に依存したモデルが推定される．そこで， $u(x; \mathbf{w})$ の滑らかさが失われるにつれて増加する罰則項 (正則化項) を対数尤度関数に付け加えた

$$l_\lambda(\mathbf{w}, \sigma^2) = \sum_{\alpha=1}^n \log f(y_\alpha | \mathbf{w}, \sigma^2) - \frac{n\lambda}{2} H(\mathbf{w})$$

の最大化，すなわち正則化法によって推定する．ここで， $\lambda (> 0)$ は平滑化パラメータと呼ばれ，モデルの当てはまりのよさと曲線の滑らかさを制御するハイパーパラメータである．また， $H(\mathbf{w})$ は正則化項と呼ばれ， \mathbf{w} の 2 次形式 $\mathbf{w}'K\mathbf{w}$ などが用いられる．ただし K は既知の $m + n_T + 1$ 次の非負定値行列である．このモデルのパラメータ \mathbf{w} と σ^2 の正則化最尤推定量は次で与えられる．

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

ただし，

$$B = \begin{bmatrix} \mathbf{b}(x_1)' \\ \mathbf{b}(x_2)' \\ \vdots \\ \mathbf{b}(x_n)' \end{bmatrix} = \begin{bmatrix} 1 & \psi_1(x_1) & \psi_2(x_1) & \cdots & \psi_m(x_1) & \tilde{\phi}_T(x_1)' \\ 1 & \psi_1(x_2) & \psi_2(x_2) & \cdots & \psi_m(x_2) & \tilde{\phi}_T(x_2)' \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \psi_1(x_n) & \psi_2(x_n) & \cdots & \psi_m(x_n) & \tilde{\phi}_T(x_n)' \end{bmatrix}$$

である．これらの推定量は，分散 σ^2 の推定を通して互いに依存しているので，以下のよ
うなアルゴリズムで計算できる．

[step1] 初期値を $\sigma^{2(0)}$ とし λ をあらかじめ与えておく． $k = 1$ とする．

[step2] $\mathbf{w}^{(k)} = (B'B + n\lambda\sigma^{2(k-1)}K)^{-1}B'\mathbf{y}$ で更新する．

[step3] $\sigma^{2(k)} = \frac{1}{n}(\mathbf{y} - B\mathbf{w}^{(k)})'(\mathbf{y} - B\mathbf{w}^{(k)})$ を求め，十分小さな δ に対して， $(\sigma^{2(k)} - \sigma^{2(k-1)})^2 < \delta$ を評価する．この条件を満たすまで step2 を繰り返し，条件を満たしたパラメータを推定値とする．

4 モデル評価基準

前節で述べた推定したモデルは，ガウス型基底関数の個数 m ，平滑化パラメータ λ ，ガウス基底関数の幅調整パラメータ ν に依存しており，この選択には以下のモデル評価基準を用いる．

一般化情報量規準 GIC (Konishi and Kitagawa, 1996)

データを発生した真の分布 $G(x)$ から得られた n 個のデータ $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ に基づく経験分布関数を $\hat{G}(x)$ とするとき，統計的汎関数 $T(\cdot)$ によって定義される一般の推定量 $\hat{\boldsymbol{\theta}} = \mathbf{T}(\hat{G}) = (T_1(\hat{G}), T_2(\hat{G}), \dots, T_p(\hat{G}))'$ に基づくモデル $f(x|\hat{\boldsymbol{\theta}})$ の評価基準として，次で与えられる一般化情報量規準 GIC (Konishi and Kitagawa (1996)) がある．

$$\text{GIC} = -2 \sum_{\alpha=1}^n \log f(x_\alpha|\hat{\boldsymbol{\theta}}) + \frac{2}{n} \sum_{\alpha=1}^n \text{tr} \left\{ \mathbf{T}^{(1)}(x_\alpha; \hat{G}) \frac{\partial \log f(x_\alpha|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \bigg|_{\hat{\boldsymbol{\theta}}} \right\}.$$

ただし， $\mathbf{T}^{(1)}(x_\alpha; \hat{G}) = (T_1^{(1)}(x_\alpha; \hat{G}), \dots, T_p^{(1)}(x_\alpha; \hat{G}))'$ は経験影響関数とよばれ，その第 i 要素 $T_i^{(1)}(x_\alpha; \hat{G})$ は次で定義される．

$$T_i^{(1)}(x_\alpha; \hat{G}) = \lim_{\epsilon \rightarrow 0} \frac{T_i((1-\epsilon)\hat{G} + \epsilon\delta_\alpha) - T_i(\hat{G})}{\epsilon}.$$

ここで， δ_α は点 x_α 上に確率 1 をもつ分布関数である．この考えを，前述の正則化最尤法によって推定された基底関数展開に基づく非線形モデルに適用すると，回帰モデルにおける GIC は

$$\text{GIC} = n(\log 2\pi + 1) + n \log \hat{\sigma}^2 + 2\text{tr}\{R^{-1}Q\} \quad (17)$$

で与えられる．ただし，

$$R = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} B'B + n\lambda\hat{\sigma}^2K & \frac{1}{\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{\hat{\sigma}^2}\mathbf{1}'_n\Lambda B & \frac{n}{2\hat{\sigma}^2} \end{bmatrix}, \quad (18)$$

$$Q = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} \frac{1}{\hat{\sigma}^2}B'\Lambda^2B - \lambda K\hat{\mathbf{w}}\mathbf{1}'_n\Lambda B & \frac{1}{2\hat{\sigma}^4}B'\Lambda^3\mathbf{1}_n - \frac{1}{2\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^4}\mathbf{1}'_n\Lambda^3B - \frac{1}{2\hat{\sigma}^2}\mathbf{1}'_n\Lambda B & \frac{1}{4\hat{\sigma}^6}\mathbf{1}'_n\Lambda^4\mathbf{1}_n - \frac{n}{4\hat{\sigma}^2} \end{bmatrix} \quad (19)$$

であり, $\mathbf{1}_n$ は要素が全て 1 の n 次元ベクトル, Λ は

$$\Lambda = \text{diag}(y_1 - \hat{\mathbf{w}}' \mathbf{b}(x_1), \dots, y_n - \hat{\mathbf{w}}' \mathbf{b}(x_n))$$

とする.

5 適用例

データはガウスノイズに対して $N(0, \eta^2)$ を仮定した次の式に従って, 100 個のデータを発生させた.

$$y_i = \sin(2\pi x_i) + I(x_i \geq x_{50}) + \epsilon_i, \quad (i = 1, \dots, n) \quad (20)$$

ただし, $I(x_i \geq a) = 0$ ($x_i < a$), $I(x_i \geq a) = 1$ ($x_i \geq a$) である. ここで, データ x は $[0, 1]$ を 100 等分した点とし, ノイズの標準偏差 η は, $\eta = 0.2$ とした. 変化点は 50 番目のデータ点 $x_{50} = 0.495$ に設定した.

(20) 式を真の曲線とするデータに対して提案手法を適用した結果を図 2 に示す. 今回は, (6) 式の基底関数の幅は十分大きい値で固定し, RVM で 0 と推定されなかった係数の中で絶対値の値が上位 2 つの係数と対応する点を変化点の候補とした. 変化点の位置, ガウス型基底関数の個数 m , 平滑化パラメータ λ , ガウス基底関数の幅調整パラメータ ν は一般化情報量規準 GIC で選択した.

100 回の実験を行った結果, 変化点の数を正確に選択できた回数は 88 回であり, その際の推定された変化点の平均値は 0.491, 標準偏差は 7.74×10^{-3} であった.

5.1 数値実験

5.2 ナイルデータ

1871 年から 1970 年までのナイル河の年間河川流量 ($10^8 m^3$) を計測したものである. cobb(1978) は, 1898 年を境に年間河川流量が突然変化していることを示している.

このナイルデータに対して, 提案手法を適用した結果を図 3 に示す. (6) 式の基底関数の幅は十分大きい値で固定し, RVM で 0 と推定されなかった係数の中で絶対値の値が上位 2 つの係数と対応する点を変化点の候補とした. 変化点の位置, ガウス型基底関数の個数 m , 平滑化パラメータ λ , ガウス基底関数の幅調整パラメータ ν は一般化情報量規準 GIC で選択した結果, 1898 年が変化点として選択された.

5.3 硬貨データ

1945 年から 1989 年までのリンカーン 1 セント硬貨の厚さ ($1 \text{mil} \doteq 0.0254 \text{mm}$) を計測したものである (Scott, 1992). 硬貨の厚さは第 2 次世界大戦中に減らされ, 1960 年ごろに元

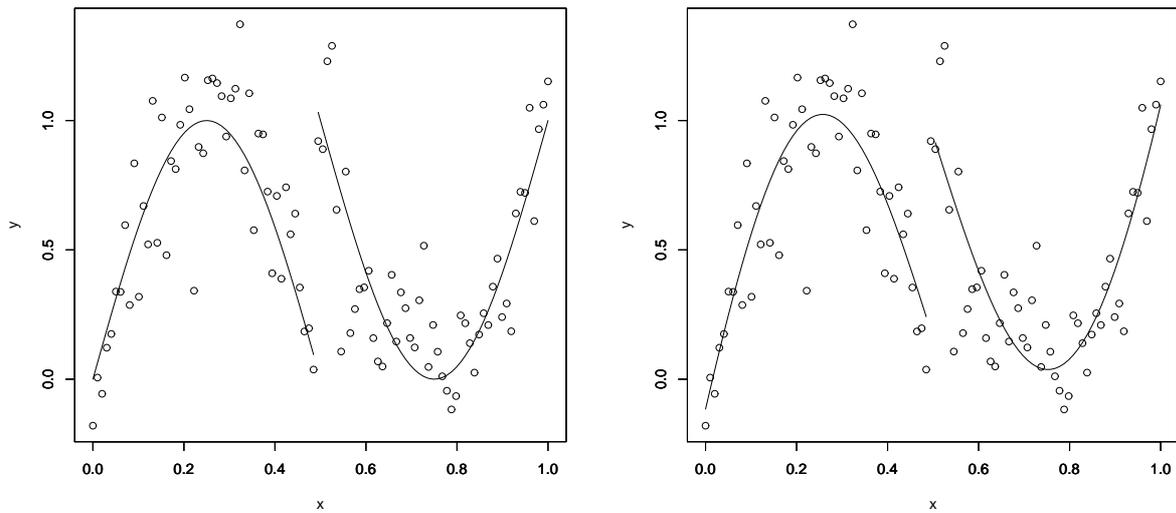


図 2: 人工データ解析例．左図は，実線が真の曲線．右図は，実線が推定曲線を示す．

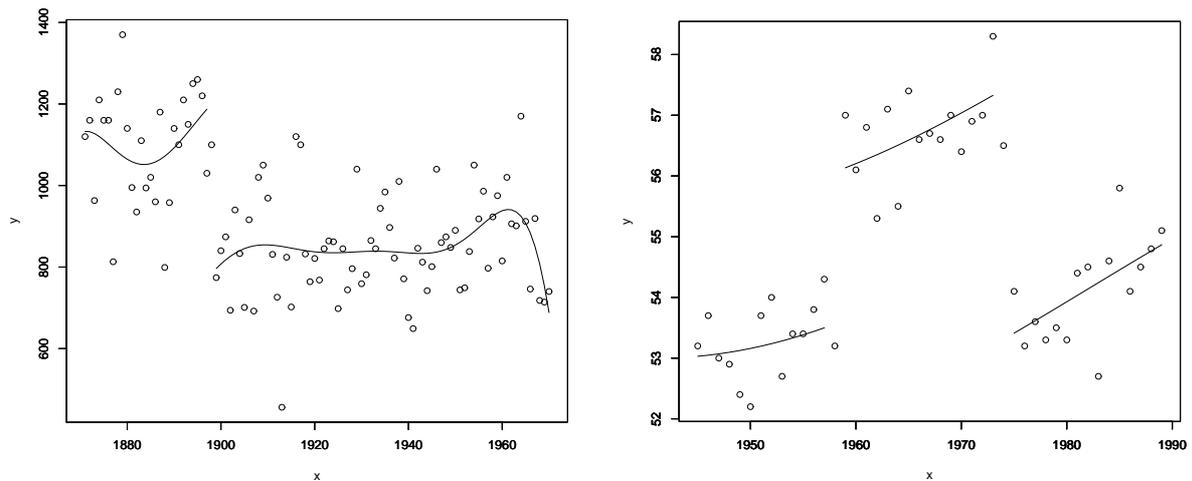


図 3: ナイルデータ．実線が推定曲線を示す．図 4: 硬貨データ．実線が推定曲線を示す．

の厚さに戻されたが，1970年代に再び減らされたことを示している．Speckman(1994) や Gijbels and Goderniaux(2004) では，1958年と1974年に変化が起こったと推定している．

この硬貨データに対して，提案手法を適用した結果を図4に示す．(6)式の基底関数の幅は十分大きい値で固定し，RVMで0と推定されなかった係数の中で絶対値の値が上位3つの係数と対応する点を変化点の候補とした．変化点の位置，ガウス型基底関数の個数 m ，平滑化パラメータ λ ，ガウス基底関数の幅調整パラメータ ν は一般化情報量規準 GIC で選択した結果，1958年と1974年が変化点として選択された．

6 まとめ

基底関数展開に基づく非線形回帰モデルを仮定し，関連ベクターマシンを適用することで，変化点をもつ非線形構造を適切に推定する方法を提案した．また，変化点の位置，ガウス型基底関数の個数と調整パラメータ，そして平滑化パラメータの選択に情報量規準 GIC を用いた．そして，平均構造に変化点を含む実データと人工データを用いた数値例を通して，提案手法の有効性を示した．

参考文献

- [1] 井元清哉，小西貞則 (2000). *B*-スプラインによる不連続曲線推定と変化点探索. 研究会”多変量解析と漸近理論に関する新展開”. 広島大学 .
- [2] 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.
- [3] Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*. **65**, 243–251.
- [4] Gijbels, I. and Goderniaux, A.-C. (2004). Bandwidth selection for change point estimation in nonparametric regression. *Technometrics*, **46**, 76–86.
- [5] Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*. **83**, 875–890.
- [6] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.
- [7] Scott, D. W. (1992). Multivariate density estimation. Theory, practice and visualization. New York: Wiley.
- [8] Speckman, P. L. (1997). Detection of change-points in nonparametric regression. *Unpublished manuscript*.
- [9] Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*. **1**, 211–244.