

Fused lasso に基づく非線形回帰モデリング

九州大学大学院数理学府 立石 正平

中央大学理工学部 小西 貞則

1 はじめに

回帰分析において、複雑な構造を有する現象を解明するために、スプライン関数や動径基底関数に基づく非線形回帰モデルが用いられる。しかし、データを発生させた真の構造が滑らかでない場合、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。

本報告では、fused lasso (Tibshirani *et al.*, 2005) を用いて構成した基底関数に基づく非線形回帰モデルを適用することで、突発的な変化をもつ非線形構造を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、基底関数の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、モデル評価基準を導出し、調整パラメータの選択を行い、提案手法の有効性を数値実験を通して検証する。

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 $\{(y_\alpha, x_\alpha); \alpha = 1, 2, \dots, n\}$ 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基底関数展開に基づくモデルを仮定する。

$$y_\alpha = \mathbf{w}^T \mathbf{b}(x_\alpha) + \epsilon_\alpha, \quad (\alpha = 1, 2, \dots, n).$$

ここで、 $\mathbf{b}(x) = (1, b_1(x), b_2(x), \dots, b_m(x))^T$ は $(m+1)$ 次元基底関数ベクトル、 $\mathbf{w} = (w_0, w_1, w_2, \dots, w_m)^T$ は $(m+1)$ 次元係数パラメータベクトルとする。ここで、基底関数として次の式で与えられるガウス型基底関数を用いる。

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

基底関数の広がりを調整パラメータ $\nu (> 0)$ によって調整することで滑らかな推定曲線が得られることから、その有効性が報告されている (Ando and Konishi, 2008)。

ガウス型基底関数に含まれる中心ベクトル μ と関数の広がり の程度を表す h^2 を決定する 必要があり, k -means 法が用いられる場合がある. すなわち, n 個のデータ $\{x_1, \dots, x_n\}$ を, 基底関数の個数である m 個のクラスタ C_1, C_2, \dots, C_m に分割し, 各クラスタ C_j に含まれる n_j 個のデータについて μ_j と h_j^2 を次のように決定する.

$$\hat{c}_j = \frac{1}{n_j} \sum_{x_\alpha \in C_j} x_\alpha, \quad \hat{h}_j^2 = \frac{1}{n_j} \sum_{x_\alpha \in C_j} (x_\alpha - c_j)^2. \quad (2)$$

しかし, このように構成された基底関数に基づく非線形回帰モデルは, 滑らかな曲線になっ てしまい, 平均構造が突然変化するデータに対する当てはめが困難となる. この問題を克服するために, 本報告では, fused lasso の特殊な形である fused lasso signal approximator (FLSA; Friedman *et al.*, 2007) を用いて基底関数を構成することを提案する.

3 FLSA とガウス型基底関数の構築

n 組のデータ $\{(y_\alpha, x_\alpha); \alpha = 1, 2, \dots, n\}$ に対して, FLSA は次の目的関数の最小化を 行う.

$$\frac{1}{2} \sum_{\alpha=1}^n (y_\alpha - \beta_\alpha)^2 + \lambda_1 \sum_{\alpha=1}^n |\beta_\alpha| + \lambda_2 \sum_{\alpha=1}^{n-1} |\beta_\alpha - \beta_{\alpha+1}|, \quad \alpha = 1, \dots, n. \quad (3)$$

ここで, β_α は, 目的変数 y_α の推定量であり, λ_1 は β を真に 0 に推定するための正則化 パラメータである. λ_2 は隣り合う β の違いを真に 0 に推定するための正則化パラメータ であり, λ_2 の値を十分大きくすることによって, $x^{F_j} = \{x_\alpha | \alpha \in F_j\}$ に関する β は全て $\hat{\beta}_{F_j}$ と推定される ($\cup_{j=1}^{n_F} F_j = \{1, 2, \dots, n\}$, $F_k \cap F_l = \emptyset$ ($k \neq l$)). FLSA 推定の結果, 同じ推 定値をもつ区間によってデータは $m = n_F$ 個のクラスタに分割され, 対応する説明変数の クラスタを C_1, \dots, C_m とする. ここで, FLSA 推定曲線に基づき, 次の式で与えられる ガウス型基底関数を構築することを考える.

$$b_j(x) = \begin{cases} \exp \left\{ -\frac{(x - c_j)^2}{2\nu h_{1j}^2} \right\}, & (x < c_j) \\ 1, & (x = c_j) \\ \exp \left\{ -\frac{(x - c_j)^2}{2\nu h_{2j}^2} \right\}, & (x > c_j) \end{cases}, \quad j = 1, 2, \dots, m. \quad (4)$$

指数関数 $\phi(x) = \exp \{-(x - \mu)^2/(2h^2)\}$ の 1 次導関数は

$$\phi(x)' = -\left(\frac{x - \mu}{h^2} \right) \exp \left\{ -\frac{(x - \mu)^2}{2h^2} \right\} \quad (5)$$

であるため, $x = \mu - kh$, $\mu + kh$ ($k \geq 1$) における $\phi(x)$ の傾きが, それぞれ d_1, d_2 となる
ときの h の値 h_1, h_2 は,

$$h_1 = \frac{k}{|d_1|} \exp\left(-\frac{k^2}{2}\right), \quad h_2 = \frac{k}{|d_2|} \exp\left(-\frac{k^2}{2}\right) \quad (6)$$

となる. 従って, (6) 式より, $x = \mu_j \pm k_{ij}h_{ij}$ における $\phi(x)$ の傾きが, FLSA 推定曲線の
傾き \hat{d}_{ij} と等しくなるような h_{ij} の値が次の式で与えられる.

$$\hat{h}_{ij} = \frac{k_{ij}}{|\hat{d}_{ij}|} \exp\left(-\frac{k_{ij}^2}{2}\right). \quad (7)$$

ただし, $\mathbf{x}^{F_j} = \{x_1^{F_j}, \dots, x_{n_j}^{F_j}\}$ とし, \hat{d}_{ij} は次のように定義する.

$$\hat{d}_{1j} = \frac{\hat{\beta}_{F_j} - \hat{\beta}_{F_{j-1}}}{x_{n_j}^{F_j} - x_{n_{j-1}}^{F_{j-1}}}, \quad \hat{d}_{2j} = \frac{\hat{\beta}_{F_{j+1}} - \hat{\beta}_{F_j}}{x_1^{F_{j+1}} - x_1^{F_j}}. \quad (8)$$

また, n_j を j 番目のクラス F_j に含まれるデータ数とし, ガウス型基底関数の中心パラ
メータ c_j を次の式で定義する.

$$\hat{c}_j = \frac{1}{n_j} \sum_{x_\alpha \in C_j} x_\alpha. \quad (9)$$

(4) 式の c_j と h_{ij} をそれぞれ \hat{c}_j と \hat{h}_{ij} で置き換えることによって, 以下の m 個の基底関
数が得られる.

$$b_j(x; \hat{c}_j, \hat{h}_{ij}^2) = \begin{cases} \exp\left\{-\frac{(x - \hat{c}_j)^2}{2\nu\hat{h}_{1j}^2}\right\}, & (x < \hat{c}_j) \\ 1, & (x = \hat{c}_j) \\ \exp\left\{-\frac{(x - \hat{c}_j)^2}{2\nu\hat{h}_{2j}^2}\right\}, & (x > \hat{c}_j) \end{cases}, \quad j = 1, 2, \dots, m. \quad (10)$$

(10) 式の基底関数に基づく非線形回帰モデルは

$$y_\alpha = \mathbf{w}^T \mathbf{b}(x_\alpha) + \epsilon_\alpha, \quad \alpha = 1, \dots, n, \quad (11)$$

で表され, 誤差項が独立に平均 0, 分散 σ^2 の正規分布に従うと仮定すると, 密度関数は
以下の式で与えられる.

$$f(y_\alpha | \mathbf{w}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y_\alpha - \mathbf{w}^T \mathbf{b}(x_\alpha)\}^2}{2\sigma^2}\right], \quad \alpha = 1, \dots, n. \quad (12)$$

4 正則化最尤法

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、 $u(x; w)$ 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

$$l_\lambda(w, \sigma^2) = \sum_{\alpha=1}^n \log f(y_\alpha | w, \sigma^2) - \frac{n\gamma}{2} H(w) \quad (13)$$

の最大化、すなわち正則化法によって推定する。ここで、 $\gamma (> 0)$ は正則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調整パラメータである。また、 $H(w)$ は正則化項と呼ばれ、 w の 2 次形式 $w'Kw$ などが用いられる。ただし K は既知の $m+1$ 次の非負定値行列である。このモデルのパラメータ w と σ^2 の正則化最尤推定量は次で与えられる。

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

ただし、 $B = (\mathbf{b}(x_1)^T, \dots, \mathbf{b}(x_n)^T)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$ である。

5 モデル評価基準

推定したモデルは、正則化パラメータ γ 、ガウス基底関数の幅調整パラメータ ν に依存しており、適切な値を決定することが重要となる。この問題に対し、モデル評価基準を用いて客観的に選択を行う。

統計的汎関数によって定義される一般の推定量の評価基準として、一般化情報量規準 GIC (Konishi and Kitagawa, 1996, 2008) がある。Konishi and Kitagawa (1996) 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

$$\text{GIC} = n(\log 2\pi + 1) + n \log \hat{\sigma}^2 + 2\text{tr}\{R^{-1}Q\} \quad (14)$$

で与えられる。ただし,

$$R = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} B^T B + n\gamma\hat{\sigma}^2 K & \frac{1}{\hat{\sigma}^2} B^T \Lambda \mathbf{1}_n \\ \frac{1}{\hat{\sigma}^2} \mathbf{1}_n^T \Lambda B & \frac{n}{2\hat{\sigma}^2} \end{bmatrix}, \quad (15)$$

$$Q = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} \frac{1}{\hat{\sigma}^2} B^T \Lambda^2 B - \gamma K \hat{\mathbf{w}} \mathbf{1}_n^T \Lambda B & \frac{1}{2\hat{\sigma}^4} B^T \Lambda^3 \mathbf{1}_n - \frac{1}{2\hat{\sigma}^2} B^T \Lambda \mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^4} \mathbf{1}_n^T \Lambda^3 B - \frac{1}{2\hat{\sigma}^2} \mathbf{1}_n^T \Lambda B & \frac{1}{4\hat{\sigma}^6} \mathbf{1}_n^T \Lambda^4 \mathbf{1}_n - \frac{n}{4\hat{\sigma}^2} \end{bmatrix} \quad (16)$$

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

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

である。GIC の値が最小となる γ と ν の値を適切な値とする。

6 適用例

データは, $y_\alpha = u(x_\alpha) + v(x_\alpha) + \epsilon_\alpha$ から発生させ, 真の曲線には次の関数を仮定した。

$$\begin{aligned} \text{(a)} \quad & \begin{cases} u(x) = \exp(-x)(x - 0.5)^3, \\ v(x) = -1.5 \cdot \mathbf{I}(0.50 \leq x \leq 0.51), \end{cases} \\ \text{(b)} \quad & \begin{cases} u(x) = \sin(2\pi x), \\ v(x) = -2 \cdot \mathbf{I}(0.25 \leq x \leq 0.26) + 2 \cdot \mathbf{I}(0.75 \leq x \leq 0.76), \end{cases} \\ \text{(c)} \quad & \begin{cases} u(x) = \exp(-x) \sin(5\pi \exp(-x)), \\ v(x) = 0. \end{cases} \end{aligned}$$

ただし, $\mathbf{I}(\cdot)$ は定義関数とする。データ点 x_α は区間 $[0, 1]$ を等間隔に分割した点であり, 誤差項 ϵ_α に対しては平均0, 分散 τ^2 の正規分布を仮定し, $\tau = 0.2R_u$ と設定した。ただし, R_u は定義域 $x \in [0, 1]$ に対する $u(x)$ の値域の幅を表す。本節の数値実験では, $\lambda_1 = 0$ と固定し, λ_2 の適切な値を交差検証法によって選択した。また, 正則化パラメータ γ と調整パラメータ ν は, 情報量規準 GIC で選択した。(13) 式の正則化項として $H(\mathbf{w}) = \mathbf{w}^T \mathbf{w}$ を用いた。データ数 $n = 100$ に対して, FLSA に基づく提案手法 (FLSA-G) と, k -means 法によって構成された基底関数に基づく非線形回帰モデル (Km-G) と, FLSA 推定 (FLSA) の3つの手法を比較する実験を行った。

図1は真の曲線と推定曲線を比較したものである。表1は50回の数値実験に対する結果を示す。表中の $\text{mean } m$, $\text{mean } \gamma$, $\text{mean } \nu$, AMSE はそれぞれ, 基底関数の個数 m の平均値, 選択された正則化パラメータ γ の平均値, 選択された調整パラメータ ν の平均, 平均2乗誤差 $\text{MSE} = \sum_{\alpha=1}^n \{(u(x_\alpha) + v(x_\alpha)) - \hat{y}_\alpha\}^2 / n$ の平均値を表す。

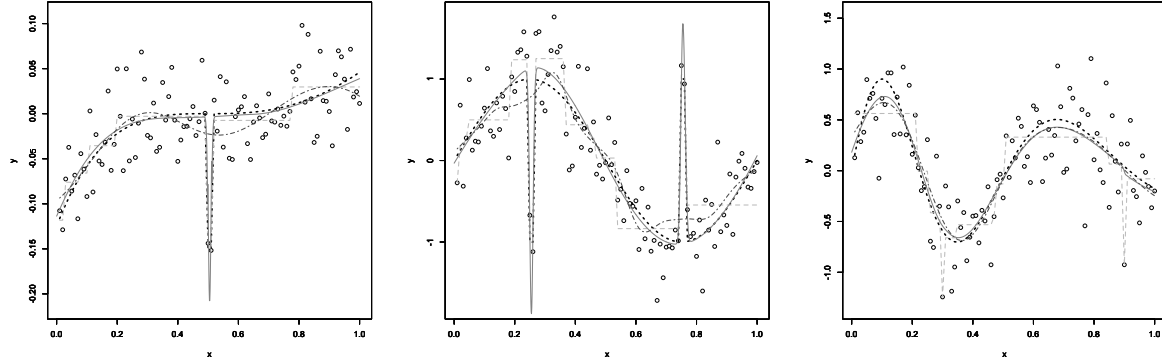


図 1: 真の曲線 (点線) と, FLSA-G (実線), Km-G (鎖線), FLSA (破線) の比較. 左上: (a). 右上: (b). 下: (c).

表 1: 数値実験結果 .

function		mean m	mean γ	mean ν	AMSE
(a)	FLSA-G	8.20	6.48×10^{-4}	4.90	1.48×10^{-4}
	Km-G	6.72	7.65×10^{-2}	34.16	4.89×10^{-4}
	FLSA	--	--	--	2.41×10^{-4}
(b)	FLSA-G	14.00	3.00×10^{-4}	402	3.84×10^{-2}
	Km-G	6.72	1.86×10^{-2}	2.22	1.77×10^{-1}
	FLSA	--	--	--	7.26×10^{-2}
(c)	FLSA-G	13.20	6.48×10^{-4}	62.00	9.20×10^{-3}
	Km-G	7.54	1.00×10^{-2}	20.86	1.18×10^{-2}
	FLSA	--	--	--	3.88×10^{-2}

参考文献

- [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] de Boor, C. (2001). *A Practical Guide to Splines*. Springer.
- [4] Eilers, P. and Marx, B. (1996). Flexible smoothing with B -splines and penalties (with discussion). *Statist. Sci.*, **11**, 89–121.
- [5] Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Statist.*, **1**, 302–332.
- [6] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. 2nd ed. New York: Springer.
- [7] Imoto, S. and Konishi, S. (2003). Selection of smoothing parameter in B -spline non-parametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics* **55**, 671–687.
- [8] Kawano, S. and Konishi, S. (2007). Nonlinear regression modeling via regularized Gaussian basis functions. *Bull. Inform. Cybern.*, **39**, 83–96.
- [9] Konishi, S., and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, 875–890.
- [10] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.
- [11] Moody, J. and Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*. **1**, 281–294.
- [12] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. Roy. Statist. Soc. Ser. B*, **67**, 91–108.