

SCAD 正則化法による非線形回帰モデリング

Nonlinear Regression Modeling via the SCAD Regularization

数学専攻 浅香 俊成
ASAKA, Toshinari

1 はじめに

非線形回帰モデルは、複雑な非線形構造を内包する現象のモデル化に有用な手法として用いられる。特に、基底展開法による非線形回帰モデルは、様々な現象のモデル化に有用な非線形解析手法である。基底展開法において、基底関数の個数の決定は重要な問題である。従来、最尤法により推定を行い、AIC (Akaike, 1973) や BIC (Schwarz, 1978) などの情報量規準やベイズアプローチにより評価・選択を行ってきたが、基底関数の個数が多数になる場合は計算コストが膨大になり実行が難しく、さらに推定が不安定となる。その不安定性を改善する方法として、正則化法により推定を行う。特に、 L_1 正則化項を用いた推定を行うことで、推定と基底関数選択を同時に行う方法を考える。本発表では、基底展開法による非線形回帰モデリングやその推定方法について述べ、予測に有効なモデル構築のための変数選択方法として、 L_1 正則化法の一つである SCAD (Fan and Li, 2001) について述べる。また、正則化パラメーターの選択の基準として SCAD 正則化法に対する一般化情報量規準 GIC (Konishi and Kitagawa, 1996) を提案し、今後の課題について検討する。

2 基底展開法による非線形回帰モデル

複雑な非線形構造のみられるデータに対しては、特定の関数で現象の構造を捉えることは難しい。そこで、より柔軟なスプラインや B -スプライン、動径基底関数などによるモデルが有効である。

いま、観測されたデータ $\{(x_i, y_i); i = 1, 2, \dots, n\}$ に対して、基底関数に基づく回帰モデルは、基底関数 $\{b_1(x_1), \dots, b_m(x_m)\}$ を用いて

$$y_i = w_0 + \sum_{j=1}^m w_j b_j(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

と表せる。また、行列とベクトルを用いて

$$\mathbf{y} = B\mathbf{w} + \boldsymbol{\epsilon}$$

と表せる。ただし、 \mathbf{y} は、 n 次元観測値ベクトル $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ 、 B は基底関数からなる $n \times m$ 基底関数行列、 \mathbf{w} は、 m 次元パラメーターベクトル $\mathbf{w} = (w_1, w_2, \dots, w_m)^T$ 、 $\boldsymbol{\epsilon}$ は n 次元誤差ベクトル $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ とし、 ϵ_i は互いに無相関かつ、平均 0、分散 σ^2 の分布に従うものとする。ここで、切片については、データの中心化により独立して推定できる。

3 正則化最尤法

回帰モデルに含まれる未知パラメーターを推定する方法として、最小 2 乗法や最尤法がある。多くの基底関数により構成された非線形モデルでは、推定の不安定性を避けるため、対数尤度関数に正則化項 $P(\mathbf{w})$ を付与

した次の正則化対数尤度関数の最大化によって推定する.

$$\begin{aligned} \ell_p(\boldsymbol{\theta}) &= \sum_{i=1}^n \log f(y_i | \mathbf{x}_i; \mathbf{w}, \sigma^2) - \lambda P(\mathbf{w}) \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - B\mathbf{w})^T (\mathbf{y} - B\mathbf{w}) - \lambda P(\mathbf{w}) \end{aligned} \quad (3.1)$$

ここで, 式 (2.1) に対して, 誤差 ϵ が $\epsilon \sim N(0, \sigma^2 I_n)$ となるガウス型非線形回帰モデルを想定した. 推定値 $\hat{\boldsymbol{\theta}}$ は, この正則化対数尤度関数を最大とする $\boldsymbol{\theta}$ として与えられる.

4 SCAD (Smoothly Clipped Absolute Derivation)

Fan and Li (2001) は, lasso (Tibshirani, 1996) が変数選択における一貫性を持たないことを指摘し, それを改善した正則化項として SCAD (Smoothly Clipped Absolute Derivation) を提案した.

4.1 SCAD 正則化項の定義

SCAD 正則化項 $q_\lambda(|w|)$ は

$$q_\lambda(|w|) = \begin{cases} \lambda|w| & \text{if } |w| \leq \lambda \\ -\frac{(|w|^2 - 2a\lambda|w| + \lambda^2)}{2(a-1)} & \text{if } \lambda < |w| \leq a\lambda \\ \frac{(a+1)\lambda^2}{2} & \text{if } |w| > a\lambda \end{cases} \quad (4.1)$$

で定義される. ただし a, λ はチューニングパラメーターであり, $a > 2, \lambda > 0$ である.

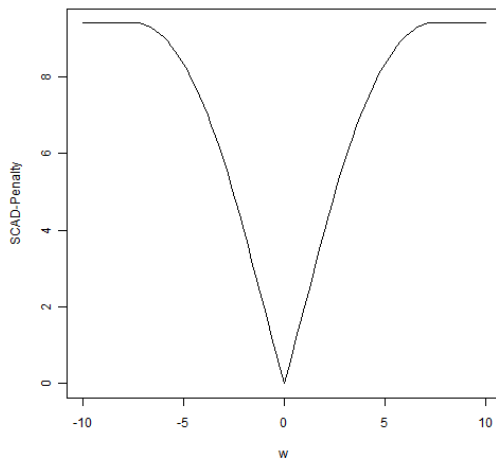


図1 SCAD-Penalty ($\lambda = 2.0, a = 3.7$)

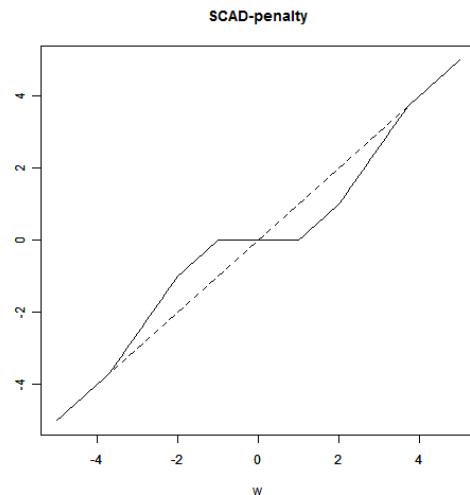


図2 SCAD-Penalty の Thresoulding rule. ($\lambda = 2.0, a = 3.7$)

SCAD 正則化項は, L_1 型正則化項であるから, 変数選択と推定を同時に行える. Lasso も同様の性質を持つが, 絶対値の大きな真のパラメーターに対して不必要なバイアスを生じさせるという問題点がある. 一方,

SCAD はこのような場合に対して、不偏性を持つという特徴がある。また、lasso は変数選択における一致性を持たないため、真に 0 であるパラメータを 0 と推定できない可能性があるのに対して、SCAD は変数選択における一致性を持つ。

4.2 SCAD 正則化法による推定

SCAD 正則化項は lasso 同様、原点で特異であるため、解析的に解を求めることができない。そこで Fan and Li (2001) では、LQA (局所 2 次近似) により罰則関数を近似することで、Newton-Raphson 法を適用可能とし、解を得る手法を提案した。

ここで、Newton-Raphson 法による反復式は

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \left[\nabla^2 \ell(\mathbf{w}^{(k)}) + nD(\mathbf{w}^{(k)}) \right]^{-1} \left[\nabla \ell(\mathbf{w}^{(k)}) + nD(\mathbf{w}^{(k)})\mathbf{w}^{(k)} \right] \quad (4.2)$$

となる。ただし、

$$\nabla = \frac{\partial}{\partial \mathbf{w}}, \quad \nabla^2 = \frac{\partial^2}{\partial \mathbf{w} \partial \mathbf{w}^T}, \quad D(\mathbf{w}^{(k)}) = \text{diag} \left\{ q'_\lambda(|w_1^{(k)}|)/|w_1^{(k)}|, \dots, q'_\lambda(|w_m^{(k)}|)/|w_m^{(k)}| \right\} \quad (4.3)$$

である。

5 SCAD 正則化非線形モデルに対する一般化情報量規準 GIC_{SCAD}

AIC や BIC は正則化最尤法には適用できない。そこで、最尤法をはじめとし、ロバスト推定やベイズ推定、そして正則化法など、様々な推定法に適用することのできる、一般化情報量規準 GIC (Konishi and Kitagawa, 1996) を用いた評価・選択を提案する。

正則化最尤法で推定した基底展開法に基づく非線形モデルに対する一般化情報量規準 GIC は

$$\text{GIC} = n(\log 2\pi + 1) + n \log \hat{\sigma}^2 + 2\text{tr}\{R(\phi_P, \hat{G})^{-1}Q(\phi_P, \hat{G})\} \quad (5.1)$$

で与えられる。ただし、 $R(\phi_P, \hat{G})$, $Q(\phi_P, \hat{G})$ は

$$R(\phi_P, \hat{G}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial \phi_P(y_i, \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_P} \quad (5.2)$$

$$Q(\phi_P, \hat{G}) = \frac{1}{n} \sum_{i=1}^n \phi_P(x, \hat{\boldsymbol{\theta}}) \frac{\partial \log f(y_i | \mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_P} \quad (5.3)$$

である。ただし、 ϕ_P は式 (3.1) の $\boldsymbol{\theta} = (\mathbf{w}, \sigma^2)^T$ に関する 1 次導関数である。正則化パラメータ λ を変化させ、GIC の値を最小とするモデルを最適なモデルとして選択する。

ここで、SCAD 正則化項が原点を除いて微分可能であるため、SCAD 正則化法により推定した基底展開法による非線形モデルに対する GIC_{SCAD} を導出するために必要な行列 R , Q は

$$\begin{aligned} R(\phi_P, \hat{G}) &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial \phi_P(y_i, \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{\text{SCAD}}} \\ &= \frac{1}{n} \begin{bmatrix} \frac{1}{\sigma^2} B^T B + nD_m & \frac{1}{\hat{\sigma}^4} B^T C \mathbf{1}_n \\ \frac{1}{\hat{\sigma}^4} \mathbf{1}_n^T C B & \frac{n}{2\hat{\sigma}^4} \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
Q(\phi_P, \hat{G}) &= \frac{1}{n} \sum_{i=1}^n \phi_P(x, \hat{\boldsymbol{\theta}}) \frac{\partial \log f(y_i | \mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{SCAD}} \\
&= \frac{1}{n} \begin{bmatrix} \frac{1}{\hat{\sigma}^4} B^T C^2 B - \frac{1}{\hat{\sigma}^2} \mathbf{q}'_{\lambda}(|\hat{\mathbf{w}}|) \mathbf{1}_n^T C B & \frac{1}{2\hat{\sigma}^6} B^T C^3 \mathbf{1}_n - \frac{1}{2\hat{\sigma}^4} B^T C \mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^6} \mathbf{1}_n^T C^3 B - \frac{1}{2\hat{\sigma}^4} \mathbf{1}_n^T C B & \frac{1}{4\hat{\sigma}^8} \mathbf{1}_n^T C^4 \mathbf{1}_n - \frac{n}{4\hat{\sigma}^4} \end{bmatrix}
\end{aligned}$$

となる。ただし

$$\begin{aligned}
D_m &= \frac{\partial^2}{\partial \mathbf{w} \partial \mathbf{w}^T} \sum_{j=1}^m q_{\lambda}(|w_j|) = \text{diag}\{q''_{\lambda}(|w_1|), \dots, q''_{\lambda}(|w_m|)\} \\
\mathbf{q}'_{\lambda}(|\mathbf{w}|) &= (q'_{\lambda}(|w_1|), \dots, q'_{\lambda}(|w_m|))^T
\end{aligned}$$

とする。これにより GIC_{SCAD} を求め、正則化パラメーターの評価・選択を行う。

6 今後の課題

本稿では、正則化法により推定した基底展開法による非線形回帰モデルについて述べ、正則化項として SCAD 正則化項を利用することで、推定と基底関数の選択を同時に行う方法を述べた。また、SCAD 正則化法による非線形モデリングや正則化パラメーターの選択基準として、 GIC_{SCAD} について理論的に述べたが、有効に機能するかどうかを数値実験を通し検証する必要がある。さらに、実データへの応用や他の評価基準との比較検証なども課題として挙げられる。

参考文献

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd Inter. Symp. on Information Theory*, Akademiai Kiado, Budapest, pp, 267-281. (Reproduced in *Breakthroughs in Statistics*, Volume 1, S. Kotz and N. L. Johnson, eds., Springer Verlag, 1992.)
- [2] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Autom. Contr.*, **AC-19**, 716-723.
- [3] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**. 1348-1359.
- [4] Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika* bf 83. 875-890.
- [5] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer, New York.
- [6] Konishi, S. (2014). *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling*. Chapman & Hall/CRC New York.
- [7] 小西貞則, 北川源四郎 (2004). 『情報量規準』. 朝倉書店.
- [8] 小西貞則 (2010). 『多変量データ解析入門-線形から非線形へ-』. 岩波書店.
- [9] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B(Methodological)*, **58**, 267-288.