Lasso を用いた非線形回帰モデルの推定と正則化の調整

九州大学大学院数理学府　松 井 秀 俊
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

正則化法は，モデル推定に伴う最適化問題において，パラメータに制約を課す推定法である．これにより，推定量にバイアスが課される一方でその分散を抑えることができ，安定した推定が可能となる．特に，Tibshirani (1996) によって提唱された，パラメータに L_1 ノルムの制約を課す lasso 推定法は，その制約の性質から変数選択の役割も担う推定法として近年盛んに研究が行われており，生命科学をはじめとした様々な分野への適用例が報告されている．

本報告では，基底関数展開に基づく非線形回帰モデルに対して，lasso を制約として用いた正則化最尤法を適用する方法について検討する．その際，適切な基底関数のみを用いてモデルを構成するために，正則化項に対して異なる重みを課し，基底関数を取捨選択する方法を提案する．正則化法による推定では，正則化の度合を調整する平滑化パラメータの選択が重要な問題となる．そこで，ベイズ型情報量規準 DIC (Spiegelhalter *et al.*, 2002) を適用して平滑化パラメータを適切に選択する．そして，提案したモデリング手法の有効性を，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 p 個の説明変数 $X_1, \cdots, X_p$ に関して，n 組のデータ $\{(y_\alpha, \boldsymbol{x}_\alpha); \alpha = 1, \cdots, n\}$ が観測されたとする．ただし $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \cdots, x_{\alpha p})'$ とする．このとき，説明変数と目的変数の関係を，次の形で与えられる基底関数展開に基づく非線形回帰モデルでモデル化する．

$$y_\alpha = \boldsymbol{w}'\boldsymbol{\phi}(\boldsymbol{x}) + \epsilon_\alpha, \quad (\alpha = 1, 2, \cdots, n). \tag{1}$$

ここで，$\boldsymbol{w} = (w_0,\ w_1,\ \cdots,\ w_m)'$ は $m+1$ 次元パラメータベクトル，$\boldsymbol{\phi}(\boldsymbol{x}) = (1,\ \phi_1(\boldsymbol{x}),\ \cdots,\ \phi_m(\boldsymbol{x}))'$ は $m+1$ 次元基底関数ベクトルとする．また，誤差 ϵ_α は互いに独立に平均 0，分散 σ^2 の正規分布に従う確率変数とする．従って，基底関数展開に基づく回帰モデルは次のように表される．

$$f(y_\alpha|\boldsymbol{w}, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(y_\alpha - \boldsymbol{w}'\boldsymbol{\phi}(\boldsymbol{x}_\alpha))^2}{2\sigma^2}\right\}. \tag{2}$$

3 モデル推定

モデルのパラメータ $\boldsymbol{w}$，σ^2 は，L^1 ノルムの制約を置いた正則化対数尤度関数

$$l_\lambda(\boldsymbol{w}, \sigma^2) = \log f(\boldsymbol{y}|\boldsymbol{w}, \sigma^2) - n\lambda \sum_{j=1}^{m} c_j |w_j|. \tag{3}$$

の最大化によって推定される．ただし，c_j は次の設定で与えられるものとする．

$$c_j = \begin{cases} \bar{h}^2/h_j^2 & (\bar{h}^2 > h_j^2) \\ 1 & (\bar{h}^2 \leq h_j^2) \end{cases}$$

ここで，$\bar{h}^2 = \sum_{j=1}^{m} h_j^2/m$ とする．すなわち，j 番目の基底関数の広がり h_j^2 が小さくなるに伴い，それに対応する平滑化パラメータにかかる重みが大きくなるように設定する．これにより，幅の狭い基底関数を適切に除去でき，通常の lasso よりも安定した推定を行うことができる．

$l_\lambda(\boldsymbol{w}, \sigma^2)$ を最大にする解を求める際，正則化項に絶対値が含まれているため，パラメータに関して微分不可能であり，解析的に推定量を導出することが困難となる．ここでは，Fu (1998) によって提案された shooting アルゴリズムを用いて推定量を導出する．

4 ベイズ型情報量規準 DIC

前節で述べた手法により得られた推定量 $\hat{\boldsymbol{w}}$ と $\hat{\sigma}^2$ は，平滑化パラメータ λ の値に依存しており，これらの値を適切に選択する必要がある．Spiegelhalter *et al.* (2002) は，データが与えられた下でのパラメータの事後分布に基づき，次の式でモデルの複雑さを定義した．

$$P_D = E_{\boldsymbol{\theta}|\boldsymbol{y}}[-2\log f(\boldsymbol{y}|\boldsymbol{\theta})] + 2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) \ . \tag{4}$$

この P_D は有効パラメータ数と呼ばれる．ただし，$f(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_\alpha f(y_\alpha|\boldsymbol{\theta})$，$\boldsymbol{\theta} = \{\boldsymbol{w}, \sigma^2\}$ とする．推定量 $\hat{\boldsymbol{\theta}}$ としては，事後平均 $\bar{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\theta}]$ や，事後分布のモード，メジアンが用いられる．有効パラメータ数 P_D で，推定されたモデルの対数尤度のバイアスを近似できるとしたのが，次の式で与えられる情報量規準 DIC である．

$$\mathrm{DIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) + 2P_D \ . \tag{5}$$

有効パラメータ数 P_D には対数の事後平均の計算が必要で，事前分布の設定によっては解析的に求めることが困難になる場合がある．このような場合でも，MCMC などを用いることで数値的に導出される．

5 数値実験

提案したモデリング手法を，人工的に発生させたデータに対して適用した．ここでは，説明変数が 1 次元および 2 次元の場合，すなわち，曲線推定と曲面推定を行った．その結果，ほぼ全ての設定において，提案手法が従来の ridge 推定，lasso 推定に比べ，予測誤差やモデルの安定性の面で優れた結果を得た．

参考文献

Ando, T., Konishi, S. and Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks. *J. Statist. Plann. Inference* **138**, 3616–3633.

Fu, W. J. (1998). Penalized Regression: The bridge versus the lasso. *J. Comput. Graph. Statist.* **7**, 397–416.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc. Ser. B* **64**, 583–616.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* **58**, 267–288.