ブリッジ型正則化法におけるチューニングパラメータ選択のための AIC

九州大学　大学院数理学府　梅津 佑太
九州大学　マス・フォア・インダストリ研究所　二宮 嘉行

概 要

L_γ 罰則を課したブリッジ型正則化法 (Frank and Friedman, 1993) において，$\gamma < 1$ の場合パラメータの推定と変数選択を同時に実行可能であることが知られている．特に，$\gamma = 1$ のときは Lasso (Tibshirani, 1996) とよばれ，統計解析や機械学習において広く研究されている．罰則付き最尤推定では，チューニングパラメータの選択がしばしば問題になるが，Ninomiya and Kawano (2014) は，一般化線形モデルにおける Lasso 推定量の漸近分布に基づき，AIC を導出してチューニングパラメータを選択することを提案している．ところが，彼らのモデルではパラメータ推定の一致性が成立しない．本報告では，パラメータ推定の一致性を保持するように，Ninomiya and Kawano (2014) とは異なる漸近論の枠組みで，ブリッジ型正則化法におけるチューニングパラメータ選択のための AIC の導出について述べる．

主結果　n 個の説明変数 $\boldsymbol{X}_i \in \mathcal{X} \subset \mathbb{R}^{r\times p}$ と目的変数 $\boldsymbol{y}_i \in \mathbb{R}^r$ に対して，自然連結関数 $\boldsymbol{\theta}_i = \boldsymbol{X}_i\boldsymbol{\beta} \in \Theta$（自然パラメータ空間）をもつ次の一般化線形モデルを考える．

$$g_i(\boldsymbol{\beta}) = \log f(\boldsymbol{y}_i; \boldsymbol{X}_i\boldsymbol{\beta}) = \boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{X}_i\boldsymbol{\beta} - a(\boldsymbol{X}_i\boldsymbol{\beta}) + b(\boldsymbol{y}_i)$$

真のパラメータ $\boldsymbol{\beta}^*$ はこのモデルに含まれているとし，$\boldsymbol{\beta} \in \mathcal{B} \subset \mathbb{R}^p$ は次で推定されるとする．

$$\hat{\boldsymbol{\beta}}_\lambda = \mathop{\mathrm{argmin}}_{\boldsymbol{\beta}\in\mathcal{B}} \left\{ -\sum_{i=1}^n g_i(\boldsymbol{\beta}) + \sqrt{n}\lambda \sum_{j=1}^p |\beta_j|^\gamma \right\},\quad 0 < \gamma < 1$$

ここで，説明関数に関して次の条件を仮定する．

(C1) $\mathcal{X}$ はコンパクト，$\mathcal{B}$ は開凸集合であり，任意の $\boldsymbol{X} \in \mathcal{X}$ と $\boldsymbol{\beta} \in \mathcal{B}$ に対して，$\boldsymbol{X}\boldsymbol{\beta} \in \Theta^{\mathrm{o}}$ である．

(C2) $\mathcal{X}$ 上の不変分布が存在する．特に，$\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} a''(\boldsymbol{X}_i\boldsymbol{\beta})\boldsymbol{X}_i/n \to \boldsymbol{J}(\boldsymbol{\beta})$ は正定値行列である．

また，$\boldsymbol{s}_n = \sum_{i=1}^n g'(\boldsymbol{\beta}^*)/\sqrt{n}$ とする．(C1) および (C2) は，一般化線形モデルにおける最尤推定量の一致性と，$\boldsymbol{s}_n$ の漸近正規性を保証するための仮定である．このとき，$\mathcal{J}^{(1)} = \{j; \beta_j^* = 0\}$, $\mathcal{J}^{(2)} = \{j; \beta_j^* \neq 0\}$ に対して，ベクトル $(\beta_i)_{i\in\mathcal{J}^{(k)}}$ を $\boldsymbol{\beta}^{(k)}$, 行列 $(\boldsymbol{J}(\boldsymbol{\beta}^*)_{ij})_{i\in\mathcal{J}^{(k)}, j\in\mathcal{J}^{(l)}}$ を $\boldsymbol{J}^{(kl)}(\boldsymbol{\beta}^*)$ $(k, l \in \{1, 2\})$ などと表せば，推定量の一致性および漸近分布

$$\sqrt{n}^{1/\gamma}\hat{\boldsymbol{\beta}}_\lambda^{(1)} = o_p(1),\quad \sqrt{n}(\hat{\boldsymbol{\beta}}_\lambda^{(2)} - \boldsymbol{\beta}^{*(2)}) = \boldsymbol{J}^{(22)}(\boldsymbol{\beta}^*)^{-1}(\boldsymbol{s}_n^{(2)} - \boldsymbol{p}_\lambda^{(2)}) + o_p(1)$$

を導出することができる．ただし，$\boldsymbol{s}_n^{(2)} \to^d N(\boldsymbol{0}, \boldsymbol{J}^{(22)}(\boldsymbol{\beta}^*))$ であり，$\boldsymbol{p}_\lambda^{(2)} = (\gamma\lambda\mathrm{sgn}(\beta_j^*)|\beta_j^*|^{\gamma-1})_{j\in\mathcal{J}^{(2)}}$ とする．さらに，Kullback-Leibler 情報量の漸近不偏推定量として，情報量規準 AIC の漸近バイアスが $|\mathcal{J}^{(2)}|$ となることを示すことができる．$\mathcal{J}^{(2)}$ は未知の量なので，その一致推定量 $\hat{\mathcal{J}}^{(2)} = \{j; \hat{\beta}_{\lambda,j} \neq 0\}$ を代用する．結果，次の値が最小となる λ を最適なものとして選べばよく，簡単にモデル選択することができる．

$$\mathrm{AIC}_\lambda = -2\sum_{i=1}^n g_i(\hat{\boldsymbol{\beta}}_\lambda) + 2|\hat{\mathcal{J}}^{(2)}|$$

数値実験の結果は当日報告する．

参考文献

- Frank, L. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109-135.
- Ninomiya, Y. and Kawano, S. (2014). AIC for the LASSO in generalized linear models. *ISM Research Memorandum,* 1187.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **58**, 267-288.