一般化線形モデルにおける Group Lasso に対する情報量規準

総合研究大学院大学 複合科学研究科統計科学専攻 小松 慧
JA 共済連 山下 雄大
統計数理研究所 数理・推論研究系 二宮 嘉行

概要:　共変量がクラスター化されてグループになっているとき, group Lasso (Yuan and Lin [1]) はグループごと係数を 0 にする傾向があるのでよく利用される. group Lasso にはペナルティの強さを調節するチューニングパラメータが含まれており, その適切な値を選択するために真の予測誤差の不偏推定量が C_p 型の情報量規準として与えられている. しかしながら, それは一般化線形モデルにおいては与えられていない. そこで本講演では, 一般化線形モデルの枠組みのもと group Lasso に対する AIC をその元来の定義に基づいて導き, 数値実験において cross-validation など他手法との性能比較を行う.

サンプル i $(i=1,\dots,n)$ に対し $\boldsymbol{y}_i, \boldsymbol{X}_i$ はそれぞれ r 次元目的変数ベクトル, $r \times p$ 説明変数行列とする. 自然パラメータ $\boldsymbol{\theta}$ を持つ密度が $f(\boldsymbol{y};\boldsymbol{\theta}) = \exp\{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\theta} - a(\boldsymbol{\theta}) + b(\boldsymbol{y})\}$ となる指数型分布族を考える. そして自然リンク関数を用いた一般化線形モデルを考え, $\boldsymbol{y}_i, \boldsymbol{X}_i$ に対する係数ベクトル $\boldsymbol{\beta} = (\beta_1,\dots,\beta_p)^{\mathrm{T}}$ の対数尤度を $g_{\boldsymbol{y}_i,\boldsymbol{X}_i}(\boldsymbol{\beta}) = \log f(\boldsymbol{y}_i;\boldsymbol{X}_i\boldsymbol{\beta})$ と書くことにする. また, $\mathcal{D}$ を添字集合の素な分割, つまり $\bigcup_{d\in\mathcal{D}} d = \{1,\dots,p\}$ で $d, d' \in \mathcal{D}$ $(d \neq d')$ に対して $d \cap d' = \phi$ とする. ここで $\mathcal{D}$ における group Lasso 推定量として以下を考える.

$$\hat{\boldsymbol{\beta}}_\lambda \equiv \underset{\boldsymbol{\beta}}{\mathrm{argmin}} \left\{ -\sum_{i=1}^{n} g_{\boldsymbol{y}_i,\boldsymbol{X}_i}(\boldsymbol{\beta}) + n\lambda \sum_{d\in\mathcal{D}} \|\boldsymbol{\beta}_d\| \right\}$$

ただし $\|\cdot\|$ は L_2 ノルム, $\boldsymbol{\beta}_d = (\beta_j)_{j\in d}$ は $\boldsymbol{\beta}$ の部分ベクトルである. このとき正則化パラメータである λ の妥当な値を与えるための AIC を以下で与える (Komatsu et al [2]).

$$\mathrm{AIC}_\lambda^{\mathrm{GL}} = -2\sum_{i=1}^{n} g_{\boldsymbol{y}_i,\boldsymbol{X}_i}(\hat{\boldsymbol{\beta}}_\lambda) + 2\mathrm{tr}\{\hat{\boldsymbol{J}}_n^{*(22)}(\hat{\boldsymbol{J}}_n^{**(22)} + \lambda\hat{\boldsymbol{L}}^{**(22)})^{-1}\}$$

ただし $\hat{\boldsymbol{\beta}}_{\lambda,d}$ は $\boldsymbol{\beta}_d$ の group Lasso 推定量であり, $\hat{\mathcal{D}}^{(2)} = \{d \mid \hat{\boldsymbol{\beta}}_{\lambda,d} \neq 0\}, \hat{\mathcal{J}}^{(2)} = \{j \in d \mid d \in \hat{\mathcal{D}}^{(2)}\}$ とする. また $\boldsymbol{J}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{X}_i^{\mathrm{T}} a''(\boldsymbol{X}_i\boldsymbol{\beta})\boldsymbol{X}_i/n, \boldsymbol{L}_d(\boldsymbol{\beta}) = (\|\boldsymbol{\beta}_d\|^2 I_{|d|} - \boldsymbol{\beta}_d\boldsymbol{\beta}_d^{\mathrm{T}})/\|\boldsymbol{\beta}_d\|^3$ に対し, $|\hat{\mathcal{J}}^{(2)}| \times |\hat{\mathcal{J}}^{(2)}|$ 行列 $\hat{\boldsymbol{J}}_n^{*(22)}$, $\hat{\boldsymbol{J}}_n^{**(22)}$, $\hat{\boldsymbol{L}}^{**(22)}$ をそれぞれ $(\boldsymbol{J}_n(\hat{\boldsymbol{\beta}}_0)_{jk})_{j\in\hat{\mathcal{J}}^{(2)},k\in\hat{\mathcal{J}}^{(2)}}$, $(\boldsymbol{J}_n(\hat{\boldsymbol{\beta}}_\lambda)_{jk})_{j\in\hat{\mathcal{J}}^{(2)},k\in\hat{\mathcal{J}}^{(2)}}$,$\mathrm{diag}_{d\in\hat{\mathcal{D}}^{(2)}}(\boldsymbol{L}_d(\hat{\boldsymbol{\beta}}_\lambda))$ としている. ここで $\mathcal{D}$ の要素が全て 1, つまり Lasso を考えると上記で与えた AIC は Ninomiya and Kawano [3] で与えられたものと一致する.

参考文献

[1] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, Vol. 68, No. 1, pp. 49–67, 2006.

[2] Satoshi Komatsu, Yuta Yamashita, and Yoshiyuki Ninomiya. AIC for the group Lasso in generalized linear models. *Japanese Journal of Statistics and Data Science*, 2019 (accepted).

[3] Yoshiyuki Ninomiya and Shuichi Kawano. AIC for the Lasso in generalized linear models. *Electronic Journal of Statistics*, Vol. 10, No. 2, pp. 2537–2560, 2016.