LASSO のための AIC 型情報量規準

二宮　嘉行　数理・推論研究系　教授

問題

LASSO

- パラメータ $\boldsymbol{\beta} = (\beta_j)_{j\in\{1,\ldots,p\}}$ に関する推定関数に ℓ_1 罰則 $\lambda||\boldsymbol{\beta}||_1$ を課す正則化法
- λ (>0): 正則化パラメータ

λ の適切な値を与えることを問題と考える

背景

既存の結果

- 計算負荷を考えないなら，交差検証法 (CV) や stability selection といった計算機的手法がよいかもしれない
- 解析的手法が発展しているが，それらのほとんどには恣意性が残る

弱点のない解析的手法

- 正規線形回帰ならば，予測誤差の不偏推定量が “一般化 $\mathrm{C_P}$” として導かれている (Efron et al. 2004, Zou et al. 2007)
- 言い換えれば，AICc (AIC の有限補正) が得られている
- それは容易に求まる形をしているが，Stein 法に基づいているため一般化されない

目的とモデル

一般化線形モデルにおける LASSO のため，元来の定義に基づいて（つまり漸近理論に基づいて）AIC を導出することが目的

データ $\{(\boldsymbol{y}_i, \boldsymbol{X}_i) : 1 \le i \le n\}$

- 目的変数 $\boldsymbol{y}_i$: 独立な r 次元確率ベクトル
- 説明変数 $\boldsymbol{X}_i$: $r \times p$ 定数行列

自然リンクを用いた一般化線形モデル

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

- $\boldsymbol{\beta}$ の真値を $\boldsymbol{\beta}^*$ と書くことにする

仮定

(C1)　$\{\boldsymbol{X}_i\}$ はコンパクト集合上にある

(C2)　各 $\boldsymbol{\beta}$ に対し，$\sum_{i=1}^n a(\boldsymbol{X}_i\boldsymbol{\beta})/n$ や $\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} a'(\boldsymbol{X}_i\boldsymbol{\beta})/n$，$\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} a''(\boldsymbol{X}_i\boldsymbol{\beta})\boldsymbol{X}_i/n$ はレート $\mathrm{o}(n^{-1/2})$ で収束し，特に最後の極限は正定値となる

(C3)　$h(\boldsymbol{\beta})$ を $\lim_{n\to\infty}\sum_{i=1}^n\{a(\boldsymbol{X}_i\boldsymbol{\beta}) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i^{\mathrm{T}} a'(\boldsymbol{X}_i\boldsymbol{\beta}^*)\}/n$ としたとき，$h(\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1$ は唯一の最小値をもつ

漸近論

LASSO 推定量

$$\hat{\boldsymbol{\beta}}_\lambda \equiv \mathrm{argmin}_{\boldsymbol{\beta}}\{-\textstyle\sum_{i=1}^n \log f(\boldsymbol{y}_i; \boldsymbol{X}_i\boldsymbol{\beta}) + n\lambda||\boldsymbol{\beta}||_1\}$$

補題 (Knight & Fu 2000)

$$\hat{\boldsymbol{\beta}}_\lambda \xrightarrow{\mathrm{p}} \boldsymbol{\beta}^{**} \equiv \mathrm{argmin}_{\boldsymbol{\beta}}\{h(\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1\}$$

記法

- $\mathcal{J}^{(1)} \equiv \{j : \beta_j^{**} = 0\}$, $\mathcal{J}^{(2)} \equiv \{j : \beta_j^{**} \ne 0\}$
- $\boldsymbol{\beta}^{(1)} \equiv (\beta_j)_{j\in\mathcal{J}^{(1)}}$, $\boldsymbol{\beta}^{(2)} \equiv (\beta_j)_{j\in\mathcal{J}^{(2)}}$
- $\boldsymbol{J}_{\boldsymbol{\beta}}^{(22)} \equiv ((\lim_{n\to\infty}\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} a''(\boldsymbol{X}_i\boldsymbol{\beta})\boldsymbol{X}_i/n)_{jk})_{j,k\in\mathcal{J}^{(2)}}$

定理 1 (極限分布)

$$n(\hat{\boldsymbol{\beta}}_\lambda^{(1)} - \boldsymbol{\beta}^{**(1)}) \xrightarrow{\mathrm{p}} \boldsymbol{0}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\lambda^{(2)} - \boldsymbol{\beta}^{**(2)}) \xrightarrow{\mathrm{d}} \mathrm{N}(\boldsymbol{0}, \boldsymbol{J}_{\boldsymbol{\beta}^{**}}^{(22)-1}\boldsymbol{J}_{\boldsymbol{\beta}^*}^{(22)}\boldsymbol{J}_{\boldsymbol{\beta}^{**}}^{(22)-1})$$

バイアス評価

定理 2 (バイアス評価)

AIC 導出の際に現れる漸近バイアスは今の場合では $\mathrm{tr}(\boldsymbol{J}_{\boldsymbol{\beta}^*}^{(22)}\boldsymbol{J}_{\boldsymbol{\beta}^{**}}^{(22)-1})$ で与えられる

一致推定量による置き換えのための準備

- $\hat{\mathcal{J}}^{(2)} \equiv \{j : \hat{\beta}_{\lambda,j} \ne 0\}$: アクティブセット
- $\hat{\boldsymbol{J}}_{\boldsymbol{\beta}}^{(22)} \equiv ((\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} a''(\boldsymbol{X}_i\boldsymbol{\beta})\boldsymbol{X}_i/n)_{jk})_{j,k\in\hat{\mathcal{J}}^{(2)}}$
- $\hat{\boldsymbol{\beta}}_0$: 最尤推定量

LASSO のための AIC (Ninomiya & Kawano 2016)

$$-2\sum_{i=1}^n \log f(\boldsymbol{y}_i; \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_\lambda) + 2\mathrm{tr}\big(\hat{\boldsymbol{J}}_{\hat{\boldsymbol{\beta}}_0}^{(22)}\hat{\boldsymbol{J}}_{\hat{\boldsymbol{\beta}}_\lambda}^{(22)-1}\big)$$

- 正規線形回帰では，これは

$$\mathrm{AICc} = -2\textstyle\sum_{i=1}^n \log f(\boldsymbol{y}_i; \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_\lambda) + 2|\hat{\mathcal{J}}^{(2)}|$$

となり，つまり AICc の一般化とみなせる

数値実験

Yuan & Lin (2007) で扱われているグラフィカル LASSO
(n: サンプルサイズ, q: 変数の数)

(n,q)	method	KL mean (sd)	FP	FN
(25,5)	CV	4.934 (0.979)	1.92	1.43
	AIC	4.907 (1.006)	2.10	0.99
	AICc	4.949 (1.059)	2.35	0.57
(50,5)	CV	4.410 (0.603)	2.21	0.54
	AIC	4.397 (0.607)	2.19	0.43
	AICc	4.393 (0.611)	2.36	0.26
(50,10)	CV	8.749 (0.893)	16.47	1.31
	AIC	8.752 (0.923)	19.67	0.80
	AICc	8.800 (0.961)	21.48	0.61
(100,10)	CV	8.077 (0.588)	19.95	0.18
	AIC	8.052 (0.590)	21.30	0.11
	AICc	8.058 (0.588)	20.82	0.13

引用文献

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), Least angle regression, *Ann. Statist.* 32, 407–499.

Knight, K. and Fu, W. (2000), Asymptotics for lasso-type estimators, *Ann. Statist.* 28, 1356–1378.

Ninomiya, Y. and Kawano, S. (2016), AIC for the LASSO in generalized linear models, *Electronic Journal of Statistics* 10, 2537–2560.

Yuan, M. and Lin, Y. (2007), Model selection and estimation in the Gaussian graphical model, *Biometrika* 94, 19–35.

Zou, H., Hastie, T. and Tibshirani, R. (2007), On the “degrees of freedom” of the lasso, *Ann. Statist.* 35, 2173–2192.