Oracle inequalities for sign constrained generalized linear models

田上 悠太　リスク解析戦略センター　特任助教

背景

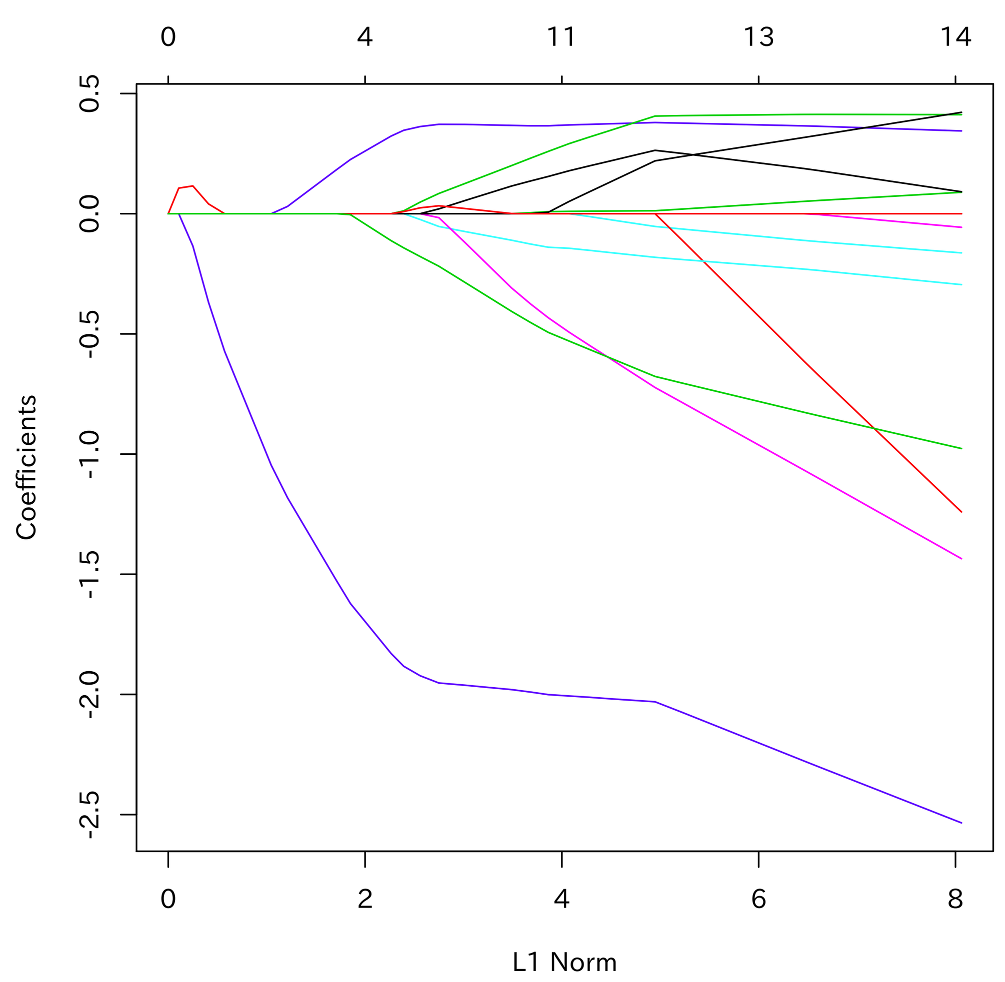
- データ取得技術の向上により、高次元データの研究利用が可能になっている。
  - マイクロアレイデータ
  - 画像データ
  - 財務データ
- 高次元データに対して、スパース推定という方法が用いられる。
  - Lasso (Tibshirani [1996])
  - SCAD (Fan and Li [2001])
  - MCP (Zhang et al. [2010])
  - ブリッジ回帰 (Frank and Friedman [1993])
  - Elastic net (Zou and Hastie [2005])
- これらの方法は、チューニングパラメータを含んでおり、それを選択しなければいけない問題がある。
  - Lasso

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(||\boldsymbol{y} - \boldsymbol{X\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1) \tag{1}$$

  - Elastic net

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(||\boldsymbol{y} - \boldsymbol{X\beta}||_2^2 + \lambda_2||\boldsymbol{\beta}||^2 + \lambda_1||\boldsymbol{\beta}||_1) \tag{2}$$

- チューニングパラメータの選び方によって大きく結果が変わってしまう問題がある (以下、Lassoの場合)。



- 本研究ではチューニングパラメータを含まない、Sign constrained regressionという方法の統計的性質に関して研究した。

設定

確率空間 $(\Omega, \mathcal{F}, P)$ を考える。$\mathbf{F}$ を実内積空間、 $L(f,\omega) : \mathbf{F}\times\Omega \to \mathbb{R}$ とする。また、$L(f) = L(f,\cdot) \in L^1(P)$ とする。この時、$\mathbf{F}$ をパラメータ空間、$L(f)$ を $f \in \mathbf{F}$ のロス関数と考える。そして、$\bar{L} : \mathbf{F} \to \mathbb{R}$

$$\bar{L}(f) = E[L(f)], \qquad f \in \mathbf{F}.$$

で期待損失を定義する。また、$f^0 \in \mathbf{F}$

$$\bar{L}(f^0) = \min_{f\in\mathbf{F}} \bar{L}(f)$$

とする。更に、$\mathcal{E} : \mathbf{F} \to [0,\infty)$

$$\mathcal{E}(f) = \bar{L}(f) - \bar{L}(f^0), \qquad f \in \mathbf{F}.$$

で超過ロスを定義する。

この時、以下の f_β で f^0 を推定する問題を考える。

$$f_\beta := \sum_{j=1}^p \beta_j\psi_j$$

ただし、$\psi_1,\dots,\psi_p \in \mathbf{F}$。更に、以下の様に符号制約 (sign-constraint) を課した場合を考える。

$$L(f_{\hat\beta}) = \min_{\beta\succeq 0} L(f_\beta).$$

結果

仮定1 (Convexity) $\beta \mapsto L(f_\beta,\omega)$ は任意の $\omega\in\Omega$ に対して、凸である。

仮定2 (Quadratic margin) 以下を満たす $c > 0$、$\mathbf{F}_{\mathrm{local}} \subset \mathbf{F}$ が存在する。

$$\forall f \in \mathbf{F}_{\mathrm{local}},\ \mathcal{E}(f) \ge c\|f - f^0\|^2$$

- $S \subset \{1,\cdots,p\}$
- $\Sigma = (\Sigma^{ij})_{1\le i,j\le p} = (\langle\psi_i,\psi_j\rangle)_{1\le i,j\le p}$
- $\phi^2(L,S) = \min\{|S|\beta^\top\Sigma\beta : \|\beta_S\|_1 = 1, \|\beta_{S^c}\| \le L\}$
- $\beta^* = \arg\min_{\beta\succeq 0}\bar{L}(f_\beta)$.
- $\phi^2_{pos}(\Sigma) = \min\{\|f_\beta\|^2 : \|\beta\|_1 = 1, \beta \succeq 0\}$
- $\mathbf{Z}_M = \sup_{\beta:\|\beta-\beta^*\|_1\le M}|v(\beta) - v(\beta^*)|$
- $f^* = f_{\beta^*}$, $\hat{f} = f_{\hat\beta}$, $S_* = \{j : \beta_j^* \neq 0\}$ and $s_* = |S_*|$

定理

仮定1、2、さらに $\nu := \phi^2_{pos}(\Sigma) > 0$、$\phi_*^2 := \phi^2(3C/\nu, S_*) > 0$、ただし、$C := \max_{1\le j\le p}\|\psi_j\|^2$ が満たされている時、$\lambda > 0$ で

$$\varepsilon_\lambda = 4\mathcal{E}(f^*) + \frac{12}{c\phi_*^2}\left(1 + \frac{3C}{\nu}\right)^2 s_*\lambda^2, \qquad M_\lambda = \varepsilon_\lambda/\lambda.$$

とする。また、$f_\beta \in \mathbf{F}_{\mathrm{local}}$ で $\beta\in\mathbb{R}^p$ が $\|\beta - \beta^*\|_1 \le M$ を満たすとする。

この時、$\{\mathbf{Z}_{M_\lambda} \le \varepsilon_\lambda\}$ において、

$$\|\hat\beta - \beta^*\|_1 \le M_\lambda = \frac{4}{\lambda}\mathcal{E}(f^*) + \frac{12}{c\phi_*^2}\left(1 + \frac{3C}{\nu}\right)^2 s_*\lambda \tag{3}$$

さらに、

$$\mathcal{E}(\hat f) \le \frac{5}{4}\varepsilon_\lambda = 5\mathcal{E}(f^*) + \frac{15}{c\phi_*^2}\left(1 + \frac{3C}{\nu}\right)^2 s_*\lambda^2 \tag{4}$$

参考文献

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

LLdiko E Frank and Jerome H Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Cun-Hui Zhang et al. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2):894–942, 2010.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.