An unibiased C_p criterion for multivariate ridge regression

広島大学　大学院理学研究科　柳原宏和
広島大学　原爆放射線医科学研究所　佐藤健一

Hoerl and Kennard (1970) によって提案されたリッジ回帰を多変量に拡張した多変量リッジ回帰 (Multivariate Ridge Regression) において, 最適なリッジパラメータと説明変数の組み合わせを選ぶための C_p 規準のバイアスを補正した修正 C_p (Modified C_p; MC_p) 規準を提案した. さらに, 提案した MC_p 規準の数理的特性を調べた.

1. モデルとリッジ推定量

$\boldsymbol{Y}$ を $n \times p$ 目的変数行列, $\boldsymbol{X}$ を $\mathrm{rank}(\boldsymbol{X}) = k$ であるような $n \times k$ 説明変数行列とする. ただし, n は標本数であるとする. ここで, j を k_j 個の成分を持つ $\omega = \{1, \ldots, k\}$ の部分集合とし, $\boldsymbol{X}_j$ を集合 j の成分に対応した $\boldsymbol{X}$ の列を取り出した $\boldsymbol{X}$ の $n \times k_j$ 部分行列とする. このとき, k_j 個の説明変数を持つ以下のような候補のモデルを考える:

$$\boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}_j \boldsymbol{\Xi}_j, \boldsymbol{\Sigma}_j \otimes \boldsymbol{I}_n). \tag{1}$$

ここで, $j = \omega$ のとき, つまり $\boldsymbol{X}_\omega = \boldsymbol{X}$ のときのモデルをフルモデルと呼ぶとする. (1) 式において, $\boldsymbol{X}_j$ のリッジ推定量は

$$\hat{\boldsymbol{\Xi}}_{j,\theta} = \boldsymbol{M}_{j,\theta}^{-1} \boldsymbol{X}_j' \boldsymbol{Y}, \tag{2}$$

となる. ただし $\boldsymbol{M}_{j,\theta} = \boldsymbol{X}_j'\boldsymbol{X}_j + \theta \boldsymbol{I}_{k_j}$ とし, θ は非負のリッジパラメータとする. (2) 式において, $\boldsymbol{\Xi}_j$ の最尤推定量 (または最小二乗推定量) は $\hat{\boldsymbol{\Xi}}_{j,0}$ となることに注意してほしい.

さらに, 目的変数 $\boldsymbol{Y}$ は以下のような真のモデルに従っているとする:

$$\boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{\Gamma}_*, \boldsymbol{\Sigma}_* \otimes \boldsymbol{I}_n). \tag{3}$$

ここで, $\boldsymbol{P}_{\boldsymbol{A}}$ を $\boldsymbol{A}$ の列ベクトルで張る空間への射影行列, つまり, $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ であるとする. このとき, 少なくともフルモデルは真のモデルを含んでいる, つまり, $\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\Gamma}_* = \boldsymbol{\Gamma}_*$ が満たされていると仮定する.

2. リスク関数と MC_p 規準

$\hat{\boldsymbol{Y}}_{j,\theta} = \boldsymbol{X}_j \hat{\boldsymbol{\Xi}}_{j,\theta}$ とし, $\boldsymbol{U}$ を $\boldsymbol{Y}$ と独立に同一の分布に従う $n \times p$ 確率変数行列とする. このとき, Fujikoshi and Satoh (1997) により提案された予測平均二乗誤差に基づくリスク関数は,

$$R(j,\theta) = E_{\boldsymbol{Y}}^* E_{\boldsymbol{U}}^* \left[\mathrm{tr}\left\{ (\boldsymbol{U} - \hat{\boldsymbol{Y}}_{j,\theta}) \boldsymbol{\Sigma}_*^{-1} (\boldsymbol{U} - \hat{\boldsymbol{Y}}_{j,\theta})' \right\} \right], \tag{4}$$

と定義される. ただし, E^* は真のモデル (3) の下での期待値を表すとする.

今, $\boldsymbol{W}_{j,\theta}$ をリッジ推定量に基づく残差行列,

$$\boldsymbol{W}_{j,\theta} = (\boldsymbol{Y} - \hat{\boldsymbol{Y}}_{j,\theta})'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_{j,\theta}),$$

とする．また, $\boldsymbol{S}$ をフルモデルの下での $\boldsymbol{\Sigma}_*$ の不偏推定量とする．つまり $\boldsymbol{S} = \boldsymbol{W}_{\omega,0}/(n-k)$ である．このとき, リスク関数 (4) の簡便推定量である C_p 規準は,

$$C_p(j,\theta) = \mathrm{tr}(\boldsymbol{W}_{j,\theta}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{j,\theta}^{-1}\boldsymbol{M}_{j,0}), \tag{5}$$

で定義することができる．しかしながら, (5) 式で定義された C_p 規準はリスク関数 (4) の不偏推定量でなく, バイアスが存在する．そのバイアスを補正した MC_p 規準は,

$$MC_p(j,\theta) = \left(1 - \frac{p+1}{n-k}\right)\mathrm{tr}(\boldsymbol{W}_{j,\theta}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{j,\theta}^{-1}\boldsymbol{M}_{j,0}) + p(p+1), \tag{6}$$

によって定義することができる．

3. MC_p 規準の数理的特性

(6) 式で定義された MC_p 規準は以下の特性を持つ．

1. $\boldsymbol{Y}$ がどのような分布であっても不等式 $MC_p(j,\theta) \leq C_p(j,\theta)$ が成り立つ．
2. $\boldsymbol{Y}$ が正規分布で $\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\Gamma}_* = \boldsymbol{\Gamma}_*$ が成り立てば, 関係式 $E_{\boldsymbol{Y}}^*[MC_p(j,\theta)] = R(j,\theta) \leq E_{\boldsymbol{Y}}^*[C_p(j,\theta)]$ が成り立つ．
3. $\boldsymbol{Y}$ がどのような分布であっても不等式 $Var[MC_p(j,\theta)] \leq Var[C_p(j,\theta)]$ が成り立つ．
4. $\boldsymbol{Y}$ が正規分布で $\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\Gamma}_* = \boldsymbol{\Gamma}_*$ が成り立てば $MC_p(j,\theta)$ はリスク関数 $R(j,\theta)$ の一様最小分散不偏推定量である．
5. $\hat{\theta}_j^{(m)}$ と $\hat{\theta}_j^{(c)}$ をそれぞれ j を固定した下で MC_p 規準と C_p 規準を最小にするリッジパラメータ θ とする．このとき, $\boldsymbol{Y}$ がどのような分布であっても不等式 $\hat{\theta}_j^{(m)} \geq \hat{\theta}_j^{(c)}$ が成り立つ．
6. $\hat{j}_\theta^{(m)}$ と $\hat{j}_\theta^{(c)}$ をそれぞれ θ を固定した下で MC_p 規準と C_p 規準を最小にする説明変数の組み合わせ j とする．このとき, $\boldsymbol{Y}$ がどのような分布であっても関係式 $\hat{j}_\theta^{(c)} \not\subset \hat{j}_\theta^{(m)}$ が成り立つ．特に, 候補のモデルが Nested モデルあれば, 関係式 $\hat{j}_\theta^{(m)} \subseteq \hat{j}_\theta^{(c)}$ が成り立つ．
7. $\hat{\theta}^{(m)}$ と $\hat{j}^{(m)}$ を MC_p 規準を最小にする θ と j とし, さらに, $\hat{\theta}^{(c)}$ と $\hat{j}^{(c)}$ を C_p 規準を最小にする θ と j とする．このとき, $\boldsymbol{Y}$ がどのような分布であっても $\hat{\theta}^{(m)} \geq \hat{\theta}^{(c)}$ か $\hat{j}^{(c)} \not\subset \hat{j}^{(m)}$ のどちらかの関係式が成り立つ．

引用文献：

[1] Fujikoshi, Y. & Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika*, **84**, 707–716.

[2] Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.