Group Lasso 型罰則項を伴う重み付き残差平方和の最小化に基づく多変量線形回帰モデルの推定

広島大・理　　小田 凌也　　柳原 宏和

複数の目的変数と説明変数からなる多変量線形回帰モデルにおいて, 回帰係数行列を推定する手法を扱う. 近年, 推定と同時に説明変数の選択も行える罰則付き推定法がしばしば用いられる. 多変量線形回帰モデルにおける説明変数の選択では, 各目的変数ごとに選択するのではなく, すべての目的変数に対してその説明変数が必要か不必要かを選択する. そのため, 説明変数の選択は回帰係数行列の列ベクトルの選択とみなされる. 回帰係数行列の推定と説明変数の選択を同時に行うための方法として, Yuan & Lin (2006) により提案された Group Lasso を適用した以下のような罰則付き推定法がある (Simila & Tikka, 2007, 参照):

$$\hat{\boldsymbol{\Theta}}_{\lambda,1} = \arg\min_{\boldsymbol{\Theta}} \left\{ \mathrm{tr}\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})'\} + \lambda \sum_{j=1}^{k} ||\boldsymbol{\theta}_j|| \right\}. \tag{1}$$

ただし, $\boldsymbol{Y}$ は $n \times p$ 目的変数行列, $\boldsymbol{X}$ は $n \times k$ 説明変数行列, $\boldsymbol{\Theta}$ は $k \times p$ 回帰係数行列, $\boldsymbol{\theta}_j$ は $\boldsymbol{\Theta}$ の第 j 行ベクトル, λ はチューニングパラメータであり, $||\cdot||$ はユークリッドノルムを表す. ここで, $\boldsymbol{Y}$, $\boldsymbol{X}$ の各列ベクトルは中心化され標準偏差で割られているとする. (1) 式による推定問題は解析的に解けないため推定アルゴリズムが必要となるが, そのアルゴリズムの 1 つとして Friedman *et al.* (2007) による座標降下法が知られている.

しかしながら, (1) 式は重みを用いない残差平方和を基にした推定法であり, 目的変数間の相関が考慮されていないため推定精度が悪くなる可能性がある. そこで本発表では標本共分散行列 $\boldsymbol{S}$ により重み付けされた残差平方和を基にした以下のような罰則付き推定法を提案する:

$$\hat{\boldsymbol{\Theta}}_{\lambda,2} = \arg\min_{\boldsymbol{\Theta}} \left\{ \mathrm{tr}\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})\boldsymbol{S}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})'\} + \lambda \sum_{j=1}^{k} ||\boldsymbol{\theta}_j|| \right\}. \tag{2}$$

(2) 式のように $\boldsymbol{S}$ により重み付けされた残差平方和に基づく推定法として, Xin *et al.* (2017) では以下の罰則付き推定法が用いられている:

$$\hat{\boldsymbol{\Theta}}_{\lambda,3} = \arg\min_{\boldsymbol{\Theta}} \left\{ \mathrm{tr}\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})\boldsymbol{S}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})'\} + \lambda \sum_{j=1}^{k} ||\boldsymbol{S}^{-1/2}\boldsymbol{\theta}_j|| \right\}. \tag{3}$$

(3) 式における座標降下法は $(\boldsymbol{Y}\boldsymbol{S}^{-1/2}, \boldsymbol{\Theta}\boldsymbol{S}^{-1/2}) \to (\boldsymbol{Y}, \boldsymbol{\Theta})$ と変換し直すことで (1) 式における座標降下法と同等となる. しかしながら, (2) 式における座標降下法は同等ではなくそのアルゴリズムは求められていない. そのため本発表では, (2) 式における座標降下法を導出する. 発表当日は, (1), (2), (3) 式による罰則付き推定法により得られた回帰係数行列の推定量の比較を行う.

引用文献：

[1] Friedman, J., Hastie, T. Hoefling, H. & Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.*, **2**, 302–332.

[2] Simila, T. & Tikka, J. (2007). Input selection and shrinkage in multiresponse linear regression. *Comput. Stat. Data An.*, **52**, 406–422.

[3] Xin, X., Hu, J. & Liu, L. (2017). On the oracle property of a generalized adaptive elastic-net for multivariate linear regression with a diverging number of parameters, *J. Multivariate Anal.*, **162**, 16–32.

[4] Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. Roy. Stat. Soc.* ***B***, **68**, 49–67.