相関の高い変数の存在下における$\ell 1$正則化による変数選択について

塘　由惟
生物統計学分野 博士課程1年

2020年5月27日

背景

回帰モデルを推定する際に，回帰係数パラメータの$\ell 1$ノルムに基づく正則化項を加えた損失関数を最小化する手法はLasso（Least Absolute Shrinkage and Selection Operator）と呼ばれている[1]．Lassoでは，正則化項によりパラメータ空間に制約が与えられることから，推定の安定化が期待される．また，Lassoはいくつかの回帰係数が0へと推定されやすくなる性質を有することから，変数選択にも使用することができる．生物統計学の分野でも，SNP（一塩基多型）データの解析や，臨床予測モデルの構築など，Lassoの適用が検討される場面は多い．

しかし，Lassoでは，相関が高い変数が存在する場合に推定が不安定になることがある．この問題に対処するため，推定時の手順や罰則項に様々な工夫がなされ，手法が拡張されてきた．

目的

Lassoの性質に関する理解を深める．また，相関の高い変数の存在に対処するためのLassoの拡張手法について理解を深める．

方法

本抄読会の前半では，Lassoの概要や数理的な性質を概観する．後半では，相関が高い変数が存在する場合への対処を目的としたLassoの拡張手法を紹介し，特に，解釈性の高いモデル推定結果を得るための手法として最近提案されたIILasso(Independently Interpretable Lasso)について文献紹介を行う[2]．

文献

[1] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288, 1996.

[2] Masaaki Takada, Taiji Suzuki, and Hironori Fujisawa. Independently interpretable lasso: A new regularizer for sparse regression with uncorrelated variables. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, Vol. 84 of *Proceedings of Machine Learning Research*, pp. 454–463, 2018.