

L1 正則化を用いた医療施設間パフォーマンス評価

慶應義塾大学大学院 理工学研究科 鳴原 成美
慶應義塾大学 理工学部 鈴木 秀男
慶應義塾大学 医学部 宮田 裕章
東京大学大学院 医学系研究科 一原 直昭

1. はじめに

患者視点での医療の質を向上させていく上で、医療システムの機能を評価し改善すべく、施設間の医療の質のばらつきを評価し、それに対処することは重要である。従来、ロジスティック回帰や階層ベイズモデルによるリスク評価（例えば、[1]）が行われてきたが、連続変数のカテゴリ化や変数選択等については解析者の判断による部分が大きく、統計手法を用いる余地があると考えられる。本研究では、医学的な事前情報を与えた LASSO により、効率的かつ解釈可能な変数選択を行い、サンプル数の多寡に関係なく精度の安定したモデルを検討する。

2. 事前情報を考慮した LASSO (pLASSO) について

患者 i に対する観測データ (\mathbf{X}_i, Y_i) に対し、pLASSO[2]を考える。ここで、 Y_i は術後 30 日状態または退院時転帰が死亡の場合 1、共に Alive の場合 0 の二値変数であり、確率 $p(\mathbf{X}_i) = P(Y_i = 1 | \mathbf{X}_i)$ のベルヌーイ分布に従うと仮定する。事前情報は、 $\hat{\mathbf{Y}}^p = (\hat{Y}_1^p, \dots, \hat{Y}_n^p)^T$ と要約できると仮定する。大きさ $p + 1$ のベクトル $\mathbf{Z}_i = (1, \mathbf{X}_i)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ を用いると、確率 $p(\mathbf{X}_i)$ を次のように表現できる。

$$p(\mathbf{X}_i) = \frac{1}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{Z}_i)}$$

さらに、pLASSO の損失関数 $\ell(\boldsymbol{\beta}; \mathbf{X}, \mathbf{Y}) = -\frac{1}{n} \sum_{i=1}^n [Y_i \ln(p(\mathbf{X}_i)) + (1 - Y_i) \ln(1 - p(\mathbf{X}_i))]$ に対し、pLASSO の基準関数を次のように定義できる。

$$\begin{aligned} L_{\lambda, \eta}(\boldsymbol{\beta}; \mathbf{X}, \mathbf{Y}, \hat{\mathbf{Y}}^p) &= -\ell(\boldsymbol{\beta}; \mathbf{X}, \mathbf{Y}) - \eta \ell(\boldsymbol{\beta}; \mathbf{X}, \hat{\mathbf{Y}}^p) + \lambda \sum_{j=1}^p |\beta_j| \\ &\propto -\frac{1}{n} \sum_{i=1}^n [\tilde{Y}_i \ln(p(\mathbf{X}_i)) + (1 - \tilde{Y}_i) \ln(1 - p(\mathbf{X}_i))] + \lambda \sum_{j=1}^p |\beta_j| \end{aligned}$$

ここで $\tilde{Y}_i = (Y_i + \eta \hat{Y}_i^p) / (1 + \eta)$ 、 λ, η はチューニングパラメータである。この基準関数を最小化するときのパラメータの推定値 $\hat{\boldsymbol{\beta}}$ を LASSO と同様のアルゴリズムで解くことが可能である。本研究では、LARS アルゴリズムにより $\hat{\boldsymbol{\beta}}$ を算出し、モデルの有効性を検証する。

参考文献

- [1] O'Brien, S. M., Cohen, D. J., Rumsfeld, J. S., Brennan, J. M., Shahian, D. M., Dai, D., & Edwards, F. H. (2016). Variation in hospital risk-adjusted mortality rates following transcatheter aortic valve replacement in the United States. *Circulation: Cardiovascular Quality and Outcomes*, CIRCOUTCOMES-116.
- [2] Jiang, Y., He, Y., & Zhang, H. (2016). Variable selection with prior information for generalized linear models via the prior Lasso method. *Journal of the American Statistical Association*, 111(513), 355-376.