数理解析研究所講究録
第2018巻 2017年 114-115

spike and slab 事前分布を用いたスパース推定

大阪大学大学院基礎工学研究科 田辺 竜ノ介
Tanabe Ryunosuke
Graduate School of Engineering Science, Osaka University

1 導入

多変量線形回帰モデルでは回帰係数パラメータの推定が主な問題である．観測を $Y=(Y_1,\ldots,Y_n)^T$，回帰係数を $\beta=(\beta_1,\ldots,\beta_d)$，説明変数 X を $n\times d$ の行列とする．このとき $Y\sim N(X\beta,\sigma^2 I_n)$ に従っているとする．σ^2 は分散パラメータ，I_n は長さ n の単位行列である．

観測に対する説明変数が多いほど推定の精度が上昇するがデータを取りなおしたときは推定の精度は大きく下がってしまう．これは実際には不要な説明変数を用いていることに起因している．そのために，不要な説明変数を使用しないことが重要になってくる．このときに用いられる手法の一つにスパース推定がある．

スパース推定は回帰係数の推定とモデル選択を同時に行える手法である．回帰係数の推定で不要なパラメータを0と推定することで，説明変数の必要か不要かの判断が可能になる．Thibishirani(1996) では罰則項を用いることで Lasso 推定を構成した．Lasso 推定量は

$$\hat{\beta}^{Lasso}=\arg\min\left\{||Y-X\beta||_2^2+\lambda||\beta||_1\right\}$$

で記される．ただし $||\cdot||_0$ は L_1 ノルムである．$||\cdot||_2$ は L_2 ノルムである．これによりスパース推定が可能となる．一方罰則項を加える手法は複数あり，その一つが Ridge 推定である．Ridge 推定量は

$$\hat{\beta}^{Ridge}=\arg\min\left\{||Y-X\beta||_2^2+\lambda||\beta||_2^2\right\}$$

で表わされる．Ridge 推定量はスパース性を持っていない．

スパース推定で求められる性質の一つに Fan and Li(2001) が提唱した oracle 性がある．それは変数選択の一致性と，推定量の漸近正規性を持つという性質である．Lasso はその oracle 性を持たないことが知られている．

頻度論におけるスパース推定の欠点の一つに信頼区間の構成が困難という点がある．これは推定量の構成が複雑な点が起因している．そのためブートストラップなどの手法を用いる必要がある．その問題点を解決する方法に Bayesian Lasso がある．Bayesian Lasso は Lasso がベイズ的に解釈可能という事実を元に構成されている．回帰係数 β の事前分布にラプラス分布を置くことで最頻値が Lasso 推定量と一致する．Park and Casella(2008) では Laplace 分布が正規分布の分散混合分布と解釈できることを用いて GIbbs sample を構成した．この手法はベイズ的手法にもとづいているので MCMC により信用区間が容易である．しかし，Bayesian Lasso はスパース性をもっていないことが知られている．そのため，信用区間に0が入っていれば0と判断するといった処理が必要でありスパース推定の利点が失われている．

spike and slab prior は一点分布と連続分布の混合事前分布である．spike and slab prior は

$$\pi(\beta_j)=(1-p)\delta_0(\beta_j)+p\pi_0(\beta_j)$$

という形でかかれる．$\pi_0(\beta_j)$ はなんらかの連続分布で，正規分布や Laplace 分布を用いる．q は混合比率パラメータ，$\delta_0(\beta_j)$ は確率1で0を取る一点分布である．

本研究では従来であれば事前分布を $\pi(\beta_j)=\pi_0(\beta_j)$ と置いていたところを，その代わりに spike and slab prior を用いて解析を行い，$X^TX=nI_d$ の条件下での oracle 性の導出を行った．

2 spike and slab Ridge

この章ではRidge推定量に対応するスパースRidgeについて考える．スパースRidgeモデルを以下のように定める

$$
\begin{aligned}
Y &\sim N(X\beta, \sigma^2 I_n) \\
\pi(\beta_j) &= (1-q)\delta_0(\beta_j) + qN(\beta_j|0, \sigma^2/\lambda), \quad j = 1, \dots, d
\end{aligned}
$$

分散 σ^2 と混合比率 q が既知の場合はそこに事前分布は置かないとする．このとき以下の事実が成立する

定理 2.1. 分散 σ^2 と混合比率 q が既知とする．また X は直交行列，すなわち $X^T X = nI_d$ とする．このときスパースRidgeモデルの事後中央値 $Med(\beta_j|Y, \sigma^2, q)$ はoracle性を持つ．

この事実は事後中央値が

$$
\hat{\beta}_j^{Med} = \begin{cases} 0 & |\hat{\beta}^{OLS}| \leq t \\ \mathrm{sign}(\hat{\beta}_j^{OLS}) \left[\frac{n}{\lambda+n} |\hat{\beta}_j^{OLS}| + (\sigma^2/n)^{1/2} \sqrt{\frac{n}{\lambda+n}} \Phi^{-1} \left(\frac{1-\min(1, \omega(\hat{\beta}_j^{OLS}))}{2} \right) \right] & |\hat{\beta}^{OLS}| > t \end{cases}
$$

$$
\omega(\hat{\beta}_j^{OLS}) = \frac{1-q}{q} \frac{\lambda+n}{\lambda} \exp\left(-\frac{n}{\lambda+n} \frac{(\hat{\beta}_j^{OLS})^2}{\sigma^2/n} \right)
$$

であることから導かれる．ただし t は

$$
\frac{q}{1-q} \sqrt{\frac{\lambda}{\lambda+n}} \left[2\Phi\left(\frac{t}{\sigma/\sqrt{n}} \sqrt{\frac{\lambda}{n+\lambda}} \right) - 1 \right] = \exp\left(-\frac{n}{n+\lambda} \frac{t^2}{2\sigma^2/n} \right)
$$

である．この手法はRidge推定にもとづいているため，β_j^{Med} が0でないときの振る舞いはRidgeに似ている．一方スパース性は保持しているためスパースなRidge推定量と解釈することが出来る．またベイズ的手法に基いているため信用区間の構成が容易である．

3 結論

spike and slab priorを用いることで従来ではスパース性やoracle性をもたなかったRidge推定に，それら性質を持たすことが可能になった．しかし，今回導出した性質は $X^T X = nI_d$ の仮定をしたままで行われているため，説明変数行列の一般化を行う必要がある．加えて $\pi_0(\beta_j)$ の部分に今回は正規分布を当てはめたが，他の分布も当てはめてどの分布が一番よいかを調査する必要がある．

参考文献

[1] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.

[2] Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.