

一般化正規分布を用いたフルベイズアン罰則回帰モデルの提案

東京大学大学院教育学研究科 / 日本学術振興会 北條大樹

東京大学大学院教育学研究科 岡田謙介

1 はじめに

変数選択は統計学の実践上非常に重要な問題である。これらは、主にスパース推定の文脈で議論されることも多いが、近年、ベイズアン罰則回帰の文脈においても多くの議論がされている。これまでベイズアン罰則回帰では、回帰係数の事前分布に spike and slab 分布や二重指数分布を置く方法(Bayesian lasso)が提案されている。ハイパーパラメータの設定法としては、通常の変差検証法に加え、経験ベイズ法、フルベイズ法が存在する(van Erp, Oberski, & Mulder, 2019)。

本研究では、既存の Bayesian lasso や spike and slab 事前分布の考え方を拡張し、一般化正規分布を事前分布に用いた新たなフルベイズアン罰則回帰モデルを提案する。

2 一般化正規分布を用いたフルベイズアン罰則回帰モデル

一般化正規分布は正規分布に形状パラメータ κ を加えた確率分布であり、その確率密度関数は、

$$f(x; \mu, \sigma, \kappa) = \frac{\kappa}{2\sigma\Gamma(1/\kappa)} e^{-\left(\frac{|x-\mu|}{\sigma}\right)^\kappa} \quad (1)$$

と定義される。 $\kappa = 1$ のときに二重指数分布、 $\kappa = 2$ のときに通常の正規分布となり、また $\kappa \rightarrow +\infty$ のとき一様分布に収束する。本研究で提案するモデルは、回帰モデルの係数パラメータの事前分布に一般化正規分布を仮定し、形状パラメータの推定も行うモデルである。本モデルの特徴は、事前分布に正規分布をおく場合の Ridge 回帰、二重指数分布をおく場合の Bayesian lasso 回帰、無情報事前分布として一様分布をおく場合の回帰モデルを含む統合的な一般化が可能になる点である。

3 方法と結果

本研究では、提案モデルと他のフルベイズアン罰則回帰モデルをシミュレーション実験によって比較し、そして、提案モデルによる実データ分析を行った。シミュレーション実験で比較対象としたモデル及び実験条件は、van Erp et al. (2019)を参考にした。各モデルパラメータのベイズ推定は、ハミルトニアンモンテカルロ法による NUTS アルゴリズムを用いた。変数間の相関を仮定しない単純な回帰モデルからデータを生成した条件では、回帰係数の事後モード値を用いた平均二乗誤差 (MSE) の観点から既存のモデルと大きな相違はなかった。

4 参考文献

[1] van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, Vol. 89, pp. 31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>