Global-Local 縮小事前分布に基づく Bayesian Convex Clustering

NTT アドバンステクノロジ株式会社, 電気通信大学　嶋村 海人
電気通信大学　川野 秀一

1. はじめに

Sparse convex clustering (Wang *et al.* 2018) は, 凸最適化問題によるクラスタ推定と変数選択を同時に行うクラスタリング手法である. Sparse convex clustering の正則化項には, 個体と特徴量ごとに重みづけされた L_q ノルムが利用されるが, この正則化項によるモデリングはデータへの依存が高くサンプルサイズが十分でない場合には推定精度が低下することが指摘されている.

本報告では, この問題を解決するために sparse convex clustering をベイズモデリングの枠組みに拡張し, global-local 縮小事前分布 (Polson and Scott 2010) によるモデリング手法を提案する. また, 数値実験を通して, 提案手法の有効性を検証する.

2. 提案手法

$X \in \mathbb{R}^{n\times p}$ を p 次元 n サンプルのデータ行列とし, 各個体に対応する行ベクトルを $\boldsymbol{x}_i$ $(i=1,\cdots,n)$ とする. $n\times p$ 特徴行列 $A=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)^T$ に対して, 次のように尤度関数および grobal–local 縮小事前分布 $\mathrm{GL}(\cdot|\nu,\tau)$ を仮定する.

$$
\begin{aligned}
f(X|A,\sigma^2) &= \prod_{i=1}^{n}(2\pi\sigma^2)^{-p/2}\exp\left\{-\frac{\|\boldsymbol{x}_i-\boldsymbol{a}_i\|_2^2}{2\sigma^2}\right\},\\
\pi(A|\sigma^2) &\propto (\sigma^2)^{-(\#\mathcal{E}+p)/2}\mathrm{GL}\left(\frac{1}{2\sigma}\mathbf{a}\;\middle|\;\nu,(\tau_{i_1,i_2};\;(i_1,i_2)\in\mathcal{E})\right)\\
&\quad\times\mathrm{GL}\left(\frac{1}{2\sigma}\widetilde{\mathbf{a}}\;\middle|\;\widetilde{\nu},(\widetilde{\tau}_j;\;j=1,\cdots,p)\right).
\end{aligned}
$$

ただし, $\mathbf{a}=(\|\boldsymbol{a}_{i_1}-\boldsymbol{a}_{i_2}\|_2;\;(i_1,i_2)\in\mathcal{E})$, $\widetilde{\mathbf{a}}=(\|\boldsymbol{a}_{\cdot j}\|_2;\;j=1,\cdots,p)$, σ^2 はサンプルの分散パラメータ, $\nu\ (>0)$, $\widetilde{\nu}\ (>0)$, $\tau_{i_1,i_2}\ (>0)$, $\widetilde{\tau}_j\ (>0)$ はハイパーパラメータ, $\mathcal{E}$ は同一のクラスタとなる可能性のあるサンプルの添字のペアを要素にもつ集合とする. ここで, 推定値 $\hat{\boldsymbol{a}}_i$ に対し, $\hat{\boldsymbol{a}}_{i_1}=\hat{\boldsymbol{a}}_{i_2}$ となった場合は, 第 i_1 番目と第 i_2 番目のサンプルは同一のクラスタに属するとする. また, ν および τ_{i_1,i_2} は推定されるクラスター数をコントロールし, $\widetilde{\nu}$ および $\widetilde{\tau}_j$ は 0 と推定する特徴量の数をコントロールする.

この事前分布のハイパーパラメータに対し, 事前分布を仮定することでモデルの階層表現が可能である. それにより, MCMC 手法の一つである Gibbs sampling を行うことができる. ベイズスパースモデリングでは, モデルの点推定を行う際に事後平均を利用することが多いが, 本報告では推定精度を向上させるために事後確率に基づく重み付き事後平均の利用を提案する. 推定アルゴリズムならびに数値実験の結果は当日紹介する.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Stat.*, **9**, 501–538.

Wang, B., Zhang, Y., Sun, W. W., and Fang, Y. (2018). Sparse convex clustering. *J. Comput. Graph. Stat.*, **27**, 393–403.