

クラスタリングを伴うテンソルデータに対する ガウシアングラフィカルモデル

大阪大学大学院人間科学研究科 宇野光平

1 ガウシアングラフィカルモデルとその発展

ガウシアングラフィカルモデルは Graphical Lasso(Friedman et al., 2008)などが提案されて以降研究が盛んに行われている。たとえば行列正規分布やテンソル正規分布を用いることによりテンソルデータへの適用も可能となった(He et al., 2014; Yin and Li, 2012)。さらには複数群からなるガウシアングラフィカルモデルを推定するために、Joint Graphical Lasso (Danaher et al, 2014)が提案され、そのグループが既知ではなく未知の場合でも推定できる手法も提案されている(Gao, et al., 2016)。本研究では、テンソルデータに対し未知の複数グループからなるグラフィカルモデルを推定するための新しい手法を提案する。

2 モデル

$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ をテンソル正規分布とする。ここで $\mathbf{x} \in \mathbb{R}^P, \boldsymbol{\mu} \in \mathbb{R}^P, \boldsymbol{\Sigma} \in \mathbb{R}^{P \times P}, \boldsymbol{\Sigma}^{-1} = \mathbf{W}$ とし、テンソルのオーダーを K 、各オーダーにおける変数の数を p_k かつ $P = \prod_{k=1}^K p_k$ とする。このとき、 $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_{K-1} \dots \otimes \boldsymbol{\Sigma}_1$ とオーダー毎に完全に分離できる。提案手法は観測された n 個のテンソルデータを $G(G < n)$ 個の混合テンソル正規分布によって分類し、各クラスターのグラフィカルモデルを推定することを目的とする手法である。提案手法の対数尤度は以下である。

$$\log L(\boldsymbol{\Theta}) = \sum_{n=1}^N \log \left\{ \sum_{g=1}^G \pi_g f_g \left(\mathbf{x}_n; \boldsymbol{\mu}_g, (\mathbf{W}_g)^{-1} \right) \right\} - P(\boldsymbol{\Theta}) . \quad (1)$$

ただし $\sum_{g=1}^G \pi_g = 1$ 。ペナルティ関数 $P(\boldsymbol{\Theta})$ は以下とする。

$$P(\boldsymbol{\Theta}) = \lambda_1 \sum_{g=1}^G \sum_{p=1}^P |\mu_{gp}| + \sum_{k=1}^K \lambda_{2,k} \sum_{g=1}^G \sum_{h \neq g} |w_{ghp,k}| + \sum_{k=1}^K \lambda_{3,k} \sum_{h \neq g} \left(\sum_{g=1}^G w_{ghp,k}^2 \right)^{\frac{1}{2}} . \quad (2)$$

パラメータの更新式と実データ解析への適用結果は当日報告する。

参考文献

- Danaher, P., Wang, P. and Witten D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society, Series B*, **76**, 373-397.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432-441.
- Gao, C., Zhu, Y., Shen, X. and Pan, W. (2016). Estimation of multiple networks in Gaussian mixture models. *Electronic Journal of Statistics*, **10**, 1133-1154.
- He, S., Yin, J., Li, H. and Wang, X. (2014). Graphical model selection and estimation for high dimensional tensor data. *Journal of Multivariate Analysis*, **128**, 165-185.
- Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis*, **107**, 119-140.