

セルサイズ外れ値に頑健なスパースグラフィカルモデリング

片山翔太

東京工業大学工学院

概要

条件付き独立性の有無を辺の有無に対応させるグラフィカルモデリングは、可視化によって変数間の関連性を捉えることができ、新しい知見の獲得につながる。特にデータに正規性を仮定した場合は、Lasso(Tibshirani, 1996)をはじめとしたスパースモデリングとの相性が良く、大規模データに対してもグラフの探索を容易に行うことができるようになった。しかしながらその一方で、データの大規模化に伴う新たな問題も生じている。例えばデータ行列に対するセルサイズな外れ値の混入である。これは特に遺伝情報やファイナンスに関するデータにおいて頻繁に観測される。セルサイズ外れ値が混入する状況では、観測ベクトルの全てを除外もしくはその重みを小さくする従来型のロバスト推定は、データの情報損失が著しく適切なものであるとは言い難い。本報告では、スパースモデリングを伴う種々のグラフ推定法が、共分散行列の推定量から構成されていることに着目し、それをセルサイズ外れ値に対して頑健なものに置き換えることで問題を解決する。なお、時間が許せば、提案手法の数値的な評価を行った後に、その理論面に関しても考察を行う。