

主成分分析法の比較に関する研究

曾我部 泰誠

スパース主成分分析は主成分分析の発展手法であり、主成分分析で推定される行列の結果のいくつかを 0 で表示することにより解釈しやすいものとする手法である。しかしスパース主成分分析には複数の種類が存在しているが、主成分分析と異なり、手法により解が異なる。しかし手法の解の比較に関する研究はあまり行われていないため、本研究ではスパース主成分分析の代表的な手法として Adachi & Trendafilov (2016)の USLPCA と Zou, Hastie & Tibshirani (2006)の elastic net を用いる sparse principal component analysis(SPCA)を比較した。

主成分分析の使用目的を2つに分け、それぞれの目的に適合するデータを用いて各手法で実行することで、各手法の効果を比較する。使用目的の 1 つはデータの構造を理解することであり、学力データを用いた。出力される結果の成分負荷量 \mathbf{A} と重み行列 \mathbf{W} 、そして寄与率を記し、そこからどのようなデータの解釈ができるかを記した。もう1つの使用目的は次元の縮約をすることであり、高次元の遺伝子データを用いた。次元縮約が正しく行われているのかを、事前情報から得た異なるフェイズに対して、異なる特徴が表れているのかをヒートマップを作成することで確認し、また寄与率を求めた。その結果学力データでは USLPCA 66.3%、SPCA 64.6%となり寄与率に大きな違いは生じなかったことに対し、遺伝子データでは USLPCA 73.4%、SPCA 61.2%となり寄与率に差が生じた。しかしどちらの手法も寄与率は 60%を超えている。そこでどちらの手法を使用するのかについては 1 つ目の目的においては主成分を表すための変数の係数を表す \mathbf{W} よりも変数と主成分の相関関係を表す \mathbf{A} が、2つ目の目的である後の利用のためには主成分がどのようなものを表す \mathbf{W} を利用する手法のほうが、それぞれの主成分分析の目的に適合した推定値の結果がスパースで得られるため良いと考えられる。(行動統計科学)