1 段階主成分回帰モデルとその応用

電気通信大学 大学院情報理工学研究科　川野 秀一

1. はじめに

主成分回帰とは，主成分分析を用いて説明変数を次元圧縮し，圧縮された変数を新たな説明変数とみなして回帰モデルを構成する方法である．計量化学の分野においては，部分的最小二乗回帰と並んでよく用いられる手法の一つである．これまで報告者は，2 段階法で推定される従来の主成分回帰の問題点を指摘し，1 段階法に基づく主成分回帰モデルを提案している．本報告では，1 段階主成分回帰モデルとその実データへの適用例を紹介するとともに，先とは異なる定式化に基づく 1 段階主成分回帰モデルについても紹介する．

2. 1 段階主成分回帰モデル

いま，目的変数 $Y\ (\in \mathbb{R})$ と p 次元説明変数 $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ に関して，n 組のデータ $\{(y_i, \boldsymbol{x}_i); i = 1, \ldots, n\}$ が得られたとする．説明変数に関するデータは中心化されているとする．このとき，Kawano *et al.* (2015) は条件付き最小化問題

$$\min_{A,B,\gamma_0,\boldsymbol{\gamma}} \left\{ \|\boldsymbol{y} - \gamma_0 \mathbf{1}_n - XB\boldsymbol{\gamma}\|_2^2 + w \sum_{i=1}^{n} \left\|\boldsymbol{x}_i - AB^T \boldsymbol{x}_i\right\|_2^2 + \lambda_\beta \|B\|_1 + \lambda_\gamma \|\boldsymbol{\gamma}\|_1 \right\} \tag{1}$$
$$\text{subject to} \quad A^T A = I_k$$

から得られるモデルをスパース主成分回帰モデルと呼んだ．ここで，$\boldsymbol{y} = (y_1, \ldots, y_n)^T$，$X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ は $n \times p$ 型計画行列，$\mathbf{1}_n$ はすべての要素が 1 の n 次元ベクトル，I_k は $k \times k$ 型単位行列，γ_0 は切片パラメータ，$\boldsymbol{\gamma}$ は $k\ (< p)$ 次元パラメータベクトル，A および B は $p \times k$ 型負荷行列である．また，$\lambda_\beta, \lambda_\gamma$ は正の値を取る正則化パラメータ，w は正の値を取る調整パラメータである．第 1 項目が回帰モデルの損失関数，第 2 項目が主成分分析モデルの損失関数であることに注意すると，これは 1 段階法による主成分回帰と捉えることができる．また，第 3, 4 項目は，それぞれパラメータ $B, \boldsymbol{\gamma}$ に対するスパース正則化項である．(1) では目的変数の値が実数を取ることを想定しているが，Kawano *et al.* (2018) は目的変数の値が 2 値や離散値の場合でも解析可能なスパース主成分回帰モデルを提案している．

一方，次の条件付き最小化問題

$$\min_{B,\gamma_0,\boldsymbol{\gamma},Z} \left\{ \|\boldsymbol{y} - \gamma_0 \mathbf{1}_n - XB\boldsymbol{\gamma}\|_2^2 + w\|X - ZB^T\|_F^2 + \lambda_\beta \|B\|_1 + \lambda_\gamma \|\boldsymbol{\gamma}\|_1 \right\} \tag{2}$$
$$\text{subject to} \quad B^T B = I_k$$

を考える．ここで，Z は $n \times k$ 型の主成分スコアからなる行列であり，$\|\cdot\|_F$ は行列のフロベニウスノルムを表す．最小化問題 (1) と (2) の主な違いは第 2 項目にあり，(2) では特異値分解による主成分分析モデルの定式化 (Shen and Huang, 2008) を用いている．したがって，最小化問題 (2) は (1) とは異なる定式化による 1 段階主成分回帰モデルである．

3. ソフトウェア

1 段階主成分回帰モデルは，統計ソフトウェア R のパッケージ `spcr` として利用可能である．

参考文献

Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T. (2015) Sparse principal component regression with adaptive loading. *Comput. Stat. Data Anal.*, **89**, 192–203.

Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T. (2018) Sparse principal component regression for generalized linear models. *Comput. Stat. Data Anal.*, **124**, 180–196.

Shen, H. and Huang, J. Z. (2008) Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.*, **99**, 1015–1034.