適応的な主成分スコアに基づくスパース回帰

藤澤　洋徳　数理・推論研究系　教授

はじめに

説明変数の数が多い回帰モデルを考えるとき，説明変数に対して主成分分析をまずは行い，固有値が大きい幾つかの主成分スコアを新しい説明変数と見直して回帰を行う，という主成分回帰は良く行われている．しかし，主成分分析で新しい説明変数をいくつか選んだとき，回帰の意味で良いものを選んでいるとは限らない．これは，目的が違う手法を組み合わせて2段階で行っているからである．

そこで，我々は，主成分分析と回帰モデルを，同時に1段階で行うことを考えた．そうすることで，回帰の意味で有望な主成分スコアを優先的に選ぶことを目指した．しかし普通はそれだけではうまく行かない．なぜなら，明らかにパラメータが冗長になるので，パラメータが推定不能になるからである．そこで，通常の二乗誤差に，主成分分析で使われるロスとスパース正則化項を罰則として取り入れて，パラメータ推定を可能にすることを試みた．提案したロス関数の最小化に対応するパラメータ推定アルゴリズムも構築した．簡単な数値実験で，その方法の良さも確認した．

主成分回帰（二段階法）

回帰

説明変数：$\boldsymbol{x} = (x_1, \ldots, x_p)^T$　（pが大きい）
目的変数：y

標本

$$(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$$

次元縮約

主成分ベクトル：$B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)$　（qは大きくない）
主成分スコア：$z_j = \boldsymbol{\beta}_j^T \boldsymbol{x}$

主成分回帰モデル

$$y = \sum_{j=1}^{q} \gamma_j z_j + e$$

パラメータ$\boldsymbol{\gamma}$は最小二乗法などで推定する．

主成分回帰（一段階法）

慣習的な主成分回帰の欠点

固有値の大きい方から主成分スコアを取り入れるため，固有値が小さい主成分スコアが回帰モデルの本質であるとき，対応できない．

二段階から一段階へ！

回帰誤差

$$L_{\mathrm{reg}}(\boldsymbol{\gamma}, B) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{q} \gamma_j \boldsymbol{\beta}_j^T \boldsymbol{x}_j \right\}^2$$

この誤差の単純な最小化は使えない．明らかにパラメータが冗長である．

主成分ロス Zou (2006)

$$L_{\mathrm{PCA}}(B, A) = \sum_{i=1}^{n} \|\boldsymbol{x}_i - AB^T \boldsymbol{x}_i\|^2 \qquad A^T A = I_q$$

最適化アルゴリズムを構築するときにスパース罰則との相性が良い．

スパース罰則を加えたロス関数

$$\begin{aligned} L(\boldsymbol{\gamma}, B, A) &= (1-w) L_{\mathrm{reg}}(\boldsymbol{\gamma}, B) + w L_{\mathrm{PCA}}(B, A) \\ &\quad + \lambda \xi \sum_{j=1}^{q} \|\boldsymbol{\beta}_j\|_1 + \lambda (1-\xi) \sum_{j=1}^{q} \|\boldsymbol{\beta}_j\|^2 + \lambda_\gamma |\boldsymbol{\gamma}|_1 \end{aligned}$$

$$(\hat{\boldsymbol{\gamma}}, \hat{B}, \hat{A}) = \arg\min L(\boldsymbol{\gamma}, B, A)$$

上記のバージョンを Sparse Principal Component Regression (SPCR) と呼び，推定値をよりスパースにするために adaptive lasso のアイデアを組み入れたものを aSPCR と呼ぶことにした．

スパース罰則でパラメータ推定値が0になりやすくする．

スパース罰則がないと，主成分スコアには，回転の不変性が残っている．スパース罰則によって，回転の不変性がなくなることが期待できる．

パラメータ推定アルゴリズム

それぞれのパラメータごとに更新する．(他のパラメータを与えた下で．)
そうすると，それぞれは，二次計画問題になる．
アルゴリズムを陽に書ける．

チューニングパラメータの選択

wとξは事前に与える．
その他のチューニングパラメータはCVで自動選択する．

w：小さすぎるとパラメータ冗長の問題に絡むのか収束が遅すぎる．
ξ: Elastic Net と同じ考えを使う．

数値実験

モデル

$$y_i = 4 \boldsymbol{x}_i^T \boldsymbol{\xi}^* + \varepsilon_i \quad \varepsilon_i \sim N(0,1) \quad i = 1, \ldots, n$$
$$\boldsymbol{x} = (x_1, \ldots, x_{20})^T \quad \sim \quad N(\mathbf{0}_{20}, \Sigma_2)$$
$$\Sigma_2 = \mathrm{blockdiag}(\Sigma_2^*, I_{11}) \qquad (\Sigma_2^*)_{ij} = 0.9^{|i-j|}$$

$$\boldsymbol{\xi}^* = (\boldsymbol{\nu}_1^*, 0, \ldots, 0)^T$$
$$\boldsymbol{\nu}_1^* = (-1, 0, 1, 1, 0, -1, -1, 0, 1)^T$$

Σ_2^*の第4固有ベクトルのスパース近似

Mean Squared Error

q	n	aSPCR	SPCR	SPLS	PLS	PCR
1	50	**1.284**	1.583	2.079×10^1	2.084×10^1	2.140×10^1
	200	**1.058**	1.120	1.568×10^1	1.695×10^1	2.086×10^1
5	50	**1.279**	1.576	2.017	3.398	2.224×10^1
	200	**1.060**	1.119	1.075	1.175	2.097×10^1

PLS: Partial Least Squares (Regression)

文献

Shuichi Kawano, Hironori Fujisawa, Toyoyuki Takada, Toshihiko Shiroishi (2014). Sparse principal component regression with adaptive loading. http://arxiv.org/abs/1402.6455

R package: spcr