スパース・モデリングの研究

藤澤　洋徳　数理・推論研究系　教授

Lasso　スパース・モデリングの基本

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i \right)^2 + \lambda \|\boldsymbol{\beta}\|_1 \qquad \|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$$

推定値の例： $\hat{\boldsymbol{\beta}} = (0, 0, 1.3, 2, 0, \ldots, 0)$
正確に0が現れる．スパースな推定．
パラメータ推定と変数選択が同時に行われている．

重要なこと

上記の推定値を得るための数値アルゴリズム （座標降下法）
一致性（や漸近正規性など）の理論保証
チューニングパラメータの自動選択（交差確認法でだいたい大丈夫）

スパース主成分回帰

R package: spcr

通常の主成分回帰の問題点

通常の主成分回帰は，説明変数に主成分分析を行い，固有値の大きいものに対応する幾つかの主成分スコアを新しい説明変数として取り入れるため，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．

回帰誤差　$\boldsymbol{\beta}_j^T \boldsymbol{x}_j$：主成分スコア．　下記だけではパラメータが冗長．

$$L_{\mathrm{reg}}(\boldsymbol{\gamma}, B) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{q} \gamma_j \boldsymbol{\beta}_j^T \boldsymbol{x}_j \right\}^2 \qquad B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q)$$

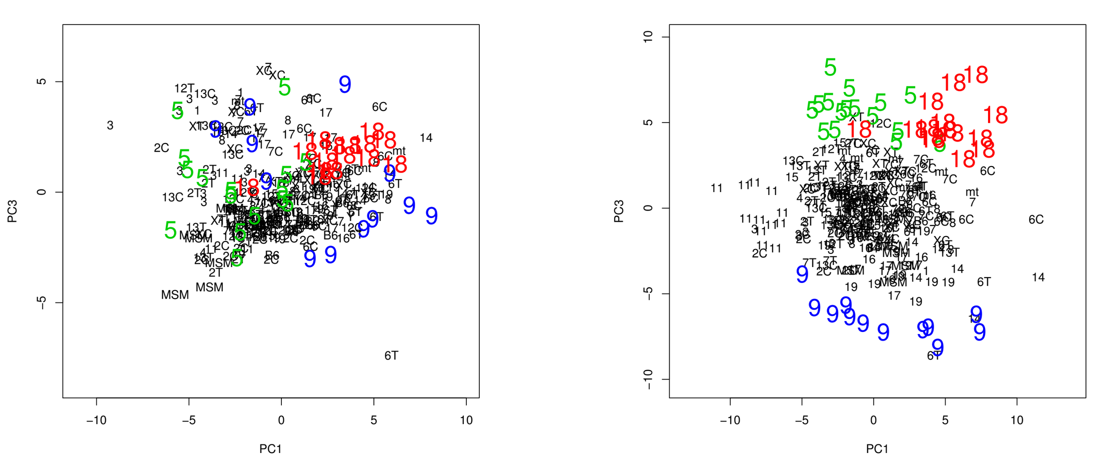
主成分ロス　$L_{\mathrm{PCA}}(B, A) = \sum_{i=1}^{n} \|\boldsymbol{x}_i - AB^T \boldsymbol{x}_i\|^2 \quad A^T A = I_q$

スパース罰則を加えた推定

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

ポイント

スパース罰則によって，冗長パラメータの原因である回転の不変性がなくなることも期待できる．
座標降下法に基づいた数値アルゴリズムが提案できる．
一般化線形回帰モデルにも拡張できる．



コンソミック系統マウスに適用した結果．(左：PCA．右：提案手法)

相関情報を罰則に導入したスパース・モデリング

R package: iilasso
詳しくは社会人大学院生の高田正彬氏のポスターをご覧ください．

Lassoの問題点

Lassoの推定量にはバイアスが内在している．元の説明変数が非常に多い時には，そのバイアスは無視できない．すると，有効な説明変数がきちんと選ばれても，その係数のLasso推定量を利用した回帰モデルでは，推定量のバイアスの分だけ説明力が不足している．有効変数と相関が高い非有効変数が存在する場合，その非有効変数も無理やりに取り込んで，回帰モデルの説明力を回復しようとする．結果的に，不要な非有効変数が取り込まれた回帰モデルが導出される．

問題点の克服方法

相関の高い変数は同時に選ばれにくくする罰則を導入する．

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2n} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda \left(\|\boldsymbol{\beta}\|_1 + \frac{\alpha}{2} |\boldsymbol{\beta}|^\top R |\boldsymbol{\beta}| \right)$$

R_{jk}は相関行列の絶対値の単調増加関数．
R_{jk}が非常に大きいとき，$R_{jk}|\beta_j||\beta_k|$のために，どちらかのβを0にする．

ポイント

座標降下法に基づいた数値アルゴリズムを提案．
非漸近バウンドを導出．適当な場合にはLassoよりも良いと言える．
推定量のバイアスを減らす手法としてはSCADやMCPが有名なので，それらとも数値的に比較して，有効性を確認．

ロバストでスパースなグラフィカル・モデリング

R package: rsggm

glasso (graphical lasso)　$\Omega = \Sigma^{-1} = (\omega_{jk})$

$$\begin{aligned}\hat{\boldsymbol{\theta}} \;&=\; \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} -\log \phi(\boldsymbol{x}_i; \boldsymbol{\mu}, \Omega^{-1}) + \lambda \sum_{j<k} |\omega_{jk}| \right\} \\ &=\; \arg\min_{\boldsymbol{\theta}} \left\{ d_{\mathrm{KL}}(\bar{g}, \phi_{\boldsymbol{\theta}}) + \lambda \sum_{j<k} |\omega_{jk}| \right\}\end{aligned}$$

$d_{\mathrm{KL}}(g, f)$ ：密度関数gとfのKL相互エントロピー
$\bar{g}$ ：データ$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$から得られる経験密度関数

ロバスト化　γ-glasso

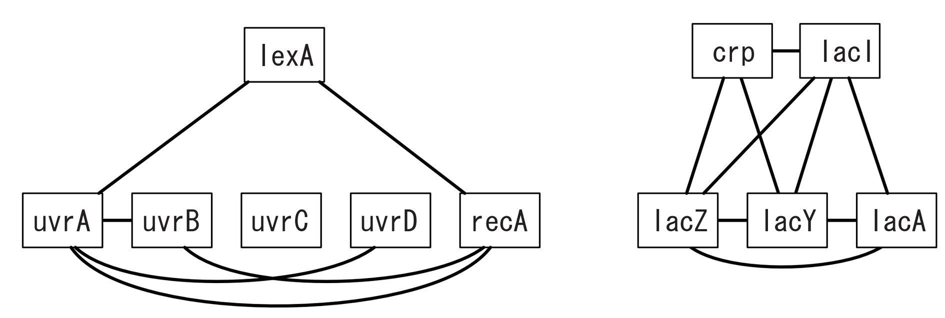
$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ d_\gamma(\bar{g}, \phi_{\boldsymbol{\theta}}) + \lambda \sum_{j<k} |\omega_{jk}| \right\}$$

γ-相互エントロピー

$$d_\gamma(g, f) = -\frac{1}{\gamma} \log \int g(x) f(x)^\gamma dx + \frac{1}{1+\gamma} \log \int f(x)^{1+\gamma} dx$$

パラメータ推定アルゴリズム： MMアルゴリズム ＋ glasso

遺伝子ネットワーク同定　γ-glasso



他のロバスト法はここまできれいにネットワーク化できなかった．
（2次元のPCAから10%程度以上の外れ値が想定された．）

そのほかの研究

ロバスト性とスパース性を併せもつ回帰モデリング
スパース・モデリングに関係した非凸最適化： 川島氏@大学院生

スパースと関係のない最近の研究： 欠測データ解析．異常検知．ロバスト統計．欠測と外れ値を同時にもつデータの解析．半教師付き学習．クラスタリング．非対称分布．ゲノムデータ解析．検定．

企業や応用分野との共同研究： 製造業・製薬・医学・遺伝学など．