Transfer Lasso

藤澤　洋徳　数理・推論研究系　教授（ものづくりデータ科学研究センター副センター長・理化学研究所AIP客員研究員・名古屋大学大学院医学系研究科客員教授）

はじめに

転移学習： 過去の情報をパラメータ推定値に縮約して現在に転移させ，その情報を有効活用して現在のパラメータを推定する手法を提案する．そこでLassoを上手く使う．

状況の変化に対応： 過去と現在の状況が同じであれば精度が上がり，過去と現在の状況が違っていれば過去の情報を上手く捨てることができる．

論文情報： arXiv:2006.14845

そのほかの情報： 東芝との共同研究．NeurIPS受理．

Lasso

$$\check{\beta} = \arg\min_{\beta} \frac{1}{2n}\sum_{i=1}^{n}(Y_i - X_i^\top \beta)^2 + \lambda\|\beta\|_1$$

L_1罰則に特徴がある．
パラメータ推定値が正確に0になりうる．
パラメータ推定と特徴量選択が同時に行える．
λは Cross Validation で選ぶ．
非線形にも基底関数展開を使えば対応可能．

Transfer Lasso

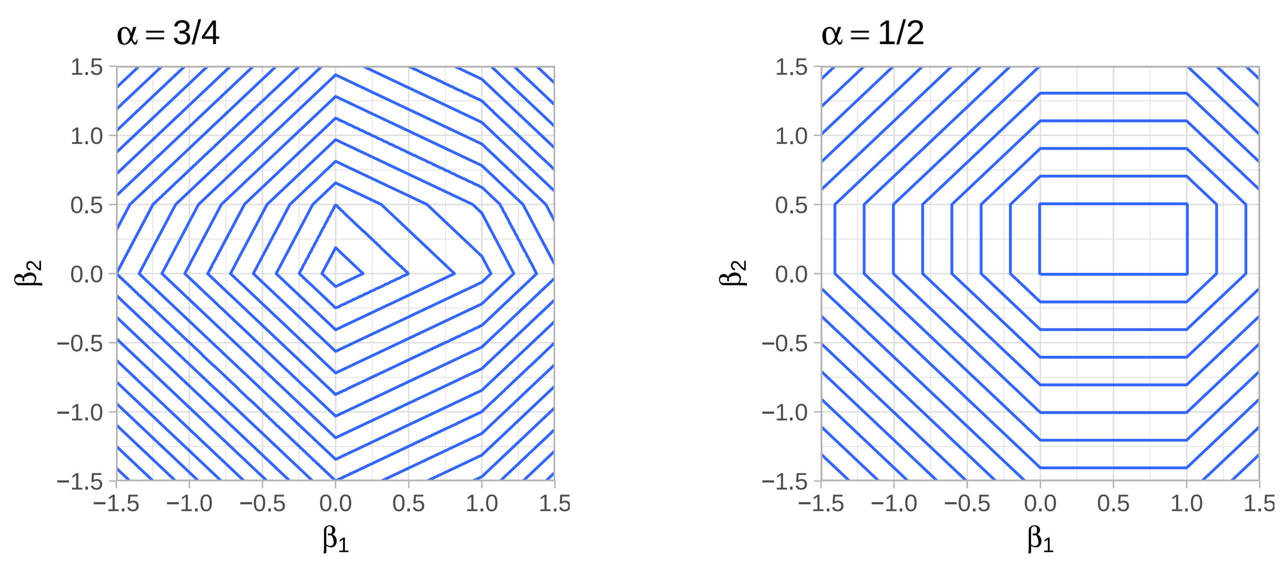
$\tilde{\beta}$：過去データでの推定値（初期推定値）
$\hat{\beta}$：現在データと$\tilde{\beta}$を組み合わせた推定値

$$\mathcal{L}(\beta;\tilde{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - X_i^\top\beta\right)^2 + \lambda P_\alpha(\beta;\tilde{\beta})$$

$$P_\alpha(\beta;\tilde{\beta}) = \alpha\|\beta\|_1 + (1-\alpha)\|\beta - \tilde{\beta}\|_1$$

$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^p} \mathcal{L}(\beta;\tilde{\beta})$$

罰則項$P_\alpha(\beta;\tilde{\beta})$の等高線：$\tilde{\beta} = (1, 1/2)^\top$



定理（二乗誤差）

Definition [Generalized Restricted Eigenvalue Condition (GRE)]　We say that the generalized restricted eigenvalue condition holds for a set $\mathcal{B} \subset \mathbb{R}^p$ if we have

$$\phi = \phi(\mathcal{B}) := \inf_{v\in\mathcal{B}} \frac{v^\top \frac{1}{n} X^\top X v}{\|v\|_2^2} > 0, \qquad X = (X_1, \ldots, X_n)^\top.$$

Theorem [Estimation Error]　Let β^* be the true parameter of β. Let $\varepsilon_i = Y_i - X_i^\top\beta^*$. Suppose that ε_i's are sub-Gaussian with scale σ. Let $S = \{j \in \{1,\ldots,p\} : \beta_j^* \neq 0\}$. Assume that $s = \#S \ll p$. Let $\Delta = \tilde{\beta} - \beta^*$. Assume that the generalized restricted eigenvalue condition holds for $\mathcal{B} = \mathcal{B}(\alpha, c, \Delta)$, where

$$\mathcal{B}(\alpha,c,\Delta) = \big\{v \in \mathbb{R}^p : (\alpha - c)\|v_{S^c}\|_1 + (1-\alpha)\|v - \Delta\|_1 \le (\alpha + c)\|v_S\|_1 + (1-\alpha)\|\Delta\|_1\big\},$$

with some constant $c > 0$. Then, we have

$$\|\hat{\beta} - \beta^*\|_2^2 \le \frac{(\alpha+c)^2\lambda_n^2 s}{\phi^2}\left(1 + \sqrt{1 + \frac{2(1-\alpha)\phi\|\Delta\|_1}{(\alpha+c)^2\lambda_n s}}\right)^2$$

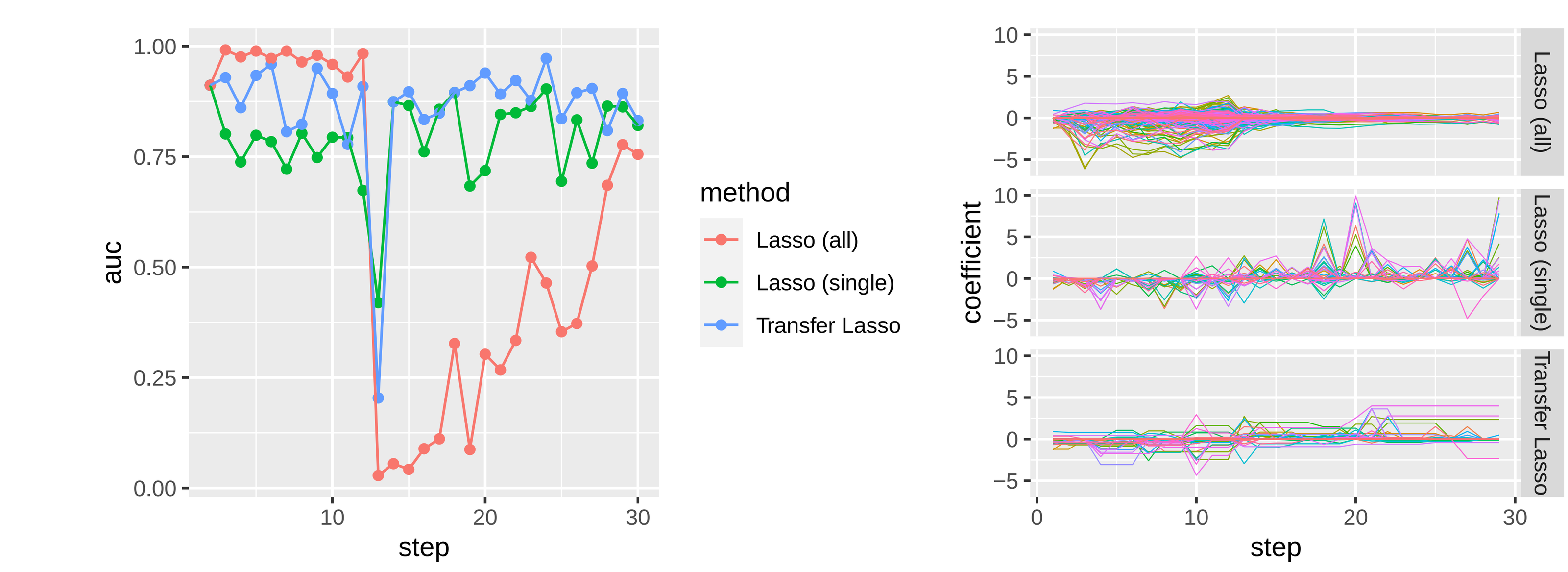
with probability at least $1 - \exp(-nc^2\lambda_n^2/2\sigma^2 + \log(2p))$.

定理の拡張として次が得られる．

過去と現在の状況が同じ場合： 過去情報が非常に大きい場合には，現在情報だけに基づいたLasso推定量よりも，αを小さく設定した Transfer Lasso に基づく推定量の方が二乗誤差が小さいことを証明できる．

過去と現在の状況が違う場合： 収束レートは遅くなるが依然として一致性は得られる．

数値結果



途中で大きな変化が起きているデータ（newsgroup message data from UCI database）に対する結果．

Lasso(all): その時点までのすべてのデータを使ったLasso．大きな変化が起きると以降も大失敗する（左図）．

Lasso(single): その時点のデータだけ使ったLasso．

Transfer Lasso: 過去情報を適度に使っている．AUCは全体的に安定してよい．急激な変化が起きたときはAUCが一時的に落ちるが，次の時点では過去情報を上手く捨ててAUCが復活している．回帰係数推定値も過去情報を利用しているため大きく変化せずに安定しやすい．

藤澤研究室の情報

研究テーマ： スパース・モデリング．ロバスト統計．ダイバージェンス．グラフィカル・モデリング．非対称分布．遺伝子発現データ．モデル選択．混合効果モデル．経時データ．欠測データ．多重検定．多重代入．クラスタリング．高欠測データ．異常検知．など

共同研究やコンサルテーション： 製薬企業・製造業など

大学院生： 思考力とやる気のある学生を歓迎します．
これまでの学生の情報はHPをご覧ください．