回帰分析における L^∞ 正則化による相関の強い変数の推定

松下 雅　(指導教員：吉田 裕亮)

1　はじめに

線形回帰モデルとは, 統計学における回帰分析の一種で, 1 つの従属変数 Y と, 1 つ以上の説明変数 X の関係を線形で記述したモデルである. 最小 2 乗法などによってモデルを推定し, 変数選択を実行することによって, 予測能力の高いモデルを構築することができる. しかし多数の説明変数をもつ線形回帰モデルなどには, 最小 2 乗法は過学習を起こしてしまう. 本研究では, これに対して, 最小二乗法に制約を付け加えたモデルを生成する. 一般的に知られている方法として L^1 ノルム, L^2 ノルム, の罰則項を課したモデルがある. 本研究では, あまり一般的ではない L^∞ の罰則項を使用し, L^∞ 正則化によってどのような推定が可能であるかを考察する.

2　線形回帰モデル

線形回帰モデル (Linear Regression) とは,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_n X_n + \epsilon$$

の回帰式を用いて, 説明変数 X_j の値から従属変数 Y の値を予測するモデルである. 特に, 説明変数 X_j が 1 つの場合を単回帰分析, 説明変数 X_j が 2 変数以上で構成される場合を重回帰分析と呼ぶ. 線形回帰モデルを推定する手法として最小二乗法などがある. 最小二乗法とは, 想定する関数が測定値に対して精度の高い近似にするために, 残差の二乗和 (RSS) を最小とするような係数決定を行うことを言う. y_i を実値データ $\hat{y_i}$ を予測値データとすると RSS は,

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

と表すことができ, RSS を最小化する最小二乗法推定は

$$\left({}^T\boldsymbol{X}\,\boldsymbol{X}\right)^{-1}\,{}^T\boldsymbol{X}\boldsymbol{Y}$$

で与えられる. 最小二乗法によって近似されたモデルは, 観測データの誤差が正規分布に従わない場合 (測定データに正規分布から大きく外れた外れ値を含む場合など) などには, 過学習を起こしてしまうことがある.

3　正則化

上記のような過学習を防ぎ, 汎化能力を高めるために使われる一つのアプローチとして正則化がある. 正則化の具体的手法は, 損失関数に罰則項を加えた式を最小化するように最適化問題を解く問題に置き換えることである. その指標として一般的には L^1 ノルム, L^2 ノルムが主に用いられる.

L^2 ノルムとは 2 乗和 (ユークリッド距離) で表されたものである. 回帰係数の L^2 ノルムを用いた正則化推定法をリッジ (Ridge) 回帰と呼び,

$$RSS + \lambda \sum_{i=1}^{p} \beta_j^2$$

を最小化するモデルを表す. Ridge 回帰は, 回帰係数は完全に 0 にはならないので, 説明変数が非常に多いモデルではモデルの解釈が複雑になるという欠点がある. そのため, 変数選択の観点から, 実データへは適していない.

L^1 ノルムはベクトル成分の絶対値の和 (マンハッタン距離) で表されたもので, 回帰係数の L^1 ノルムを用いた正則化推定法をラッソ (Lasso) 回帰と呼び,

$$RSS + \lambda \sum_{i=1}^{p} |\beta_j|$$

を最小化するモデルを表す. Lasso 回帰は, Ridge 回帰とは異なり回帰係数を 0 にするため, モデルの選択と説明変数を行えるメリットがある. ただし, 説明変数間の相関を考慮できないというデメリットもある. Lasso 回帰では, 相関が高い説明変数のグループの中から 1 つの説明変数だけがモデル選択に使用される. よって, 従属変数と関連がある説明変数をモデルに選択し損なうことが起こる.

ここで L^∞ を罰則項とする正則化を考える. L^∞ 正則化は, 回帰係数の絶対値の最大を罰則項とし,

$$RSS + \lambda \max_{1 \le i \le n} |\beta_i|$$

を最小化させる. L^∞ 正則化によってどのような推定が可能なのかの検証を行う.

4　実験

本研究では, L^∞ 正則化を用いることで, 相関関係の強い説明変数を探索する方法を実験によって調べる. 予め相関の強い組み合わせを持つようにシュミレーションデータを正規乱数で生成し, L^∞ 正則化を用いて

$$RSS + \lambda \max_{1 \le i \le n} |\beta_i|$$

を最小にする回帰係数 β の値を推定する.

その際に, λ を変化させ, β の値を観察することで相関の強い説明変数の組を見つける実験を行う.

手順を以下にする

1. 元のシュミレーションデータより L^∞ 正則化を行い, λ の値を変化させ β の値を観察する
2. 回帰係数の変化から相関が強いと思われる説明変数の組の値を, それらの回帰係数が大きくなるようにデータを調整する.
3. 上記のように置き換えた入力データで, 再度 λ の値を変化させながら L^∞ 正則化を行い β の変化を観察する

4. ・回帰係数 β_k と β_ℓ が互いに近づき β_k と β_ℓ の間で一致した時, 説明変数 X_k と X_ℓ は相関が強いと推定される

   ・回帰係数 β_k と β_ℓ ($\beta_k > \beta_\ell$) について, β_1 が β_2 に近づくのみの場合 X_k と X_ℓ には相関はないものと推定される

5. 2, 3, 4 を繰り返し, 説明変数の相関の強い組を見つけていく

実験例 1

説明変数 4 つのうち 1 組 (2 つ) の相関がある場合のシュミレーションデータを用意する. 具体的には最小二乗法を行った際に, $\beta_1 = 15, \beta_2 = 9, \beta_3 = 7, \beta_4 = 3$ の線形モデル ($Y = 17X_1 + 13X_2 + 8X_3 + 3X_4$) になるシュミレーションデータを用いる.

1. これらのデータに L^∞ 正則化を行うと, λ の値の変化に伴って図 1 のようになった.

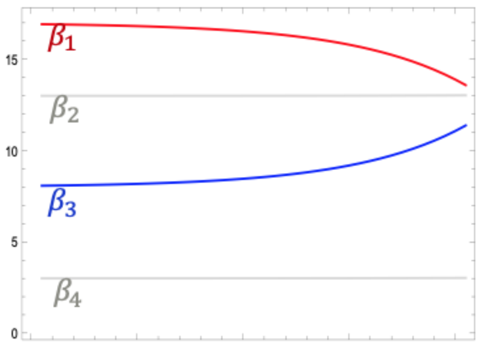

図 1: λ の値の変化に伴各 β の値の変化

   ここで最初に動き出した β の値から, X_1 と X_3 の相関が強い可能性が推測される

2. ここで説明変数 X_3 を 3 で割ったデータに置き換える. 再度, λ の値を変化させながら L^∞ 正則化を行う. その変化を図 2 に示す

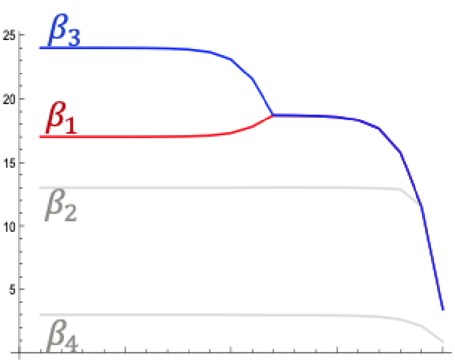

図 2: λ の値の変化に伴各 β の値の変化

3. 上記グラフより, β_1 と β_3 の関係から, X_1 と X_3 に相関関係が強いことが見えて取れる.

実験例 2

説明変数 6 つのうち 1 組 (3 つ) の相関がある場合のシュミレーションデータを作成する. 線形モデル ($Y = 19X_1 + 15X_2 + 14X_3 + 12X_4 + 7X_5 + 4X_6$) を設定する.

1. L^∞ 正則化を行い β の値の変化を観察すると, 図 3 のようになった. ここで最初に動き出した β の

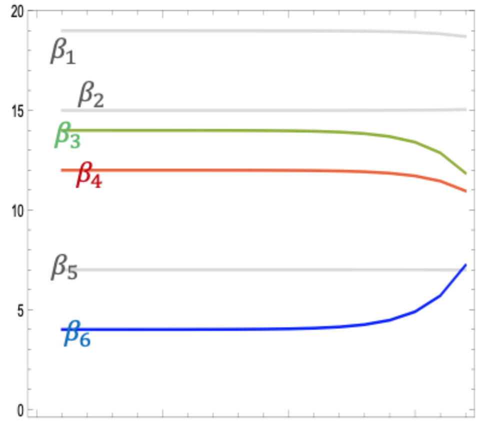

図 3: λ の値の変化に伴各 β の値の変化

   値から, X_3 と X_4 と X_6 に相関がある可能性が推測される.

2. ここで説明変数 X_3 を 2 で, X_6 を 12 で, X_4 を 5 で割ったデータに置き換える. 再度, λ の値を変化させながら L^∞ 正則化を行う. その変化を図 4 に示す.

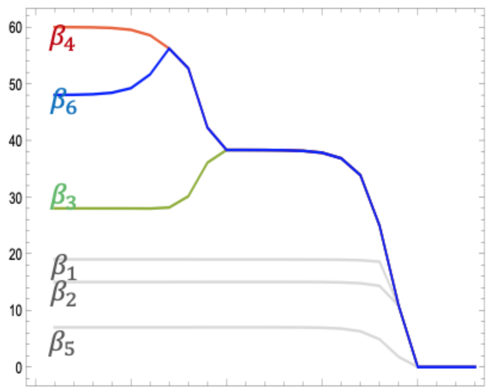

図 4: λ の値の変化に伴各 β の値の変化

3. 上記グラフより β_3 と β_4 と β_6 の関係から, X_3 と X_4 と X_6 の相関関係が強いことが見える.

5 考察

L^∞ 正則化を行い正則化パラメータ λ の値の変化に伴う回帰係数 β 変動を観察することで, 相関の強い変数達を同定することが可能と考えられる.

参考文献

Gareth James, Daniela Witten, Trevor Hastie Robert Tibshirani, An Introduction to Statistical Learning, Springer, New York Heidelberg Dordrecht London, 2013