

# Ridge と Lasso の推定

2019 年 5 月 8 日

大森 夢拓

## 1 はじめに

複雑な非線形構造を内在する現象の分析では、線形モデルや多項式モデルのような特定の関数式で捉えることは難しく、線形モデルではなく非線形モデルで考える必要がある。非線形モデルを推定する方法の一つである正則化法の中でも、本稿では Ridge 回帰と Lasso を用いた推定を紹介する。また、それらを用いて実際のデータに対して行った計算機実験についても紹介する。

## 2 Ridge 回帰と Lasso の推定

正則化項のパラメータベクトルの長さの概念を一般化した  $L_q$  ノルムと呼ばれる正則化項を用いた正則最小 2 乗法は

$$S\gamma(w) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^m w_j b_j(x_i) \right)^2 + \gamma \sum_{j=1}^m |w_j|^q \quad (1)$$

のように拡張される。(1) が  $L_1$  ノルムの場合を Lasso,  $L_2$  ノルムの場合を Ridge 回帰と呼ぶ。

### 2.1 Ridge 回帰

線形回帰モデル

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

において、切片を除く回帰係数の 2 乗和をペナルティ項としてとることで、説明変数  $\mathbf{x} = (x_1, x_2, \dots, x_p)^\top$  間の強い相関によって生じる回帰係数の推定値の不安定性を回避する方法である。Ridge 推定量の求め方は次のようになる。

各説明変数に関するデータの平均値を  $\bar{x}$ , 中心化したデータを  $z_{ij} = x_{ij} - \bar{x}_j$  とすると、線形回帰モデル (2) は次のように変形できる。

$$y_i = \beta_0 + \beta_1 \bar{x}_{i1} + \beta_2 \bar{x}_{i2} + \cdots + \beta_p \bar{x}_{ip} + \beta_1 (x_{i1} - \bar{x}_{i1}) + \beta_2 (x_{i2} - \bar{x}_{i2}) + \cdots + \beta_p (x_{ip} - \bar{x}_{ip}) + \epsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

線形回帰モデル (3) を,  $\beta_0^* = \beta_0 + \beta_1 \bar{x}_{i1} + \cdots + \beta_p \bar{x}_{ip}$ , 成分が全て 1 の  $n$  次元ベクトル  $\mathbf{1} = (1, 1, \dots, 1)^\top$ , 各説明変数

の回帰係数からなる  $p$  次元ベクトル  $\beta_1 = (\beta_1, \beta_2, \dots, \beta_p)^\top$  を用いることで次のような行列形式に変形できる。

$$\mathbf{y} = \beta_0^* \mathbf{1} + Z\beta_1 + \epsilon$$
$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{np} \end{bmatrix}, \quad z_{ij} = x_{ij} - \bar{x}_j; \quad (4)$$
$$i = 1, \dots, n; \quad j = 1, \dots, p$$

行列形式の線形モデル (4) より、回帰係数の Ridge 推定量は、

$$S\gamma(\beta_0^*, \beta_1) = (\mathbf{y} - \beta_0^* \mathbf{1} - Z\beta_1)^\top (\mathbf{y} - \beta_0^* \mathbf{1} - Z\beta_1) + \gamma \beta_1^\top \beta_1 \quad (5)$$

の偏微分による最小化によって、次のような切片の推定量と回帰係数ベクトルの Ridge 推定量が与えられる。

$$\hat{\beta}_0^* = \bar{y}, \quad \hat{\beta}_1 = (Z^\top Z + \gamma I_p)^{-1} Z^\top \mathbf{y}. \quad (6)$$

以上より、(2) の切片は  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \cdots - \hat{\beta}_p \bar{x}_p$  の切片と推定されるので、回帰係数ベクトルのリッジ推定量は、 $\beta_1 = (\beta_1, \beta_2, \dots, \beta_p)^\top$  と、中心化したデータから構成された計画行列  $Z$  を用いた  $S\gamma(\beta_1) = (\mathbf{y} - Z\beta_1)^\top (\mathbf{y} - Z\beta_1) + \gamma \beta_1^\top \beta_1$  の最小化によって与えられる。

### 2.2 Lasso

線形回帰モデル (2) において、切片を除く回帰係数の和をペナルティ項としてとることで、パラメーターの一部を完全に 0 と推定するようなモデルの推定・変数選択を同時に実行できる方法である。切片を除く回帰係数はデータを中心化することによって切片と切り離して測定できるため、

$$S\gamma(\beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^p \beta_j (x_{ij} - \bar{x}_j) \right)^2 + \gamma \sum_{j=1}^p |\beta_j|. \quad (7)$$

の最小化によって与えられる。ところが、 $L_1$  正則化項が微分不可能であるため解析的に求めることはできない。このため、Fu(1998) による shooting アルゴリズムや Efron et al.(2004) による LARS(Least Angle Regression) と呼ばれる計算方法が用いられる。

また、ラグランジュの未定乗数法を適用すると、次の様な制約条件付きのパラメータベクトルの最小化と同等になる。

$$\sum_{i=1}^n \left( y_i - \sum_{j=1}^m w_j b_j(x_i) \right)^2, \quad \text{制約条件} \quad \sum_{j=1}^m |w_j|^q \leq \eta. \quad (8)$$

### 3 計算機実験

データセットとしてアメリカのボストン州の住宅価格を、それに影響を与えていると思われる説明変数 13 個との関係を線形回帰モデルでモデル化した。データ数は 506 である。説明変数は以下の通りである。

- $x_1$ : 犯罪率

$x_3$ : 非商用地

$x_5$ : 窒素酸化物

$x_7$ : 築年

$x_9$ : ハイウェイアクセス指数

$x_{11}$ : 生徒と教師の比率

$x_{13}$ : 低所得者の割合
- $x_2$ : 宅地の割合

$x_4$ : チャールズ川流域か否か

$x_6$ : 部屋数

$x_8$ : ビジネス地域への距離

$x_{10}$ : 固定資産税

$x_{12}$ : 有色人種の割合

通常の最小 2 乗法と、正則化項の係数を  $\alpha$  として 0.01, 0.1, 1, 10, 100 を値にした 5 通りにクロス・バリデーション法を用いて決定した値 (Ridge 回帰 : 79.43, Lasso : 0.16) を合わせた計 6 通りの Ridge 回帰, Lasso で回帰係数の推定を行い、それぞれ比較した。この実験結果を図 1, 図 2, 図 3 に示し, Lasso の解パスを図 4, 決定係数の推移を図 5, 表 1 に示す。

図 1, 図 2, 図 3 より, 最小 2 乗法に比べて  $\alpha$  が大きくなるにつれて, Ridge 回帰では  $x_3 \sim x_5$  の  $\beta$  の値が 0 に向かって縮小しており, Lasso では全ての値が 0 に向かって縮小している。特に, Lasso では  $\alpha=100$  の場合は全ての  $\beta$  の値が 0 となっている。これらの結果から, Ridge 回帰および Lasso での正則化項によるペナルティの与えられ方や, Lasso ではペナルティによって変数選択が行われていることを確認することができた。

また, 図 4 の Lasso の解パスを確認することで, どの説明変数がどの大きさのペナルティによって変数選択が行われているかも確認できた。

最後の図 5, 表 1 では, 設定した  $\alpha$  の値における決定係数の推移が確認できるが, 今回の実験では  $\alpha$  が大きくなるにつれて決定係数が小さくなっていることから, 今回のデータでは Ridge 回帰及び Lasso の恩恵を受けられなかったことが分かった。

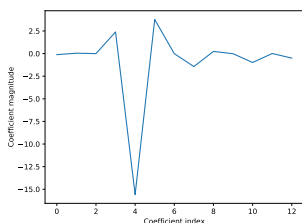


図 1: 最小 2 乗法

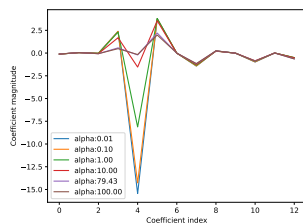


図 2: Ridge 回帰

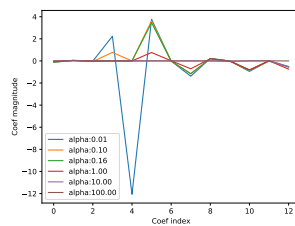


図 3: Lasso

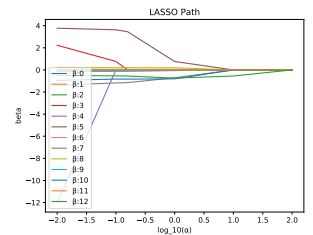


図 4: Lasso の解パス

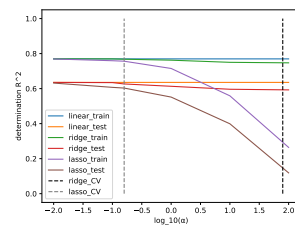


図 5: 決定係数

表 1: 決定係数の表

alpha	0.01	0.1	0.16	1	10	79	100
train							
linear	0.77						
ridge	0.77	0.77		0.77	0.76	0.75	0.75
lasso	0.77	0.76	0.76	0.72	0.56		0.25
test							
linear	0.64						
ridge	0.64	0.63		0.63	0.61	0.60	0.59
lasso	0.63	0.61	0.60	0.55	0.40		0.12

### 4 まとめ

Ridge 回帰と Lasso を中心に, 正則化法や非線形回帰モデルについて学ぶことができた。数式だけでなく, Python のパッケージを用いた計算機実験による可視化を通して, 自分の中でさらにイメージを具体化させることができた。

今後は, Lasso の計算に用いるアルゴリズムを理解して, パッケージを使わずに実装することが課題である。

### 参考文献

[1] 小西貞則, 『多変量解析入門』. 岩波書店, 2009.