論文紹介

nanimonaiyo

山本 藤也（Touya Yamamoto）
u220067@st.pu-toyama.ac.jp

富山県立大学 情報システム工学科

December 12, 2025

論文の概要

対象論文

- Regression Shrinkage and Selection via the Lasso
- Robert Tibshirani, JRSS B, 1996

この論文がやっていること

- 回帰モデルで予測精度を保ちつつ、不要な変数を自動で 0 にする方法を提案する。
- その方法を Lasso（Least Absolute Shrinkage and Selection Operator）と呼ぶ。

通常の回帰と問題点

通常の回帰（最小二乗法）

- モデル：

$$y = X\beta + \varepsilon$$

- 推定：

$$\hat{\beta}^{\mathrm{OLS}} = \arg\min_{\beta} \|y - X\beta\|_2^2$$

問題点

- 説明変数が多いと分散が大きくなり予測が不安定になる。
- すべての変数に係数が付くためどの変数が重要か分かりにくい。

Lasso の定義

L1 正則化付き回帰

- Lasso は次の問題を解く：

$$\hat{\beta}^{\mathrm{lasso}} = \arg\min_{\beta} \left\{ \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

- λ が大きいほど $\sum_j |\beta_j|$ のペナルティが強くなり、係数が 0 に近づく。

Lasso の特徴

特徴

- **一部の係数を 0 にする** ため、重要な変数だけ残る。
- Subset selection より連続的で安定であり、Ridge 回帰よりスパースになる。

L1 正則化が 0 を作る理由

直交デザインの場合の解

- $X^\top X = I$ のとき、Lasso の解は

$$\hat{\beta}_j^{\mathrm{lasso}} = \mathrm{sign}\left(\hat{\beta}_j^{\mathrm{LS}}\right)\left(|\hat{\beta}_j^{\mathrm{LS}}| - \lambda\right)_+$$

- $|\hat{\beta}_j^{\mathrm{LS}}| \le \lambda$ なら $\hat{\beta}_j^{\mathrm{lasso}} = 0$。

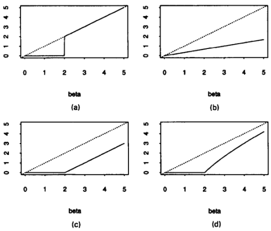

図 1: 4 手法（Subset・Ridge・Lasso・Garotte）の縮小挙動

幾何学的イメージ

L1 と L2 の違い

- Ridge：$\sum_j \beta_j^2 \leq c$（円形の制約領域）
- Lasso：$\sum_j |\beta_j| \leq t$（ひし形の制約領域）
- 等高線と制約領域の最初の接点が解になる。

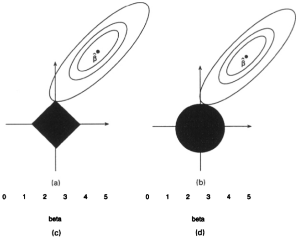

図 2: **L1 L2 の制約領域と等高線**

→ひし形の角が軸上にあり、係数が 0 の解を選びやすい。

前立腺がんデータの結果（1）

データとモデル

- PSA（前立腺特異抗原）の対数を目的変数とする。
- 8 つの説明変数（lcavol, lweight, age, ⋯）を標準化して回帰。

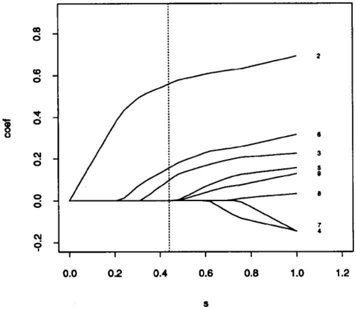

図 3: 正則化の強さによる Lasso 係数パス

前立腺がんデータの結果（2）

Lasso が選んだ変数

- GCV により非ゼロの係数は 3 つ：
  - lcavol（がん体積）
  - lweight（前立腺重量）
  - svi（精嚢浸潤）

係数の例（Table1）

- lcavol：OLS 約 0.69 → Lasso 約 0.56
- lweight：OLS 約 0.23 → Lasso 約 0.10
- svi：OLS 約 0.32 → Lasso 約 0.16

Example 1：少数の大きな効果

設定

- 真の係数：

$$\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)^\top$$

- 非ゼロは 3 つだけのスパース構造。

結果（Table 3）

- 中央 MSE の比較：
  - 最小二乗：≈ 2.8
  - 部分集合選択：≈ 2.4
  - Lasso（GCV）：≈ 1.9（最良）
  - Ridge：≈ 3.2
- Lasso はスパース構造に強い。

Example 4：高次元の例

設定と結果（Table 8）

- $N = 100,\ p = 40$ の高次元設定。
- Ridge が最良（MSE 約 57.4）。
- Lasso（GCV）は MSE 約 64.9。

ここから言えること

- 多くの小さな効果が分散している場合は Ridge が有利。
- Lasso は解釈性の高いスパース解を提供。

まとめ

Lasso のポイント

- Lasso は L1 正則化により係数の一部を 0 にする。
- Subset のスパース性と Ridge の安定性を両立する。
- スパース構造があるときに強い予測性能。

この論文の意義

- L1 正則化モデルの基礎を確立した重要論文。
- スパース推定という概念を統計学・機械学習に広めた。

記号一覧（補足）

- y：目的変数（$N \times 1$）
- X：説明変数行列（$N \times p$）
- β：回帰係数（$p \times 1$）
- $\hat{\beta}^{\mathrm{LS}}$：最小二乗推定量
- $\hat{\beta}^{\mathrm{lasso}}$：Lasso 推定量
- λ：L1 正則化パラメータ
- $(x)_+ = \max(x, 0)$：soft-thresholding
- $\mathrm{sign}(x)$：符号関数