

Penalized regression

川田恵介

keisukekawata@iss.u-tokyo.ac.jp

2025-05-06

1 罰則付き回帰

1.1 背景: 過剰適合

- 多くの推定方法が、基本的に、モデルとデータの矛盾 (empirical risk) を減らすように、予想モデル $g(X)$ を推定する
- 複雑なモデル (線型モデル $g(X) = \beta_0 + \dots + \beta_L X_L$ では、 β の数が多いモデル) を推定すると、
 - ▶ より多くのパラメタをデータに合うように調整できる
 - データへの当てはまりは改善するが、やりすぎると母平均から乖離する

1.2 背景: 予測問題の要求

- OLS において高性能の予測モデルを推定するためには、Population OLS \simeq 母平均、を目指す必要がある
- 十分に複雑なモデルを推定する必要があるが、Population OLS を高い精度で推定するためには、単純化する必要がある
 - ▶ 伝統的には、研究者の”背景知識(含む思い込み、惰性)“を用いてきた
 - ▶ 罰則付き回帰は、よりデータ主導の方法を提案する

1.3 罰則付き回帰

1. 十分に複雑なモデル $g(X)$ を研究者が設定
2. 何らかの基準 (後述) に基づいて複雑性への罰則を設定
3. 以下を最小化するように β を推定
 - $(Y - g(X))^2$ のデータ上の平均 + 複雑性への罰則

1.4 LASSO

- 複雑性への罰則 = $\underbrace{\lambda}_{\text{Tuning Parameter}} \times \text{複雑性の指標}$
 - ▶ LASSO: 複雑性の指標 = $|\beta_1| + \dots + |\beta_L|$
 - ▶ Ridge: 複雑性の指標 = $\beta_1^2 + \dots + \beta_L^2$
- どちらも $g(X) = \beta_0$ を最も単純なモデルとして評価
 - ▶ データに当てはまるように推定すると、 $\beta_0 = \text{データ上の} Y \text{の平均}$
 - 単純平均を最も単純なモデルとして評価

1.5 OLS

- $(Y - g(X))^2$ のデータ上の平均 + 複雑性への罰則

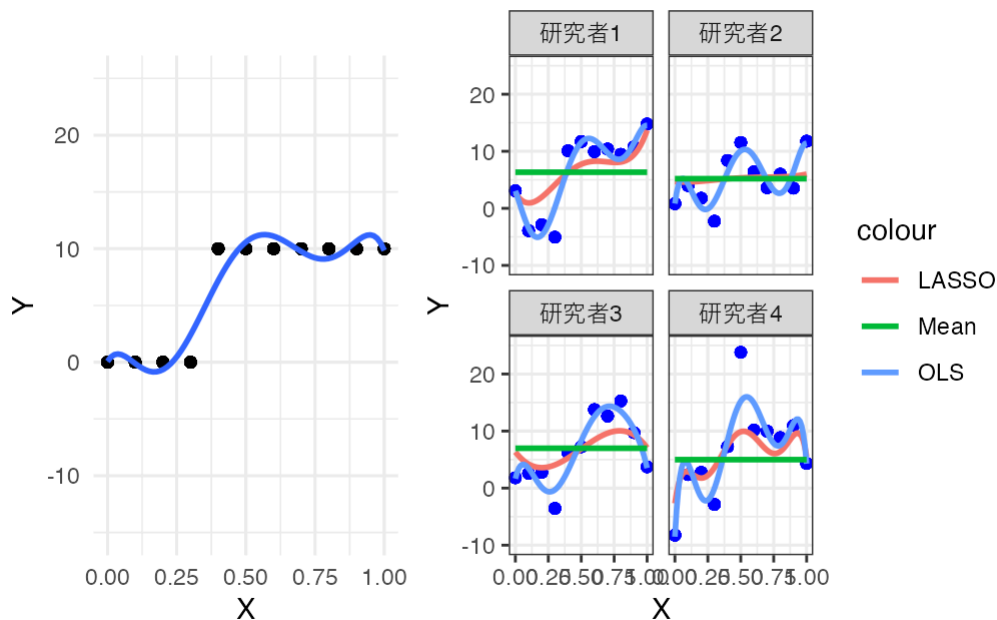
1.6 例え話: 環境税/補助金

- 企業が費用を最小化するように生産計画を立てること自体は有益だが、"自由放任"では環境負荷が過度に高まる (負の外部性)
 - ▶ 環境負荷 = 複雑性
 - ▶ 税額 = λ
 - ▶ 自由放任 = OLS
 - ▶ 環境負荷への税/補助金 = 罰則

1.7 λ の役割

- λ を無限大に大きくすると、 $\beta_1 = \dots = \beta_L = 0$
 - ▶ 推定された予測モデル $\hat{g}(X) = \text{データ上の平均}$
- $\lambda = 0$ にすると、 $\hat{g}(X) = \text{OLS}$
- λ を変更することで、OLS と単純平均の間のモデルとなる

1.8 数値例: 200 サンプル



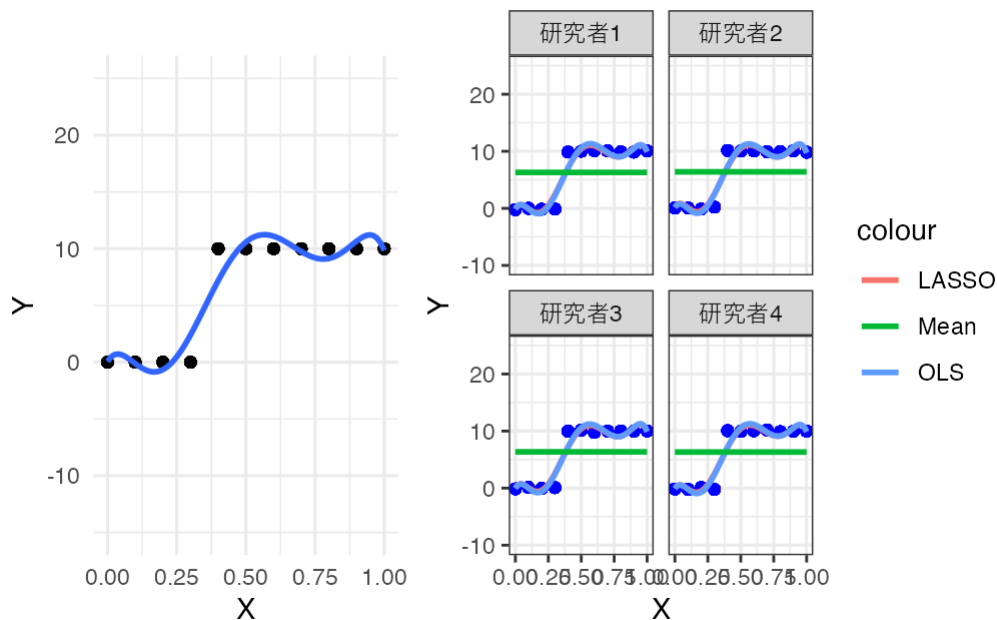
1.9 λ の選び方

- 推定されたモデル $\hat{g}(X)$ と母平均 $\mu(X) = E[Y | X]$ の乖離を小さくするように設定したい
- いくつかの方法が提案
 - ▶ Taddy (2017): 赤池情報基準の利用 (gamlr package)
 - ▶ Tibshirani (1996) : 交差推定の利用 (後述, glmnet package)
 - ▶ Belloni et al. (2014) : 理論指標の利用 (hdm package)
 - 詳細は CausalMLBook 第3章 参照

1.10 LASSO の推定対象

- 母集団 (無限大の事例数を持つランダムサンプリングデータ)に LASSO を適用すると、**緩やかな条件のもとで Population OLS が算出される**
 - ▶ LASSO は、Population OLS の一致推定量
- 緩やかな条件: 事例数が増加するにつれて、 λ が十分な速度で 0 に収束する
 - ▶ 無限大の事例数のもとでは、実質的に OLS が行われる

1.11 数値例: 200000 サンプル



1.12 LASSO の推定値

- LASSO および Ridge は、 β の数が事例数を上回る(!!!)モデルも推定できる
 - ▶ OLS では不可能
- LASSO では、 β の値が厳密に 0 になり得る
 - ▶ 変数がモデルから、"データ主導"で脱落する (変数選択が行われる)
 - ▶ Ridge/OLS では行われない

1.13 まとめ

- 罰則付き回帰：過剰適合を抑えながら、十分に複雑なモデルを推定する手法
 - ▶ LASSO はその代表例
 - 変数選択が行われることも注目されてきた

2 実践

2.1 予測問題: 他の推定方法との併用

- 常に優れた予測モデルを生み出せるとは限らない
 - ▶ β の数は多いが、その大部分は重要ではない (β の値が 0 に近い)場合に有効
 - 事前には判断できない

- ▶ OLS や RandomForest, Booting などと、(Test data を用いて)比較、ないし Stacking 法(後述)を活用する必要がある
- 上記を行うのであれば、比較的安心して用いることができる

2.2 記述問題: 嚴重注意

- LASSO により推定された β を、社会の特徴を捉えるものとして活用する
- 推定対象は OLS と同じなので、OLS と同じような解釈ができそうな気がするが
 - ▶ 一般にバイアスがあり、信頼区間計算ができない
 - ▶ 変数選択についても、誤った変数選択が生じる
- さまざまな試み (Chernozhukov et al., 2015; Fan et al., 2011; Kuchibhotla et al., 2022) があるが、母平均を、“なんとなく”近似する”Black box”モデルと捉えた方が、(私見では) 実践的

2.3 Reference

Bibliography

- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650.
- Chernozhukov, V., Hansen, C., & Spindler, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annu. Rev. Econ.*, 7(1), 649–688.
- Fan, J., Lv, J., & Qi, L. (2011). Sparse high-dimensional models in economics. *Annu. Rev. Econ.*, 3(1), 291–317.
- Kuchibhotla, A. K., Kolassa, J. E., & Kuffner, T. A. (2022). Post-selection inference. *Annual Review of Statistics and Its Application*, 9(1), 505–527.
- Taddy, M. (2017). One-step estimator paths for concave regularization. *Journal of Computational and Graphical Statistics*, 26(3), 525–536.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267–288.