LASSO
機械学習

川田恵介
東京大学
keisukekawata@iss.u-tokyo.ac.jp

2025-09-23

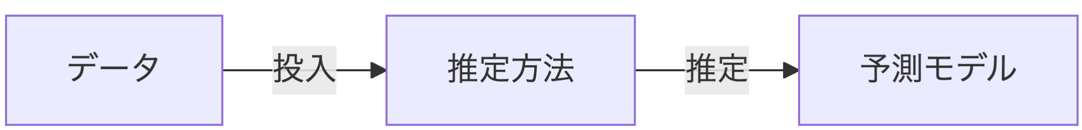
1 LASSO

1.1 目標

- データ主導の線型モデルの推定方法を学ぶ
  - 代表的な手法である LASSO を紹介
- “モデルの複雑性を調整する”というイメージを把握する

1.2 復習

- 予測モデルの構築: データをモデルに変換する
  - 変換 = 推定/学習
- 変換手順は、人間が決める必要がある
  - 変換手順 = 定式化と推定方法/アルゴリズム



1.3 定式化の例: ベンチマーク

- $Price \simeq \beta_0 + \beta_1 \times \underbrace{RoomNumber}_{\text{部屋数}}$

1.4 定式化の例: より単純なモデル

- $$Price \simeq \beta_0$$
  - $\beta_1 = 0$ と”人間”が決めている $\rightarrow$ データに決めさせる部分が減っている
    - $\beta_0 = Price$ の単純平均 (東京全体での平均価格)が予測”モデル”となる

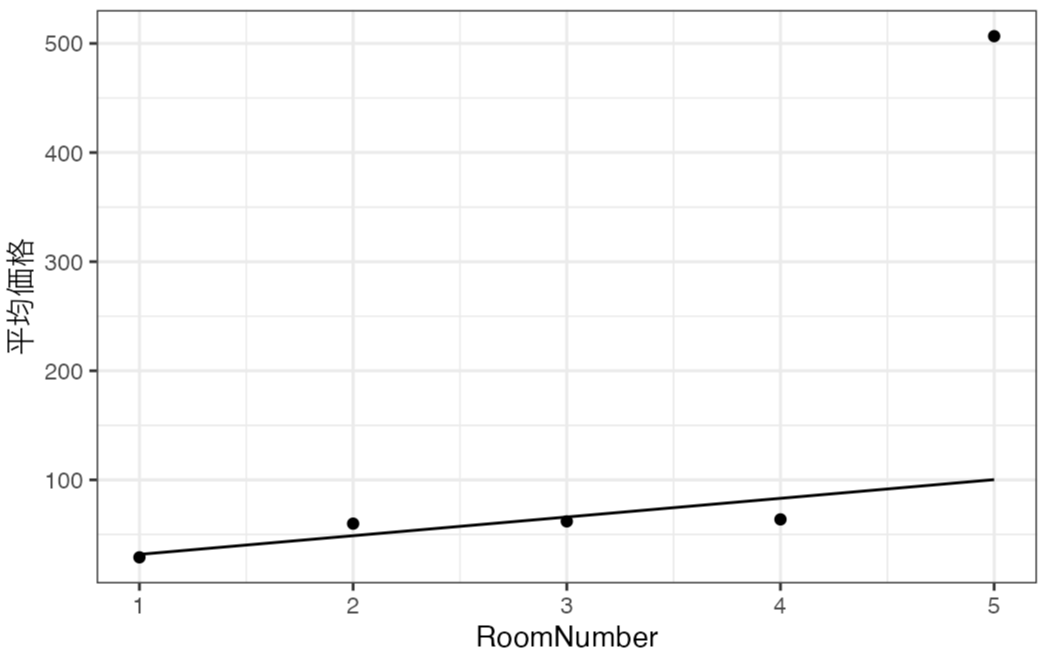
1.5 定式化の例: より複雑なモデル

- $Price \simeq \beta_0 + \beta_1 \times RoomNumber + \beta_2 \times RoomNumber^2$
$$+\beta_3 \times RoomNumber^3$$
- $\beta_2 = \beta_3 = 0$ と”人間”が決れば、ベースモデルと一致
  - より多くをデータに決定させる定式化

1.6 定式化の例: 凄く複雑なモデル

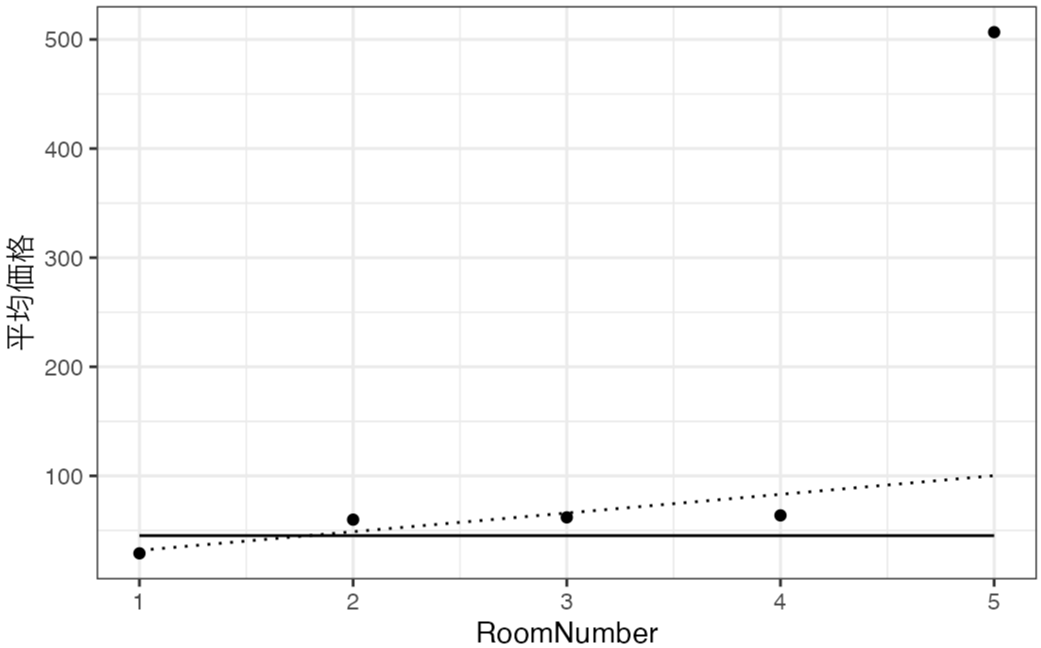
- $Price \simeq \beta_0 + \beta_1 \times RoomNumber + \beta_2 \times RoomNumber^2$
$$+\beta_3 \times RoomNumber^3 + \beta_4 \times RoomNumber^4$$
- $\beta_4 = 0$ と”人間”が決れば、より複雑なモデルと一致

1.7 例: ベンチマーク



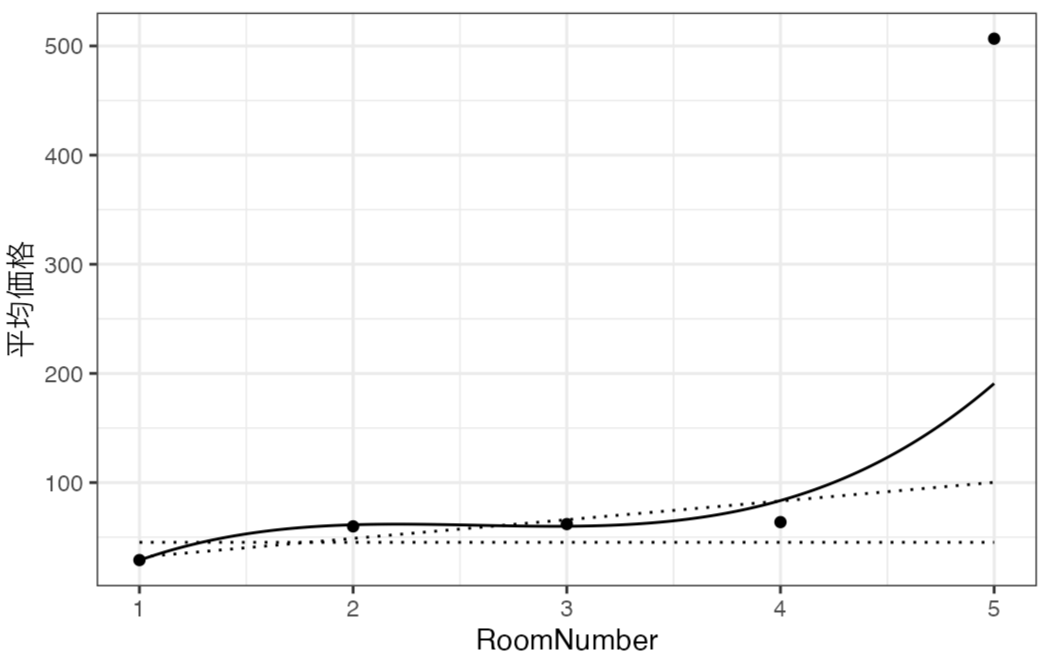
- 5 部屋物件の予測価格 $\simeq 100$

1.8 例: 単純なモデル



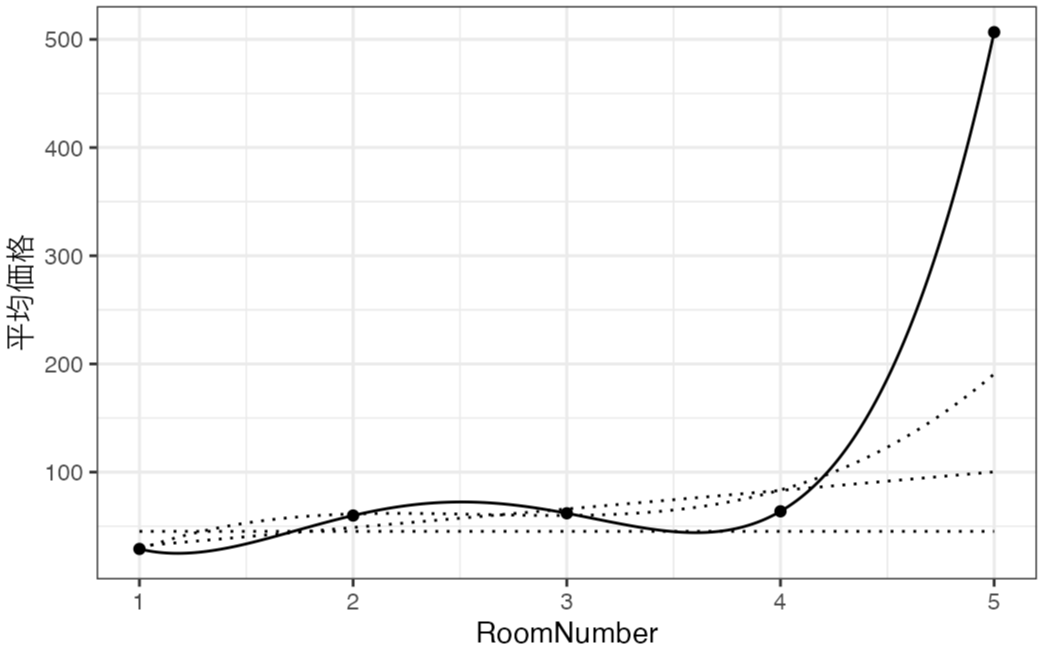
- 5 部屋物件の予測価格 $\simeq 45$

1.9 例: 複雑なモデル



- 5 部屋物件の予測価格 $\simeq 191$

1.10 例: 凄く複雑なモデル



- ５部屋物件の予測価格 $\simeq 507$

1.11 過剰適合/過学習

- 複雑なモデルは、より多くをデータが決定し、データ上の平均値とモデルが完全一致している
  - データに最も適合するが、、、
    - 予測性能が低いケースが多い
- **過学習/過剰適合:** モデルが、データと過度に適合する(事例から過度に学び過ぎてしまう)現象
  - 非常に少数の事例から生じたデータ特徴も、モデルに反映してしまう
    - ５部屋の物件は３事例しか存在しない

2 複雑性への罰則

2.1 目標

- データにより多くのことを決めさせるには、複雑なモデルを推定する必要がある
  - 複雑なモデルを OLS 推定すると、データへの過剰適合が生じる
- 複雑なモデルを「データに程よく適合させたい」

2.2 例え話

- 入門経済学: 生産方法を企業の自主的な意思決定に任せると
  - 同じ生産量を達成する方法の中で、最も費用が少ない方法が選ばれる
    - 希少な資源の利用を減らせ、それなりに望ましい
  - 一般に”負の外部性”が生じる
    - 温室効果ガスの過剰排出等
  - 社会的に望ましい水準に誘導するための政策が必要
    - 総量規制、環境税、補助金等

2.3 罰則付き回帰

- 線型モデル $Y \simeq \beta_0 + \beta_1 X_1 + ..$ を推定
- 以下を最小にするようにβを推定

$$\underbrace{\text{データへの不適合度}}_{(Y-\text{予測値})^2\text{のデータ上の平均}} + \text{複雑性への罰則}$$

2.4 OLS の再解釈

- OLS: 以下を最小にするように推定

$$\text{データへの不適合度}$$

- 定式化を前提に、データへの適合**のみ**を目指す推定方法
  - 過剰適合の原因

2.5 罰則の定式化

- $$\text{複雑性への罰則} = \underbrace{\lambda}_{\text{税率}} \times \text{複雑性の測定値}$$
  - LASSO においては、

    $\text{複雑性の測定値} = |\beta_1| + |\beta_2|..$
- λ は予測性能が良くなりそうな値を設定する
  - いろいろな方法が提案されている
  - 本講義では、hdm package で実装されている理論的指標を用いる

2.6 λ の影響

- 最も単純なモデル (複雑性の測定値 $= 0$)は、

$$\beta_1 = \beta_2 = .. = 0$$

- $Y \simeq \beta_0 =$ **単純平均**

- $|\beta_1|$ が大きくなれば、予測値が X_1 により反応する
  - より複雑な予測モデル

2.7 λ の影響

- $\lambda \to \infty$ であれば、複雑性の低下が最優先
  - $\beta_1 = \beta_2 = .. = 0$
  - 予測モデル $=$ 単純平均

2.8 λ の影響

- $\lambda = 0$ であれば、データへの当てはまりのみを追求
  - OLS!!!
  - 予測モデル $=$ OLS で推定されたモデル

2.9 例

```
hdm::rlasso(Price ~ Size + Tenure + Distance + District, data, post = FALSE)
```

```


Call:
rlasso.formula(formula = Price ~ Size + Tenure + Distance + District,
    data = data, post = FALSE)

Coefficients:
     (Intercept)             Size           Tenure         Distance
         15.2453           1.1703          -0.6395          -0.9943
  District中央区     District中野区       District北区   District千代田区
          2.8331           0.0000         -13.0874          18.3989
  District台東区     District品川区     District大田区     District文京区
         -4.2591           0.6435          -8.3461           0.3467
  District新宿区     District杉並区     District板橋区   District江戸川区
          3.4565           0.0000         -15.6118         -26.6841
  District江東区     District渋谷区       District港区     District目黒区
         -9.8651          15.1645          32.4377           8.7597
  District練馬区     District荒川区     District葛飾区     District豊島区
        -15.0456         -18.2211         -22.4133          -0.7486
  District足立区     District墨田区
        -23.5088          -9.4562
```

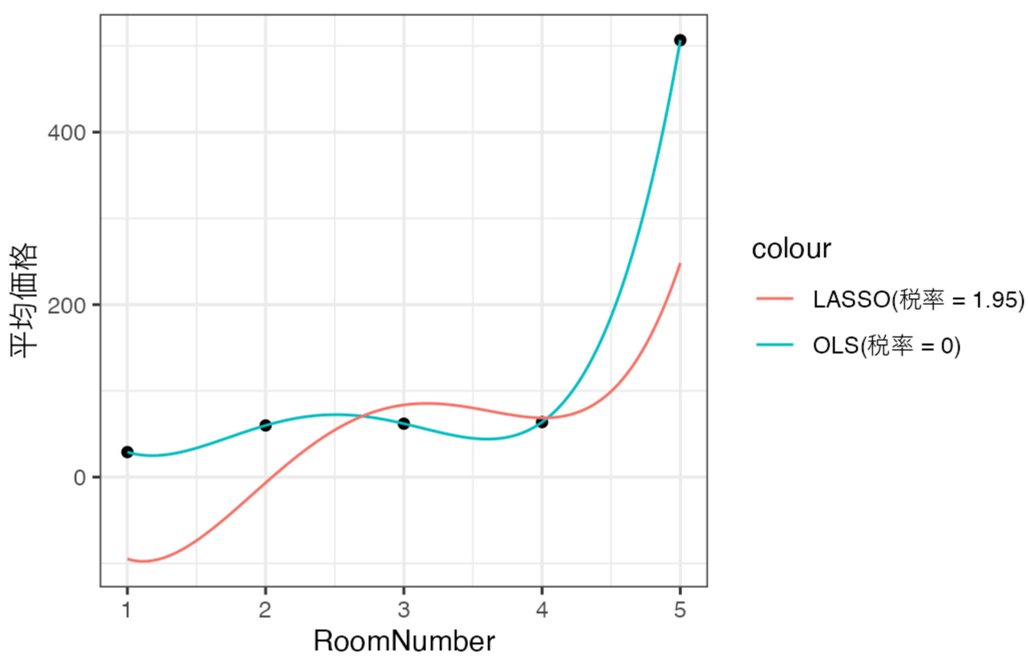
2.10 例: OLS

```
lm(Price ~ Size + Tenure + Distance + District, data)
```

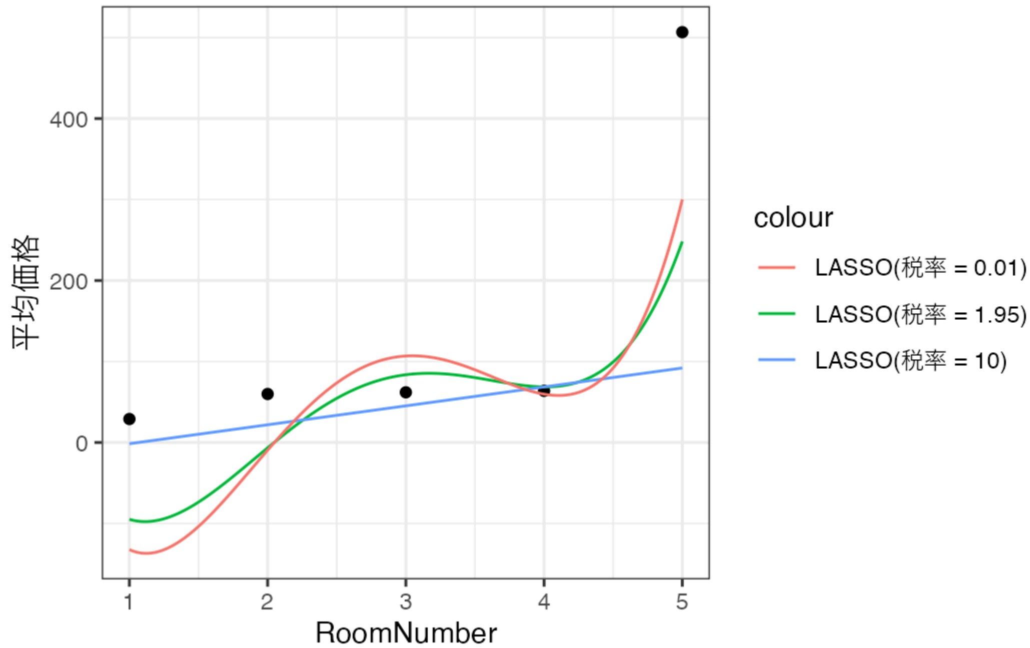
```
Call:
lm(formula = Price ~ Size + Tenure + Distance + District, data = data)

Coefficients:
     (Intercept)                 Size               Tenure             Distance  
          9.9762               1.2373              -0.6782              -1.0522  
  District中央区     District中野区       District北区   District千代田区  
          7.4307               2.1518             -11.3431              28.5354  
  District台東区     District品川区     District大田区     District文京区  
         -2.3232               5.8732              -5.3455               5.4088  
  District新宿区     District杉並区     District板橋区   District江戸川区  
          8.6121               1.7546             -13.2401             -25.5362  
  District江東区     District渋谷区       District港区     District目黒区  
         -8.0226              24.2997              41.3445              15.7670  
  District練馬区     District荒川区     District葛飾区     District豊島区  
        -12.8446             -16.9899             -20.9135               1.7057  
  District足立区     District墨田区  
        -21.9901              -6.8787  
```

2.11 例



2.12 例



2.13 LASSO の性質: 変数選択

- LASSO の重要な性質は変数選択
  - モデルから変数を除外する
- 例: モデル

  $$Price \simeq \beta_0 + \beta_1 Size + \beta_2 Tenure$$

  から、Tenure を除外する

  - $Price \simeq \beta_0 + \beta_1 Size + \underbrace{\beta_2}_{=0} Tenure$

2.14 LASSO の性質

- 罰則の定式化より、β の推定値が厳密に 0 となりうる
  - 該当する変数が除外される
- OLS など他の推定方法では、一般に厳密に 0 とはならない
  - 変数が除外されない
    - 除外したいのであれば、人間が推定前に除外しないといけない
- LASSO は、“データ主導の変数選択”を行う

2.15 Takeaway

- 一般に、OLS は複雑なモデルの推定が苦手
  - 過剰適合を避けるために、多くの事例数を要求
    - 経験則として、β の数が事例数の 1/3 を超えると予測性能が悪化しやすい
- 機械学習 = 複雑なモデルを、過剰適合を避けながら推定する方法を提案
  - LASSO は代表格