LinearModel

機械学習の経済学への応用

川田恵介

Non-Penalized Linear Model

- スムーズな予想木と並ぶ人気モデル
  - NonPenalized Empirical Risk Minimization (OLS など): 漸近性質が" ほぼほぼ" 解明されている
  - Penalized Empirical Risk Minimization (LASSO などなど): そもそもの X が多いデータにおいて現実的な選択肢

Linear Model

- "線形" モデルを事前に設定

$$f(X) = \beta_0 + \beta_1 X_1 + .. + \beta_L X_L$$

- 非線形モデルも設定可能

$$f(X) = \beta_0 + \beta_1 X + .. + \beta_L X^L$$

OLS

- Empirical Risk Minimization として、線形モデルを推定

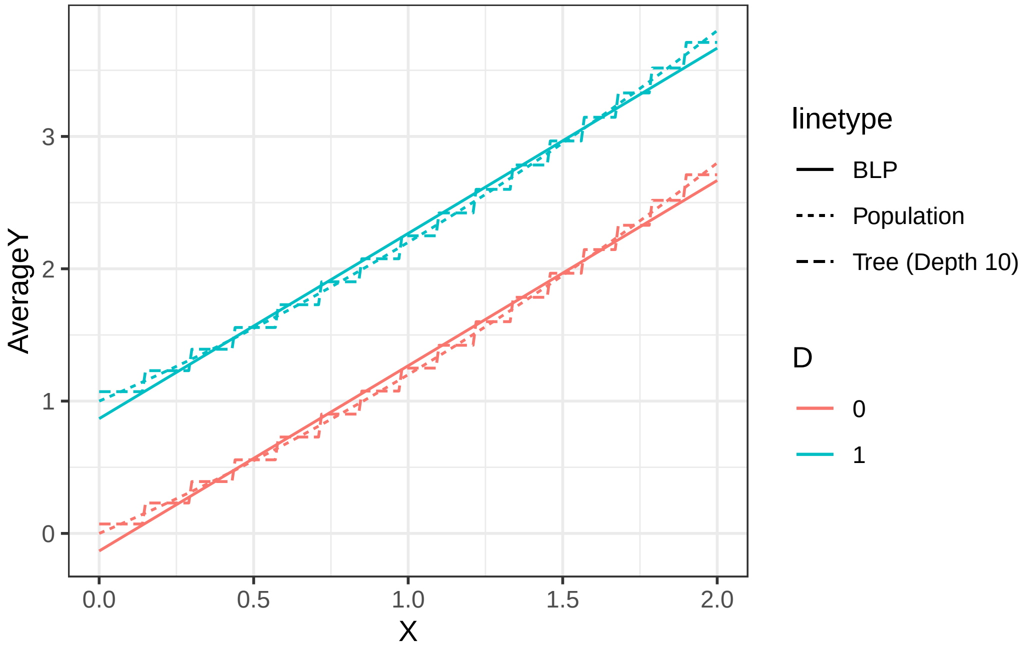
$$\min \sum_i (Y_i - f(X_i))^2$$

- 解釈:: 母集団における Risk(MSE) を最小化する線形近似 f_P の" 優れた" 推定量

$$\min_{\beta_0,..,\beta_L} E[(\mu_Y(X) - f_P(X))^2]$$

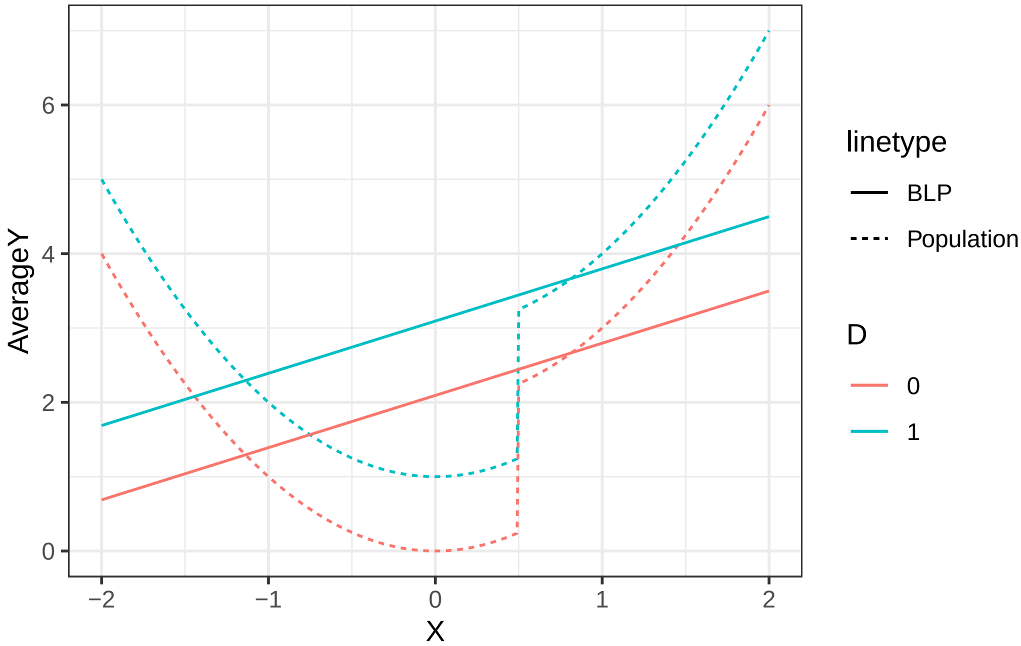
例: Approximation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X$



例: Approximation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X$



モデルの複雑化

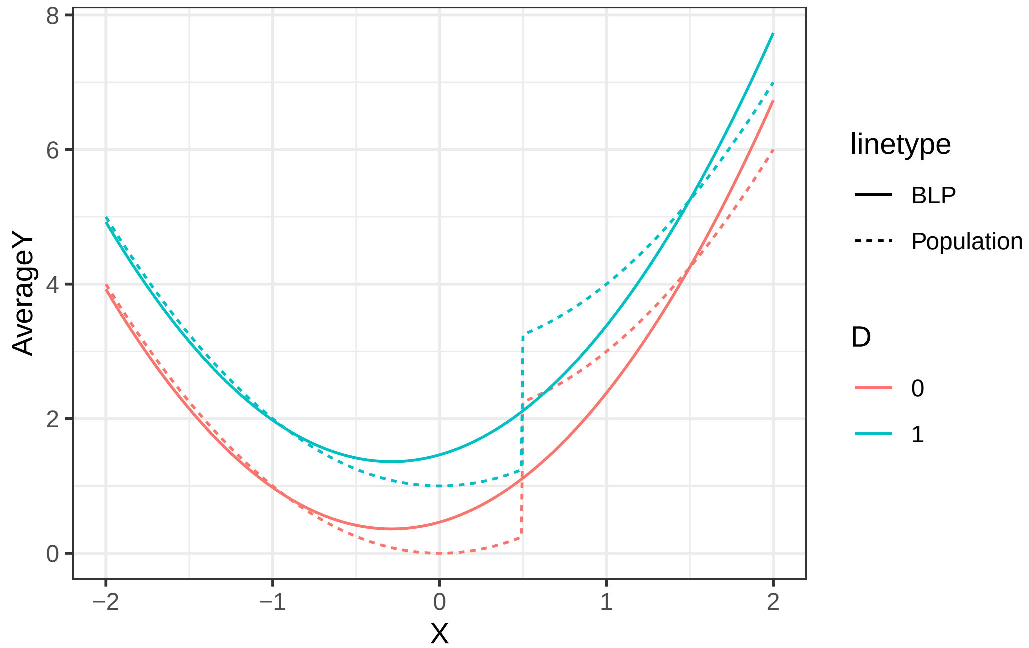
- 複雑なモデルは容易に想定可能
- 2 次モデル

$$f(X, D) = \beta_0 + \beta_D D + \beta_1 X + \beta_2 X^2$$

- "線形モデル" $= \beta_2 = 0$ という特殊ケース

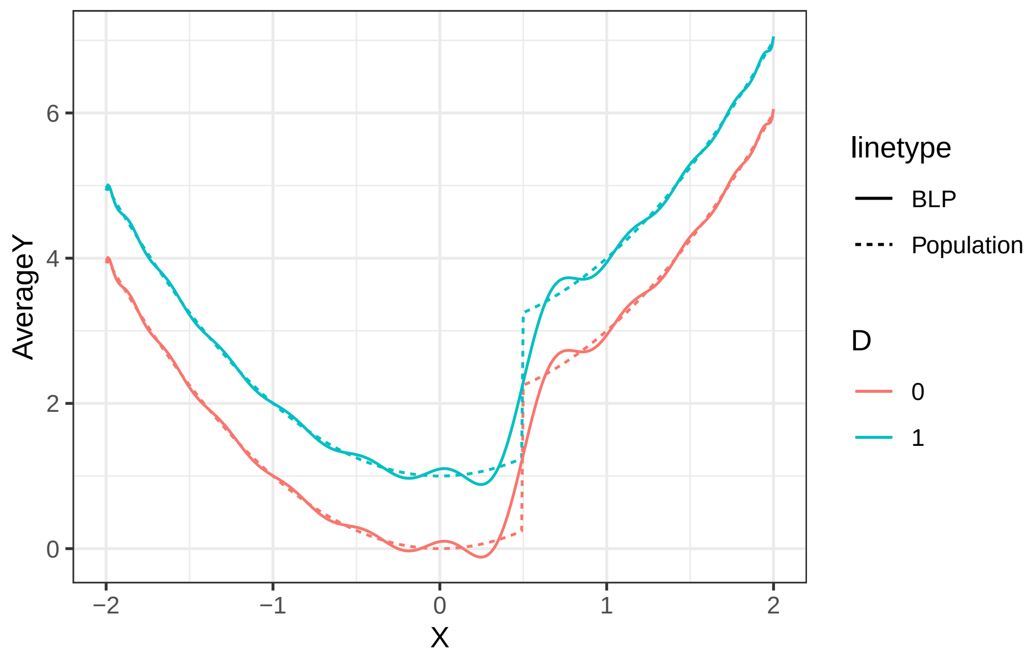
例 Approximation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X + \beta_2 X^2$



例 Approximation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X + .. + \beta_{25} X^{25}$



復習

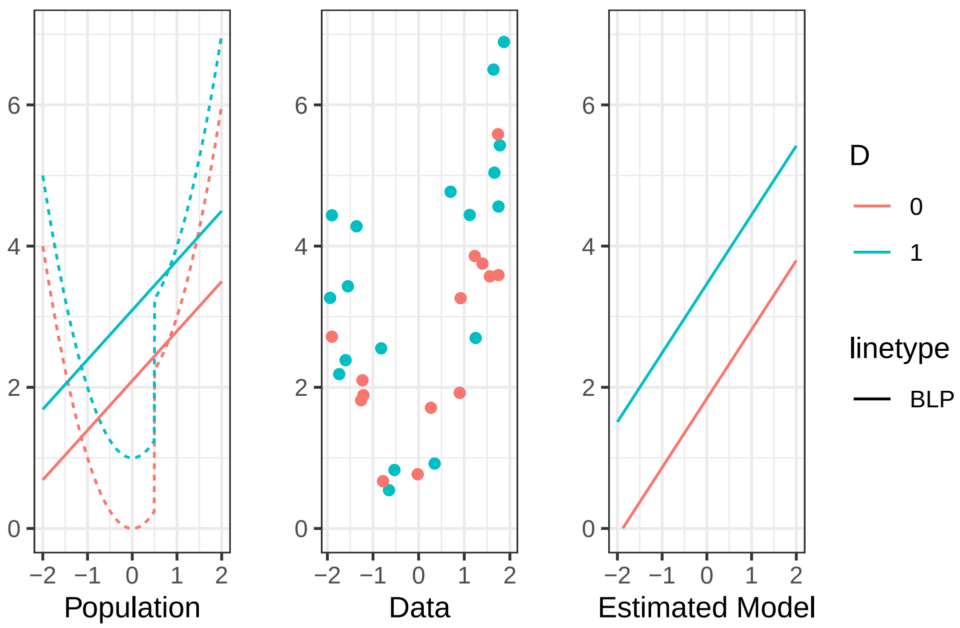
- 予測誤差の分解

$$Y - f(X) = \underbrace{Y - \mu_Y(X)}_{IrreducibleError}$$

$$+ \underbrace{\mu_Y(X) - f_\infty(X)}_{ApproximationError} + \underbrace{f_\infty(X) - f(X)}_{EstimationError}$$

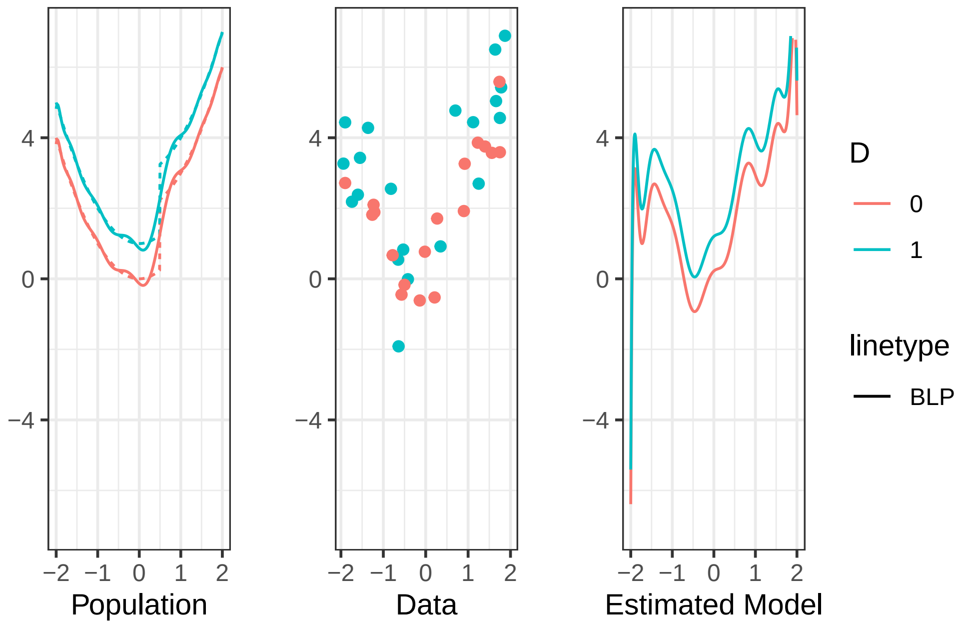
例 Estimation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X$



例 Estimation Error

- $f(X, D) = \beta_0 + \beta_D D + \beta_1 X + .. + \beta_{25} X^{25}$



まとめ

- “適度” な複雑性をもつモデルを” 事前” 設定できれば、OLS や最尤法、ベイズ推定は極めて実用的
  - 社会科学では (おそらく BioMedical Science でも) 困難
- 複雑すぎると EstimationError、単純すぎると ApproimationError が深刻
  - 古典的な予測モデルは、単純すぎる場合が多いとも (Breiman 2001)

Penalized Linear Model

- 複雑すぎる線形モデルからスタート
- 複雑性へのペナルティーをつけて推定

LASSO

- 以下の Penalized Empirical Minimization の解

$$\min \sum_i (Y_i - f(X_i))^2 + \lambda[|\beta_1| + .. + |\beta_L|]$$

- λ は CrossValidation で決定
  - 代替: AIC で決定 (gamlr パッケージ)

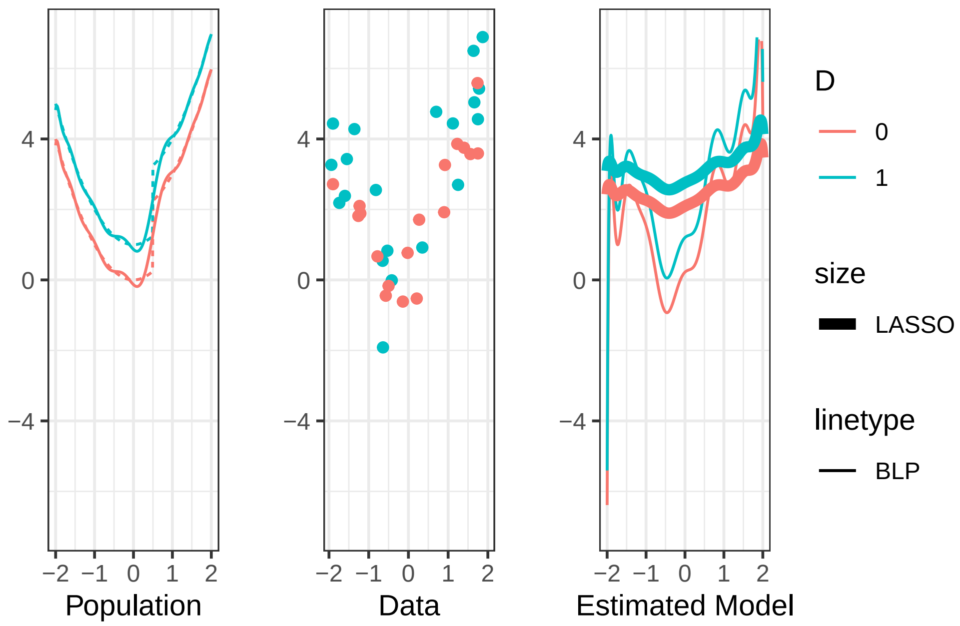
解釈: 条件付き最適化

- 以下と同値

$$\min \sum_i (Y_i - f(X_i))^2 \ s.t. \ |\beta_1| + .. + |\beta_L| \le A$$

- $A \to 0(\lambda \to \infty)$ ならば, $f(X) \to$ サンプル平均
- $A \to \infty(\lambda \to 0)$ ならば, $f(X) \to$ OLS
- 一般に、サンプル平均と OLS の間

例



係数値

```
17 x 1 sparse Matrix of class "dgCMatrix"
                      s1
(Intercept)    2.7578375
poly(X, 15)1   5.9612713
poly(X, 15)2   7.8721161
poly(X, 15)3   .
```

```
poly(X, 15)4  .
poly(X, 15)5  0.8987052
poly(X, 15)6  .
poly(X, 15)7  .
poly(X, 15)8  .
poly(X, 15)9  .
poly(X, 15)10 .
poly(X, 15)11 .
poly(X, 15)12 .
poly(X, 15)13 .
poly(X, 15)14 .
poly(X, 15)15 .
D             0.1884852
```

Sparcity

- OLS: X の数 $>$ サンプルサイズとなると推定不可能
- Penalized Linear model や Tree 系は、可能
  - LASSO や Tree 系は、係数値の一部を厳密に “0” として推定
  - ノイズ的 X が大量に含まれていることが予想される場合、特に重要
- RandomForest などと比べても、X の数が多い時は LASSO に比較優位
  - 目安: サンプルサイズ/4 > X の数 (Taddy 2019)

発展: ElasticNet

- α , λ を CrossValidation で決定

$$\min \sum_i (Y_i - f(X_i))^2$$

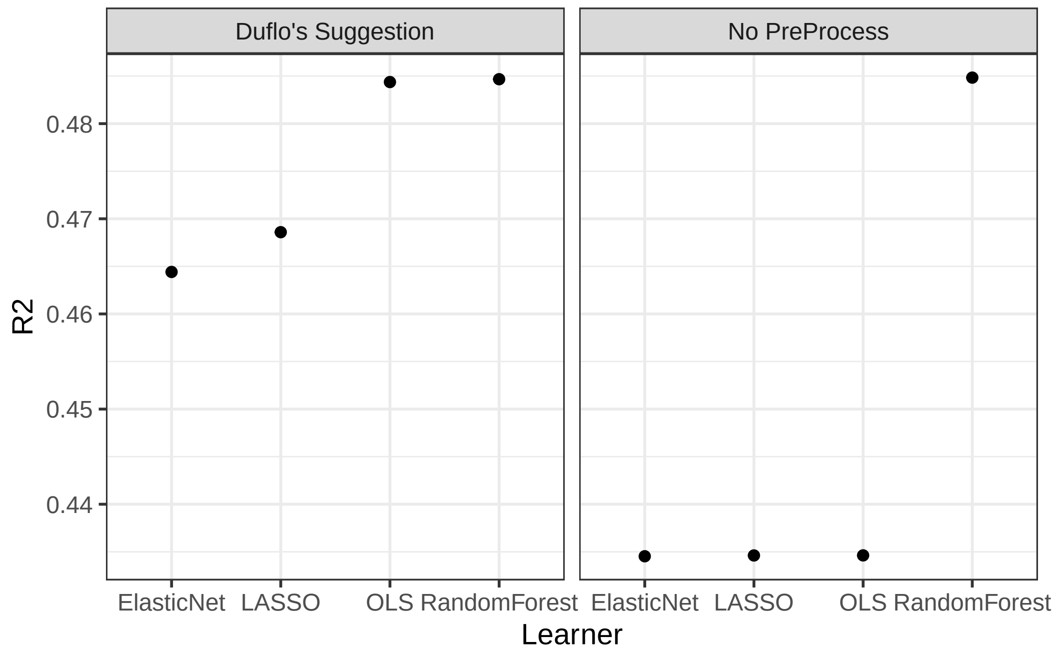
$$+ \underbrace{\lambda\alpha[|\beta_1| + .. + |\beta_L|]}_{L1ペナルティー(LASSO)}$$

$$+ \underbrace{\lambda(1-\alpha)[\beta_1^2 + .. + \beta_L^2]}_{L2ペナルティー(Ridge)}$$

Practice

- Tree 系とは異なり、完全な Nonparametric モデルの推定ではない
  - 十分に複雑なモデル (OverParametric モデル) からスタートしたい
- Duflo's suggestion
  - 全ての X を投入
  - 連続変数については二乗項、Option として交差項
  - 欠損値は欠損ダミーを作成した後に、0 を補完
  - Variation がない変数や完全な多重共線を起こしている変数は除外

実例



まとめ

- 母平均の複雑すぎるパラメトリックモデルからスタートし、ペナルティー項を用いて単純化
- 決めうちの深い予測木 + Pruning と同じ戦略

Reference

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures." *Statistical Science* 16 (3): 199–231. https://doi.org/10.1214/ss/1009213726.

Taddy, Matt. 2019. *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. McGraw Hill Professional.