PSY9511: Seminar 3

Regularization and variable selection

Esten H. Leonardsen

11.09.24

Shrinkage

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function

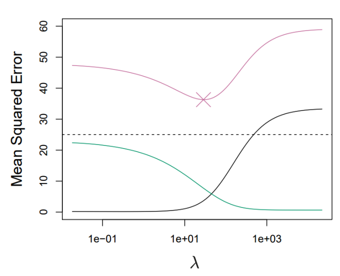
$$mse = bias^2 + variance + irreducible\ error$$

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function



mse=*bias*^{2}+*variance*+*irreducible error*

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function
2. A little more bias $\implies$ A lot less variance

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function
2. A little more bias $\implies$ A lot less variance

$$y \sim \beta_0 + \beta_1 x_1 + \overset{\uparrow}{\beta_2 x_2} + \beta_3 x_3 + \underset{\downarrow}{\beta_4 x_4} + \beta_5 x_5 + \beta_6 x_6$$

Shrinkage: Outline

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$
$$\beta_n \to 0$$

1. $\beta_1 = 0 \implies$ One less degree of freedom in our function
2. A little more bias $\implies$ A lot less variance
3. Parameters depend on eachother $\implies$ Fewer degrees of freedom

Shrinkage: Ridge regression

$$loss_{mse} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2$$

Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} \beta_j^2$$

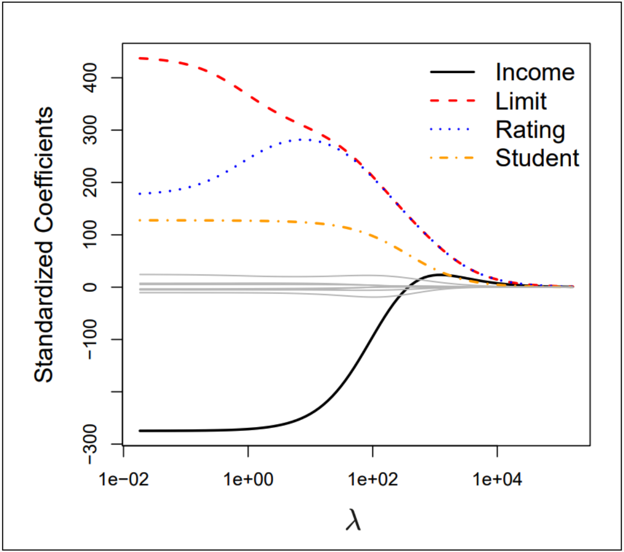
Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 \boxed{+\lambda \sum_{j=0}^{p} \beta_j^2}$$

$$\Downarrow$$

$$\lambda \rightarrow \infty \Rightarrow \beta \rightarrow 0$$

Shrinkage: Ridge regression

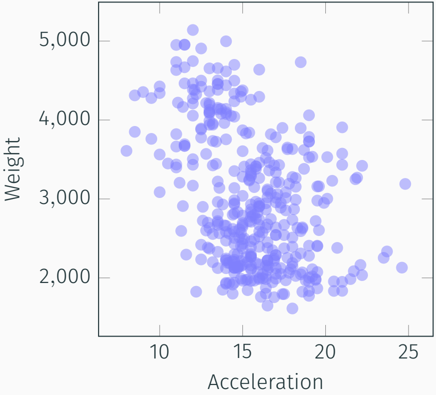


Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n}\left(y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right)^2 + \lambda \sum_{j=0}^{p}\beta_j^2$$

$$y \sim \beta 0 + \beta_1 x_1 + \beta_2 x_2, x_1 \in [0,1],_2 \in [0,1000]$$

Shrinkage: Feature standardization



Shrinkage: Feature standardization

z-score standardization

Shrinkage: Feature standardization

z-score standardization

$$x = \frac{x - \mu_X}{\sigma_X}$$

Shrinkage: Feature standardization

z-score standardization

$$x = \frac{x - \mu_x}{\sigma_x}$$

In[1]:

```
for col in predictors:
    print(f'{col}: {np.mean(df[col]):.2f} ({np.std(df[col]):.2f})')

# z-score standardization
for col in predictors:
    df[col] = (df[col] - np.mean(df[col])) / np.std(df[col])

for col in predictors:
    print(f'{col} after: {np.mean(df[col]):.2f} ({np.std(df[col]):.2f})')
```

Out[1]:

```
cylinders: 5.47 (1.70)
displacement: 194.41 (104.51)
horsepower: 104.47 (38.44)
cylinders after: -0.00 (1.00)
displacement after: -0.00 (1.00)
horsepower after: -0.00 (1.00)
```

Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} \beta_j^2$$

Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} \beta_j^2$$

Blackboard!

Shrinkage: Ridge regression

$$loss_{ridge} = \sum_{i=0}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} \beta_j^2$$

http://localhost:8888/notebooks/notebooks/Live%20coding.ipynb