Multiple Linear Regression

Linear Regression is an Example of a Linear Model

Input instance – feature vector: $\boldsymbol{x} = (x_0, x_1, \ldots, x_n)$

Predicted output: $\hat{y} = \widehat{w_0}x_0 + \widehat{w_1}x_1 + \cdots \widehat{w_n}x_n + \hat{b}$

Parameters to estimate:

- $\widehat{\boldsymbol{w}} = (\widehat{w_0}, \cdots, \widehat{w_n})$: feature weights/ model coefficients
- $\widehat{\boldsymbol{b}}$: constant bias term / intercept

Linear Models

- **A linear model is a <u>sum of weighted variables</u> that predicts a target output value given an input data instance.** **<u>Example</u>:** ***predicting housing prices***
  - *House features: taxes per year (X_{TAX}), age in years (X_{AGE})*

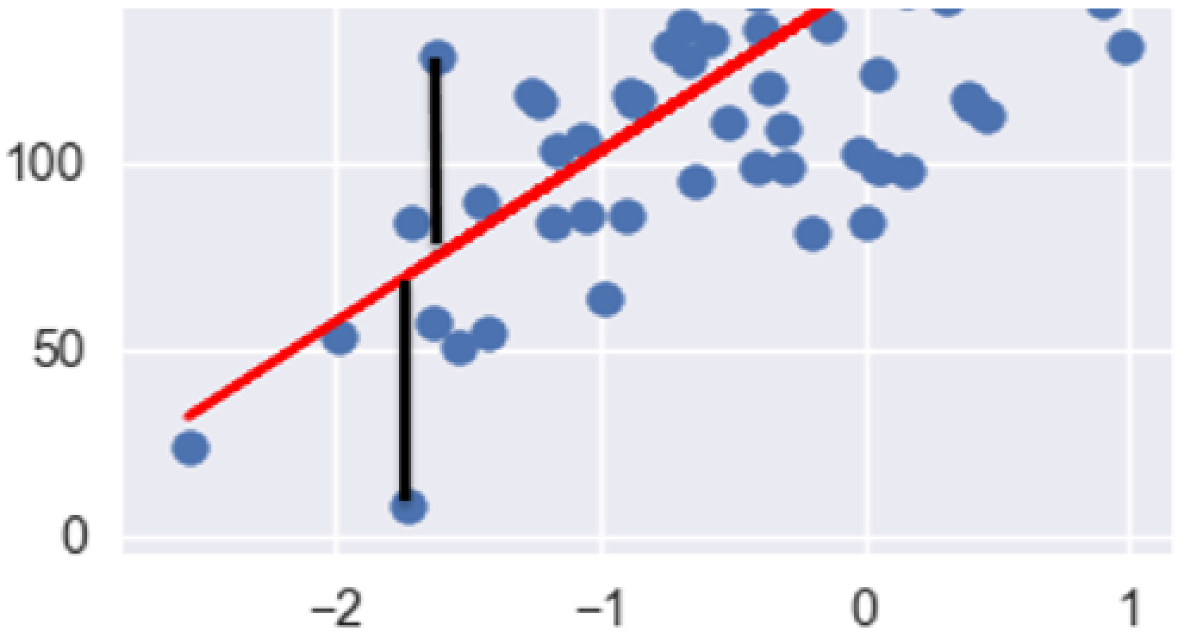
$$\widehat{Y_{PRICE}} = 212000 + 109\, X_{TAX} - 2000\, X_{AGE}$$

  - *A house with feature values (X_{TAX}, X_{AGE}) of (10000, 75) would have a predicted selling price of:*

$$\widehat{Y_{PRICE}} = 212000 + 109 \cdot \mathbf{10000} - 2000 \cdot \mathbf{75} = \mathbf{1{,}152{,}000}$$

Least-Squares Linear Regression ("Ordinary least-squares")

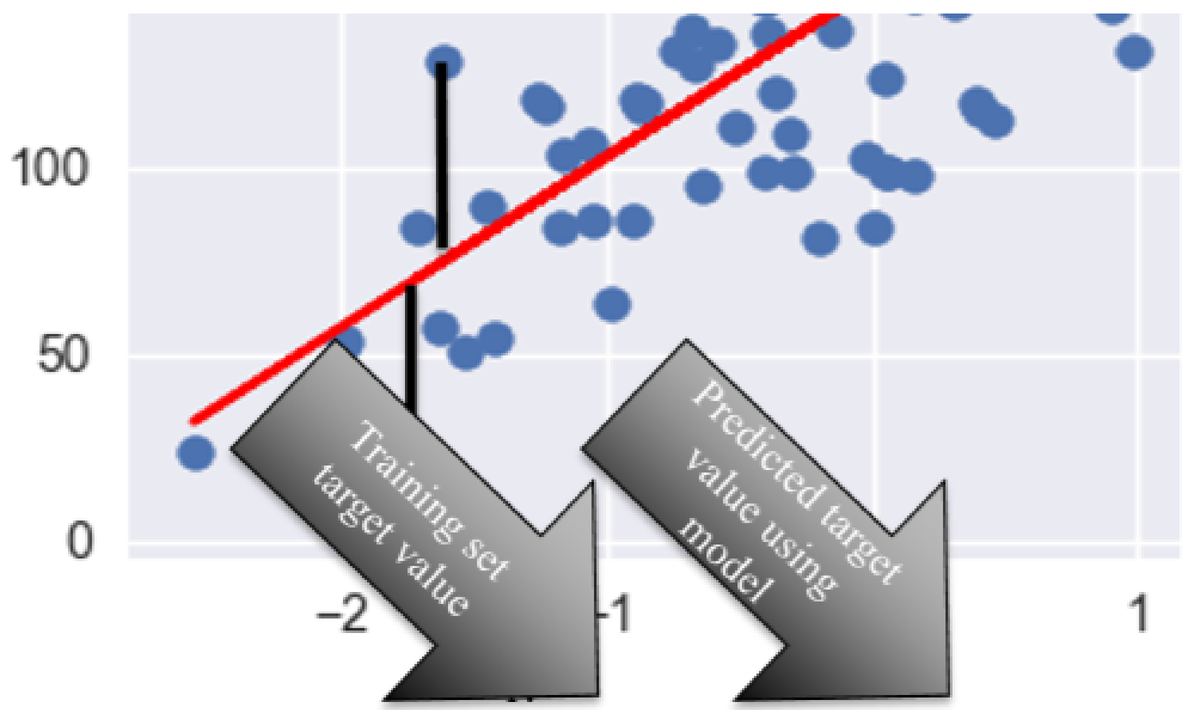
- Finds *w* and *b* that minimizes the <u>sum of squared differences</u> (RSS) over the training data between predicted target and actual target values.
- a.k.a. mean squared error of the linear model
- No parameters to control model complexity.



$$RSS(\boldsymbol{w}, b) = \sum_{\{i=1\}}^{N} (\boldsymbol{y}_i - (\boldsymbol{w} \cdot \boldsymbol{x}_i + b))^2$$

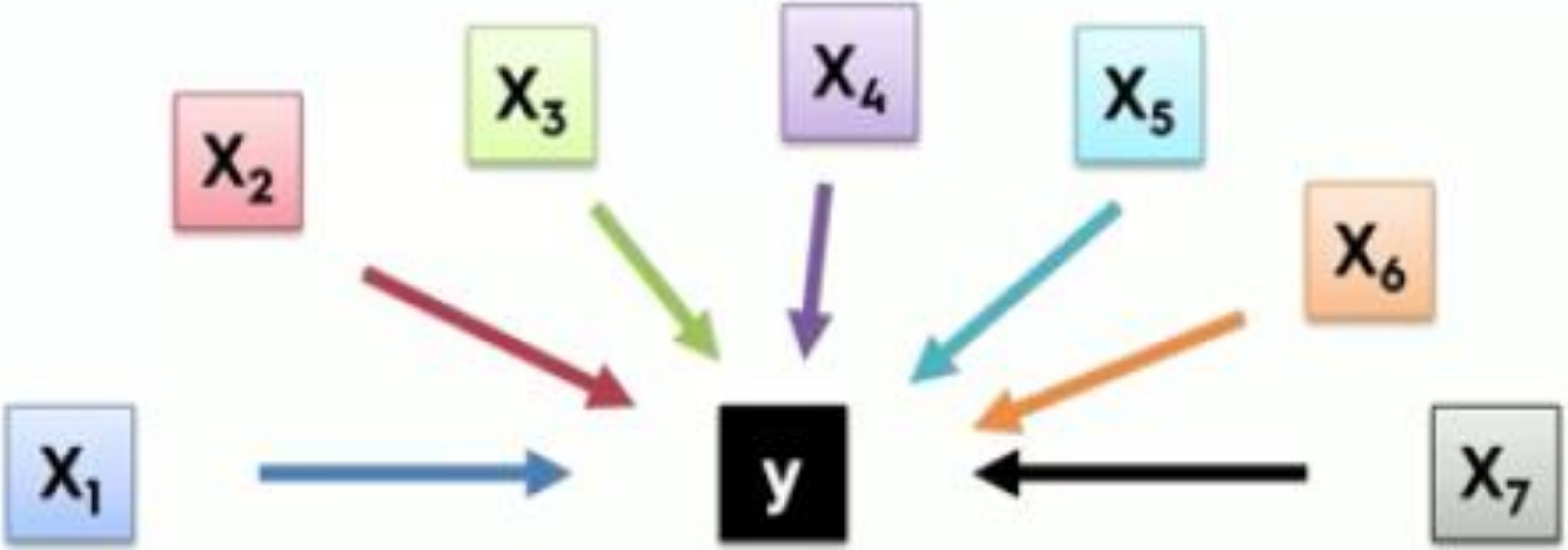
Least-Squares Linear Regression ("Ordinary Least-Squares")

- Finds *w* and *b* that minimizes the <u>sum of squared differences</u> (RSS) over the training data between predicted target and actual target values.
- a.k.a. mean squared error of the linear model
- No parameters to control model complexity.

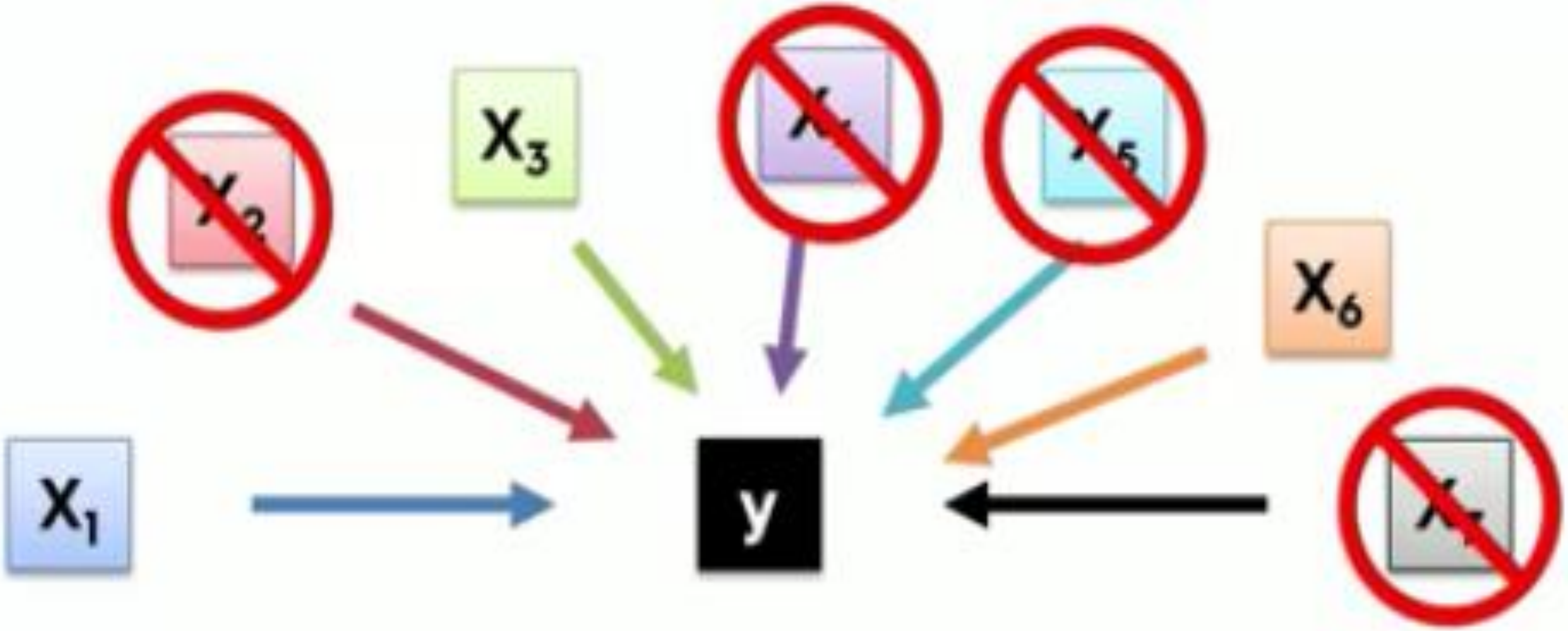


$$RSS(\boldsymbol{w}, b) = \sum_{\{i=1\}} (\boldsymbol{y}_i - (\boldsymbol{w} \cdot \boldsymbol{x}_i + b))^2$$

Building A Model



Building A Model



P-Value

H_0: This is a fair coin
H_1: This is not a fair coin

0.5
0.25
0.12
0.06
0.03
0.01

H_0: This is a fair coin
H_1: This is not a fair coin

P-Value

0.5
0.25
0.12
0.06
0.03
0.01

H_0: This is a fair coin

H_1: This is not a fair coin

P-Value

0.5

0.25

0.12

0.06

$\alpha = 0.05$

0.03

0.01

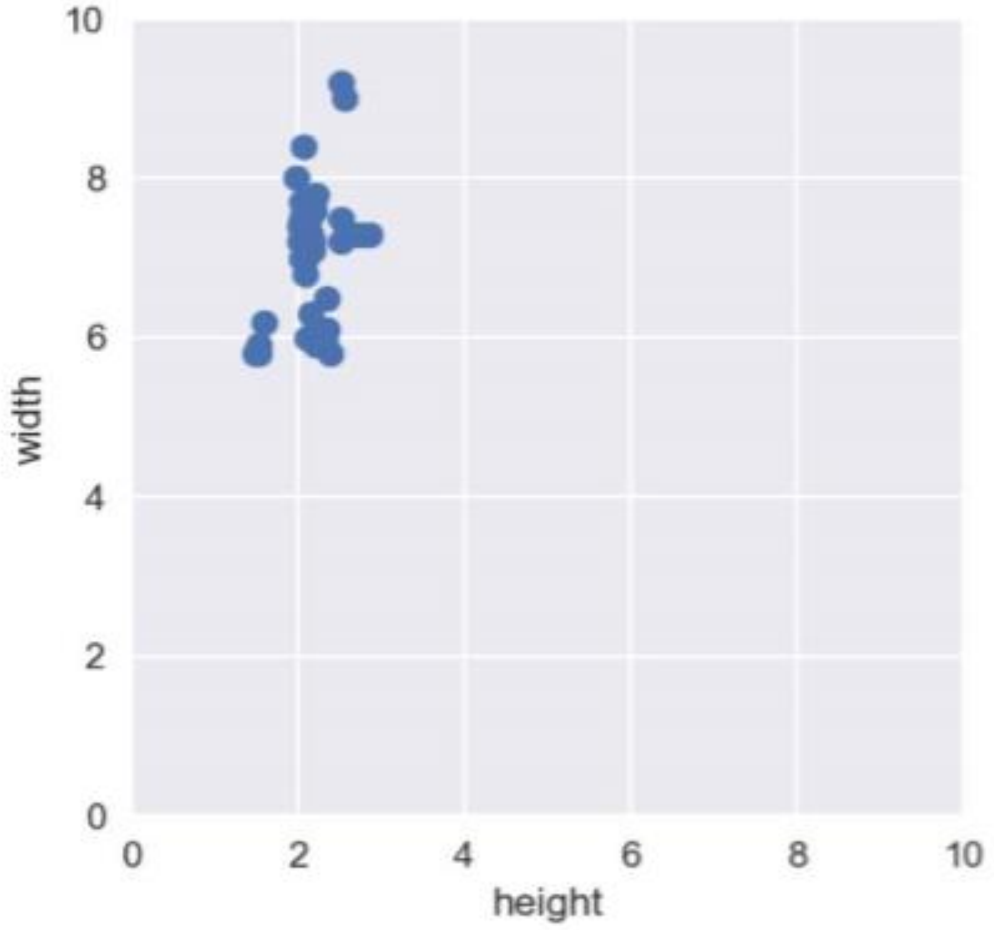
Ridge Regression

- Ridge regression learns *w*, *b* using the same least-squares criterion but adds a penalty for large variations in *w* parameters

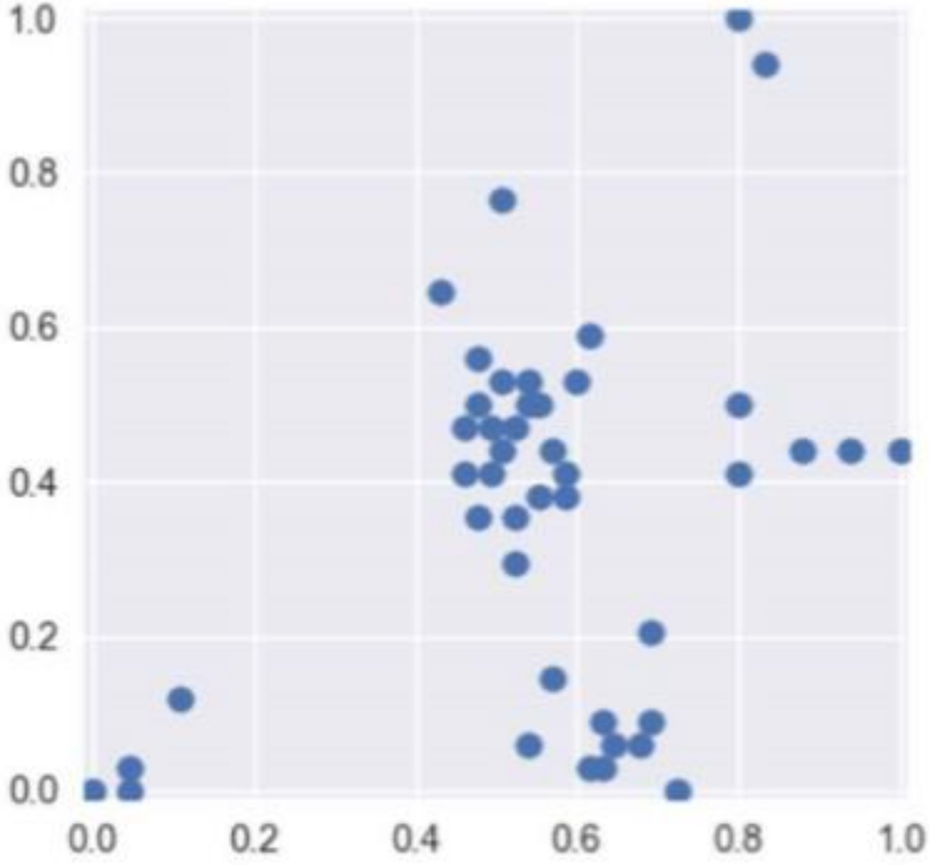
$$RSS_{RIDGE}(\mathbf{w}, b) = \sum_{\{i=1\}}^{N} (\mathbf{y}_i - (\mathbf{w} \cdot \mathbf{x}_i + b))^2 \boxed{+ \alpha \sum_{\{j=1\}}^{p} w_j^2}$$

- Once the parameters are learned, the ridge regression <u>prediction</u> formula is the <u>same</u> as ordinary least-squares.
- The addition of a parameter penalty is called <u>regularization</u>. Regularization prevents overfitting by restricting the model, typically to reduce its complexity.
- Ridge regression uses <u>L2 regularization</u>: minimize sum of squares of *w* entries
- The influence of the regularization term is controlled by the α parameter.
- Higher alpha means more regularization and simpler models.

Feature Normalization with MinMaxScaler



Unnormalized data points



Normalized with MinMaxScaler

Lasso regression is another form of regularized linear regression that uses an <u>L1 regularization</u> penalty for training (instead of ridge's L2 penalty)

- **L1 penalty: Minimize the sum of the <u>absolute values</u> of the coefficients**

$$RSS_{LASSO}(\boldsymbol{w}, b) = \sum_{\{i=1\}}^{N} (\boldsymbol{y}_i - (\boldsymbol{w} \cdot \boldsymbol{x}_i + b))^2 \boxed{+ \alpha \sum_{\{j=1\}}^{p} |w_j|}$$

- **This has the effect of setting parameter weights in *w* to <u>zero</u> for the least influential variables. This is called a <u>sparse</u> solution: a kind of feature selection**
- **The parameter α controls amount of L1 regularization (default = 1.0).**
- **The prediction formula is the same as ordinary least-squares.**
- **When to use ridge vs lasso regression:**
  - *Many small/medium sized effects: use ridge.*
  - *Only a few variables with medium/large effect: use lasso.*