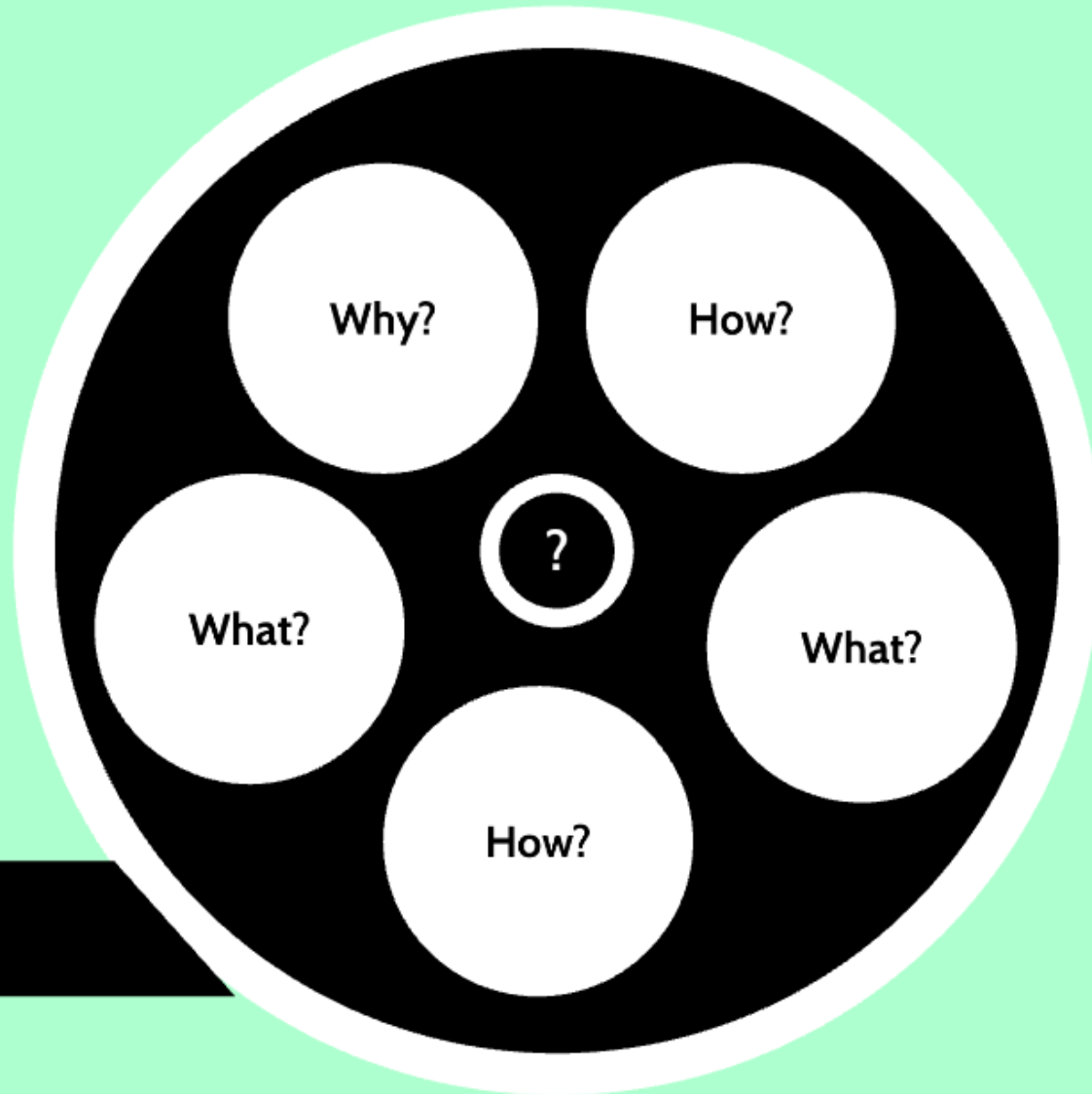


# Predicting the IMDB rating of a movie before its release

Kimia Esmaili  
610398193  
Dr. Arta Jamshidi  
Feb. 2023



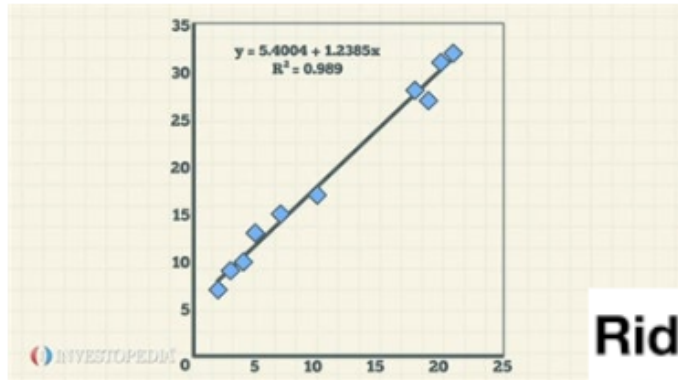
# Why should we care about this topic?

- Multi-million-dollar industry
- IMDB being the most popular and authoritative source for movie
- better decision making for how to make a movie more (commercially) successful

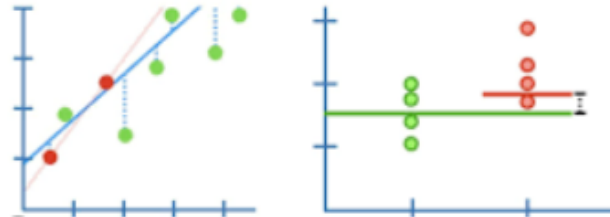


# How are we going to do that(mathematically)?

- Linear Algebra and regression
- $R^2$
- Ridge Regression



**Ridge Regression....**



# What does each tool do?

- Explaining our mathematical tools

**Ridge  
regression**

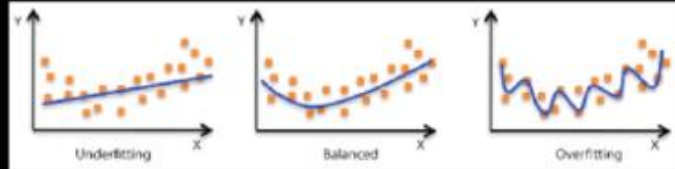
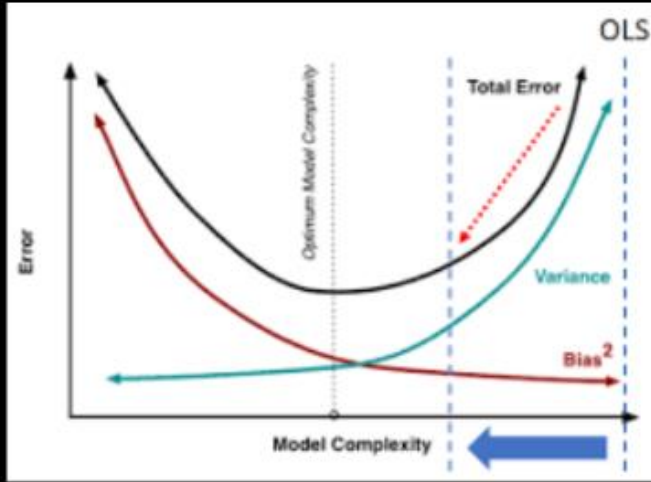
**R-squared**

# Multiple linear regression

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n$$

$$y = X\beta + \epsilon$$

$$\|\epsilon\|_2^2 = \|y - X\beta\|_2^2$$



# OLS

$$\operatorname{argmin} ||y - \hat{y}||_2^2 = \operatorname{argmin} \sum [y_i - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)]^2$$

Ridge

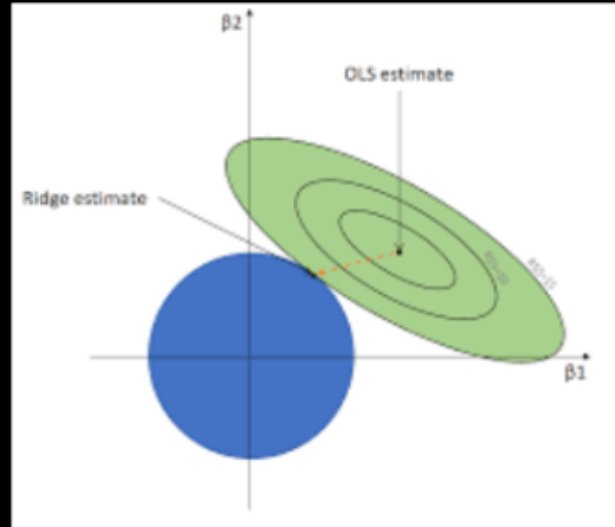
# parameter estimation in ridge regression and its geometrical interpretation

$$\beta_0^2 + \beta_1^2 + \dots + \beta_p^2 \leq C^2$$

$$\|\beta\|_2^2 \leq C^2$$

$$\operatorname{argmin} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

$$\hat{\beta}^{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y$$



# What does each tool do?

- Explaining our mathematical tools

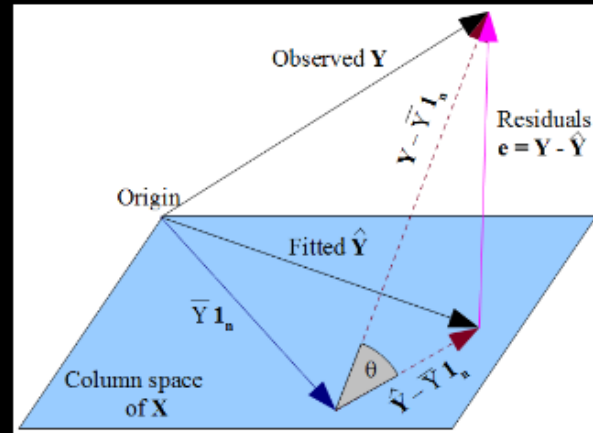
**Ridge  
regression**

**R-squared**



# How the metric works?

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}$$



# How Do we do that in code?

Step 1: Data acquisition & cleaning

Step 2: Models and features

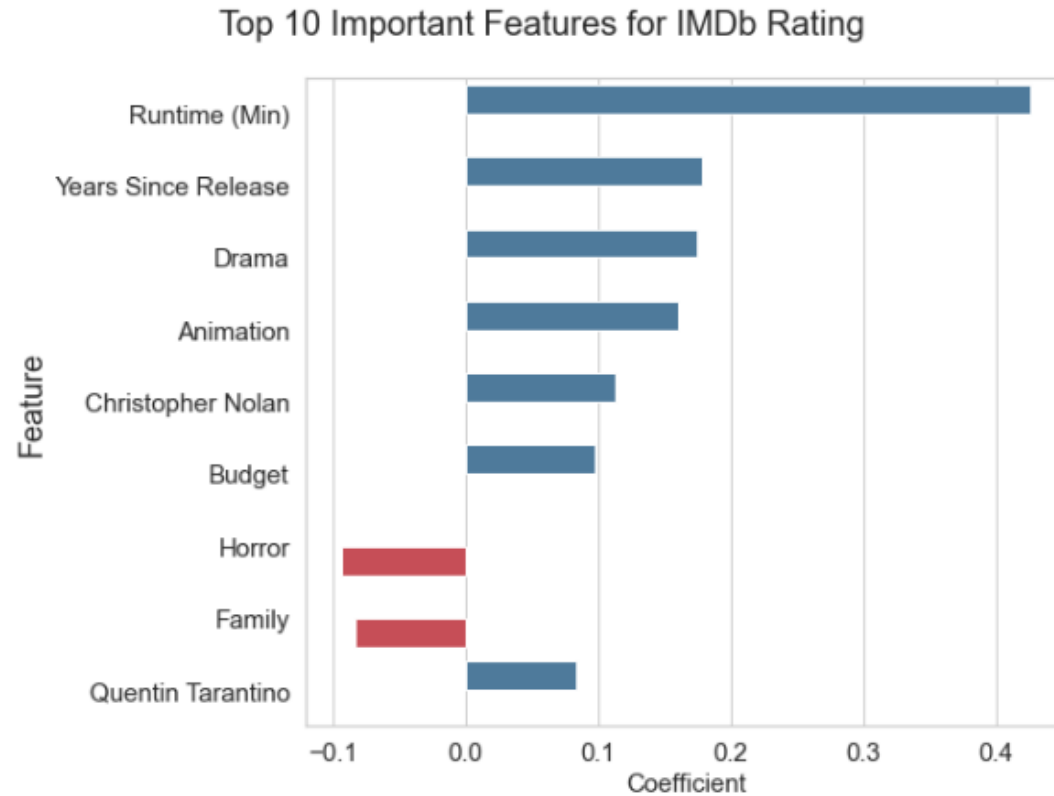
Step 3: Testing and training

Skills and tools:

- Web scraping (requests, HTML, BeautifulSoup)
- Exploratory Data Analysis (EDA) (pandas, numpy)
- Linear regression (scikit-learn)
- Data visualization (seaborn, matplotlib)



# What was the result?



This chart was taken from the the Jupyter notebook of the project

Thank you  
for your attention!



Click the [link](#) for my online Prezi presentation.