Regression problem

Python for AI

Types of learning problems

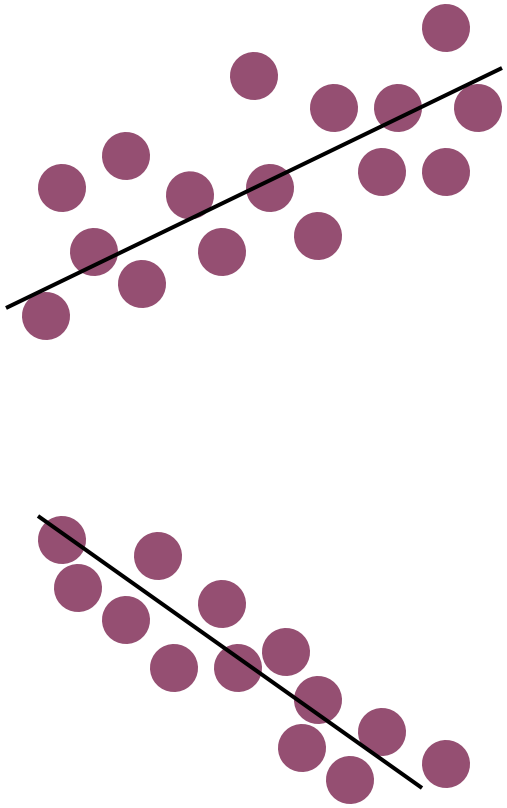
1. Supervised
    a. Classification
    b. Regression
2. Unsupervised
    a. Clustering
    b. Association
3. Reinforcement

Correlation vs. Regression

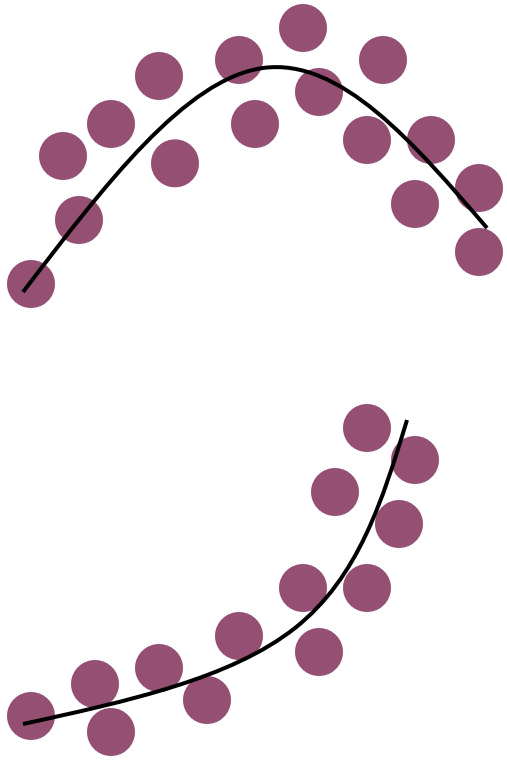
- A scatter plot can be used to show the relationship between two variables
- Correlation analysis is used to measure the strength of the association (linear relationship) between two variables
  - Correlation is only concerned with strength of the relationship
  - No causal effect is implied with correlation
  - Scatter plots
  - Correlation

12.1 Regression Models

Linear relationships

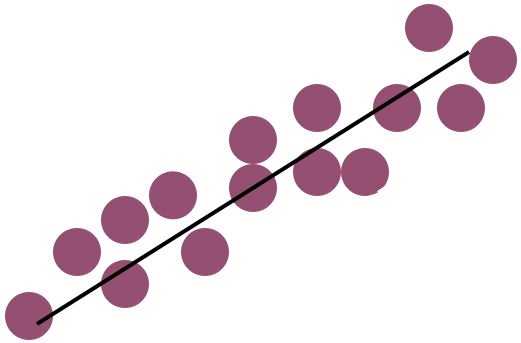


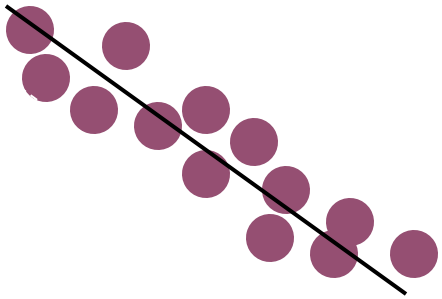
Curvilinear relationships



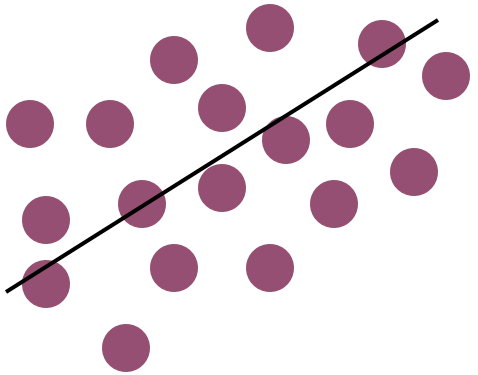
Types of Relationships

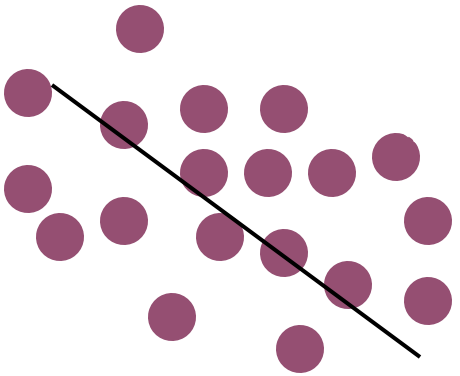
Strong relationships





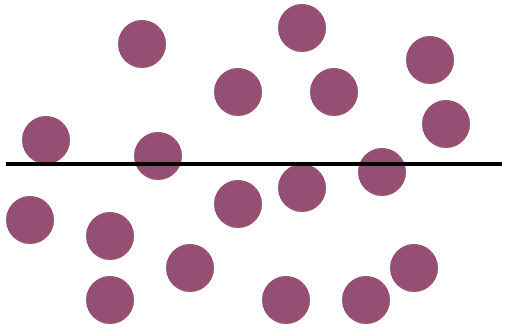
Weak relationships

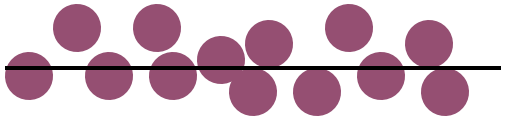




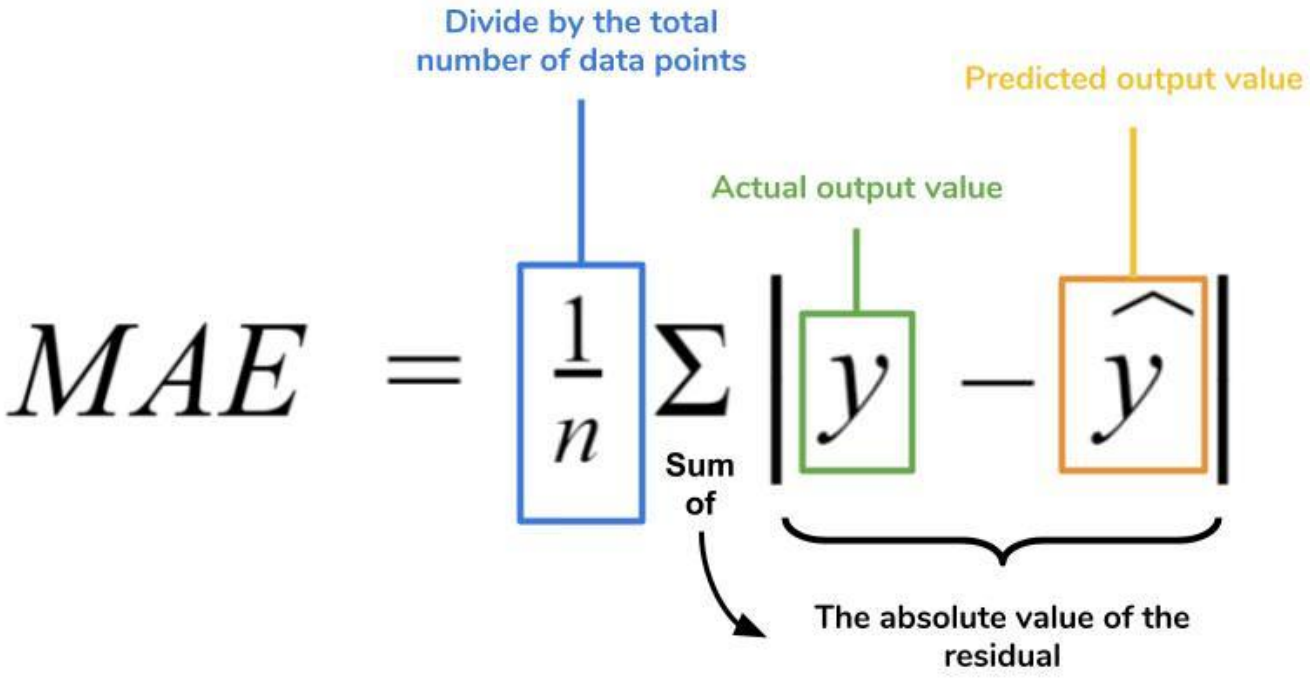
Types of Relationships

No relationship





Metrics


Divide by the total number of data points
Predicted output value
Actual output value
$MAE = \frac{1}{n} \Sigma \left| y - \hat{y} \right|$
Sum of
The absolute value of the residual

Metrics

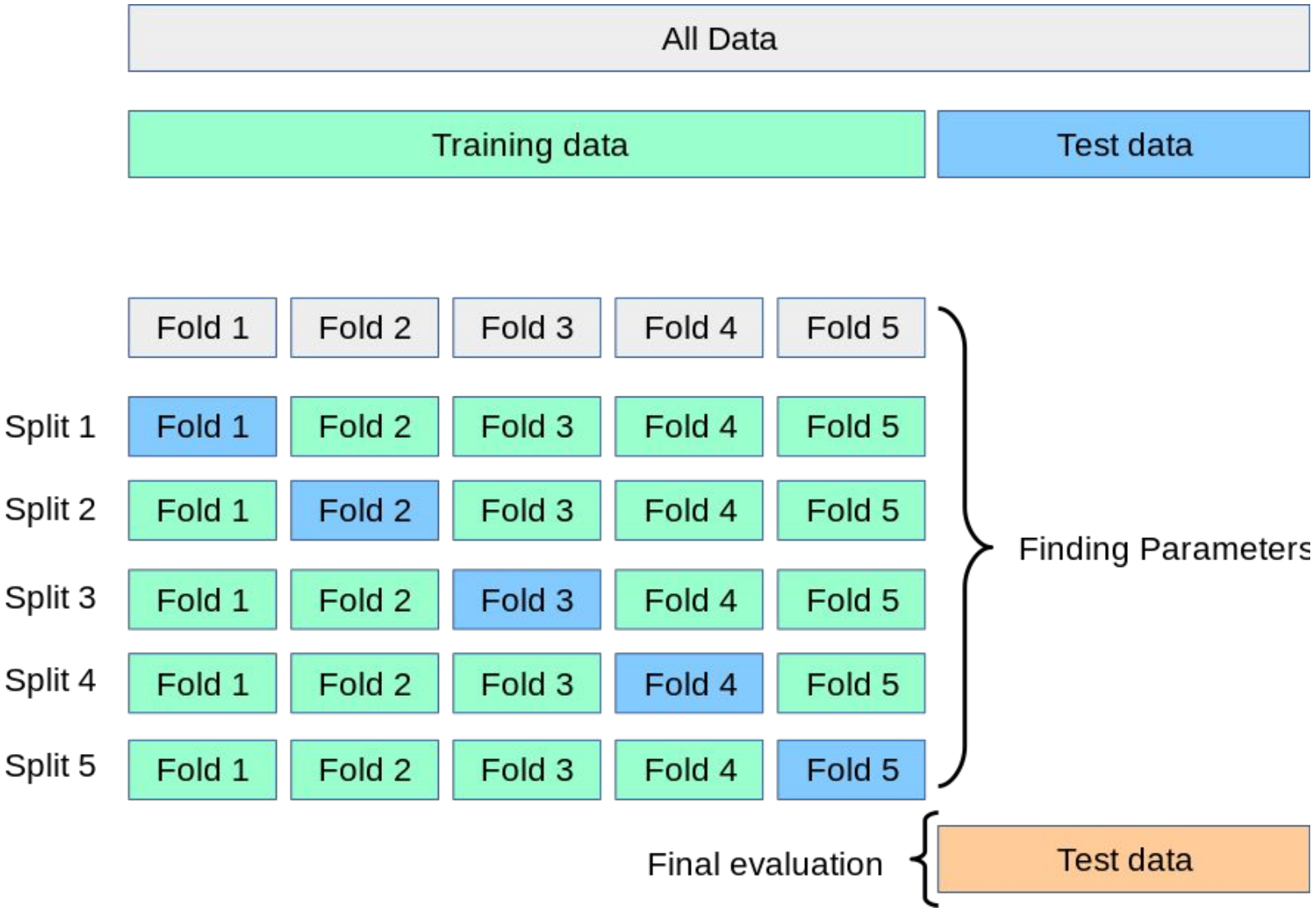
$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \hat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Metrics

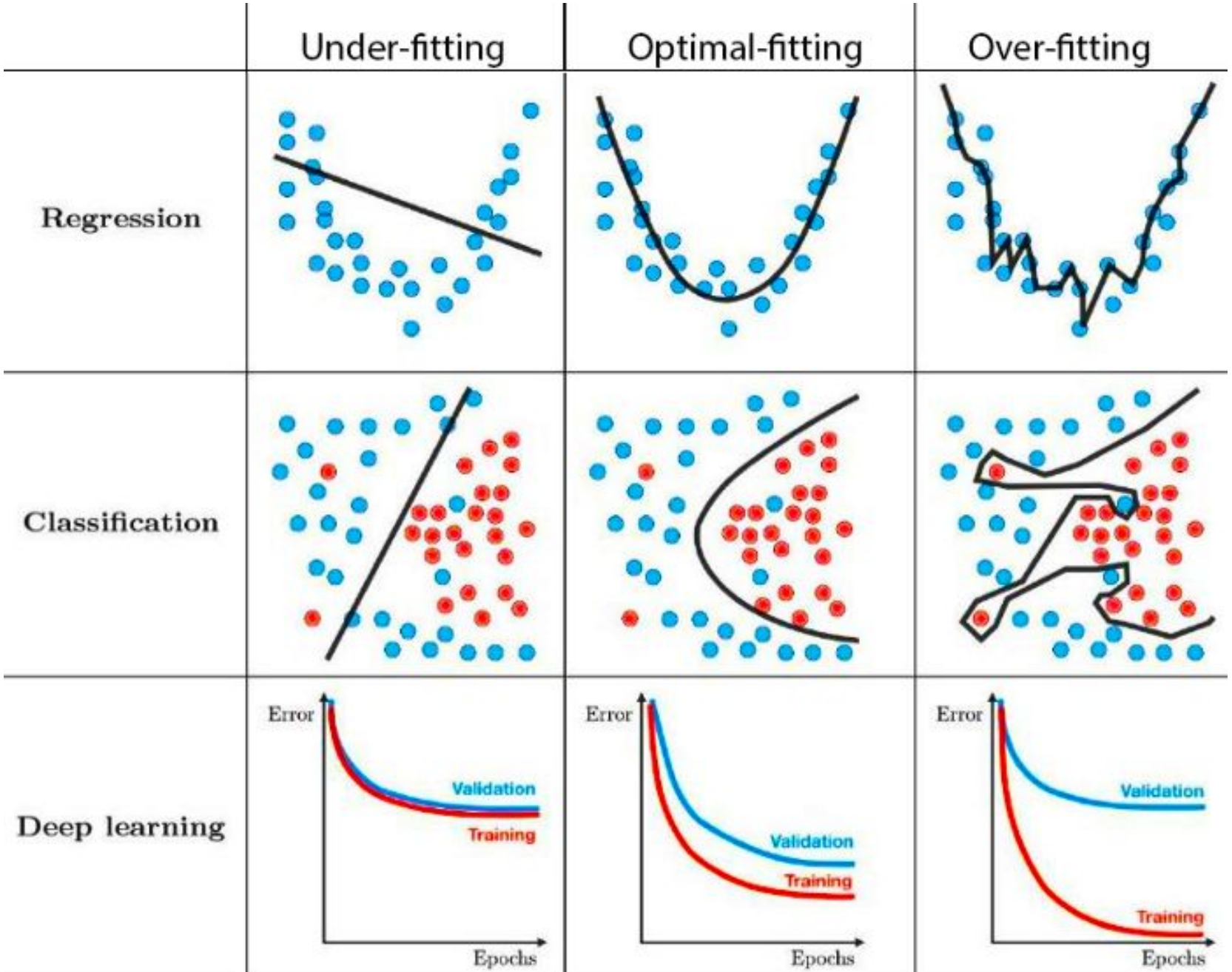
$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

- **0% represents a model that does not explain any of the variation in the response variable around its mean. The mean of the dependent variable predicts the dependent variable as well as the regression model.**
- **100% represents a model that explains all the variation in the response variable around its mean.**

Cross-validation



Challenges: Underfitting & Overfitting



Example and code

- Download code in the classroom
- On class: follow a step by step tutorial