Classification

Kookjin Lee

(kookjin.Lee@asu.edu)

Introduction to classification

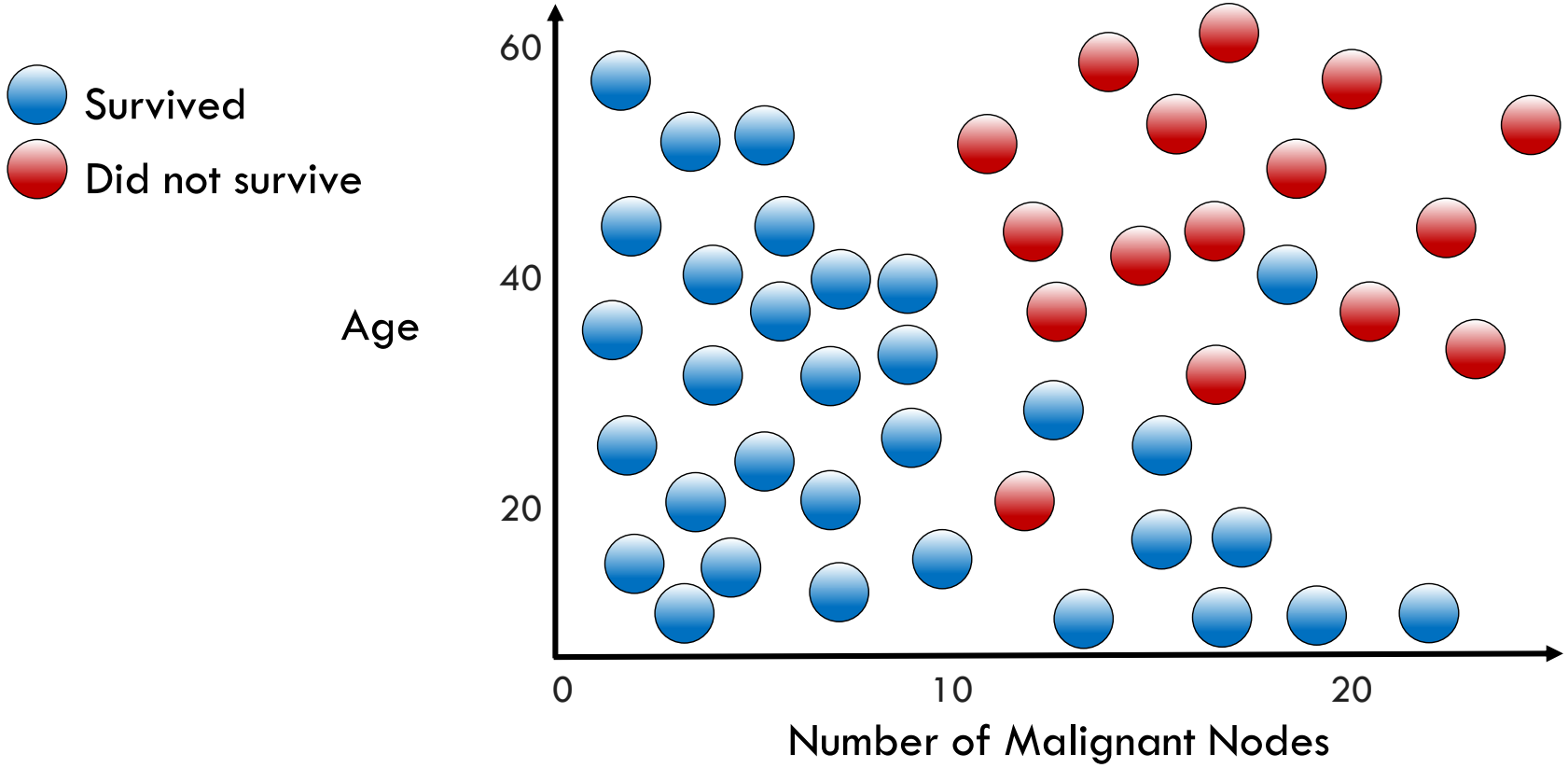
- New task!
- Example:
  A flower shop wants to guess a customer's purchase from similarity to most recent purchase.

What is needed for classification

- Data with
  - Features that can be quantitated
  - Labels that are known ➔ supervised learning
- Method to measure similiarity

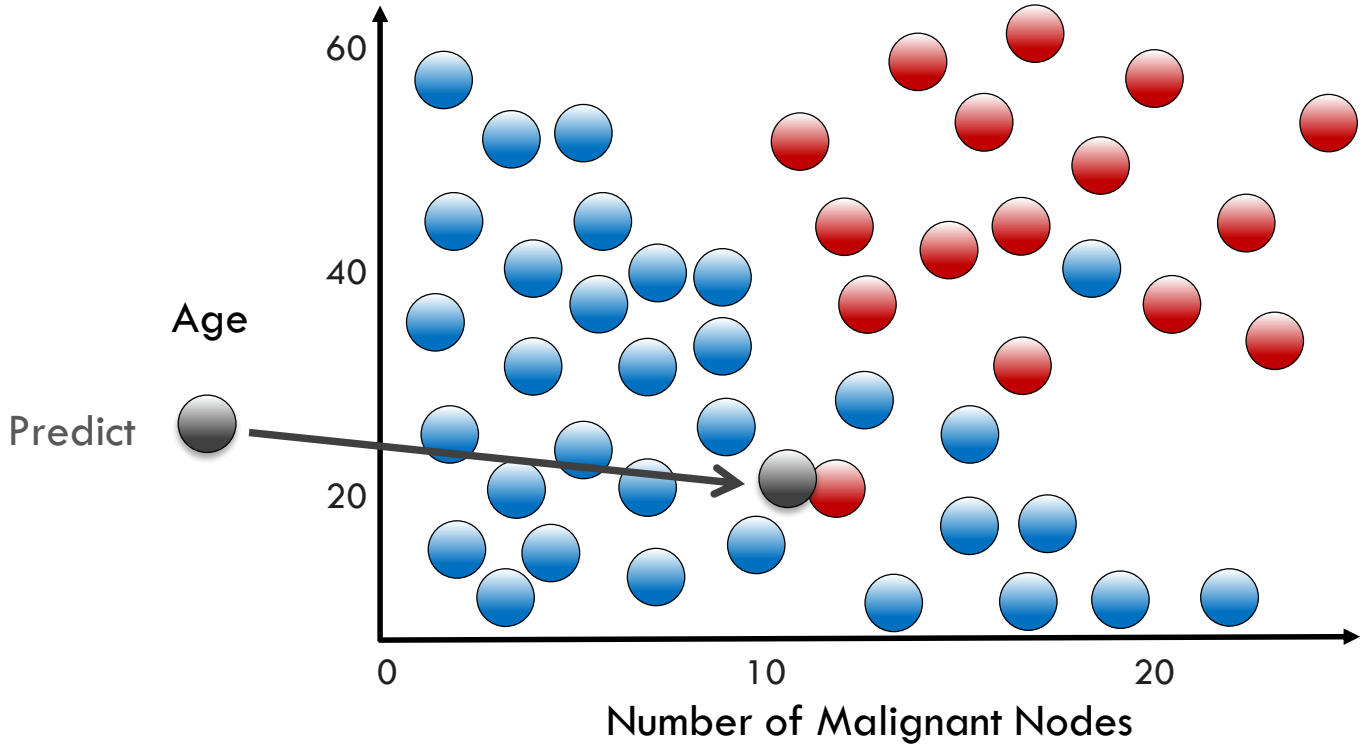
Your first classification algorithm

- K Nearest Neighbors (KNN) classification



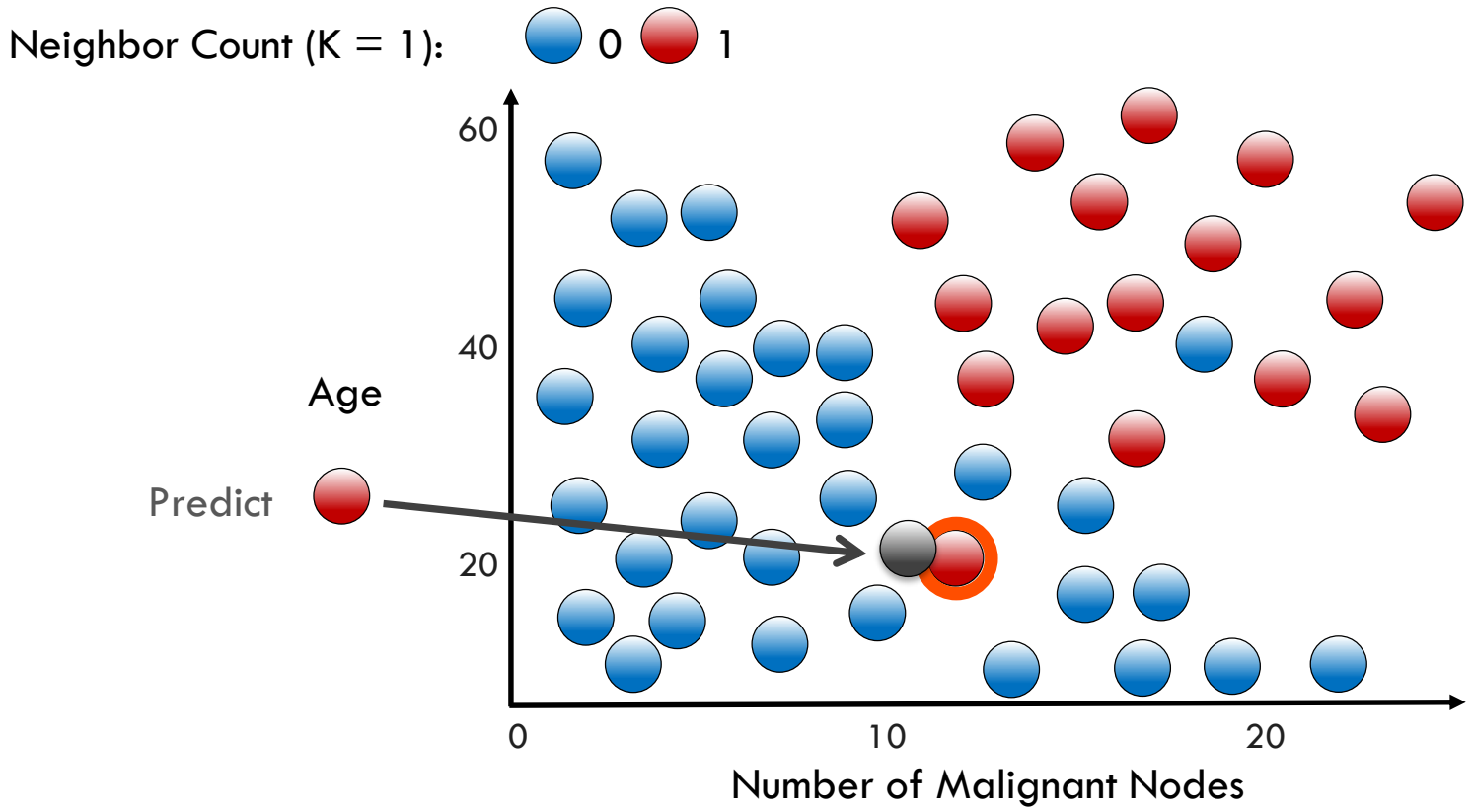
Your first classification algorithm (continued)

- KNN classification
  - How to make prediction?



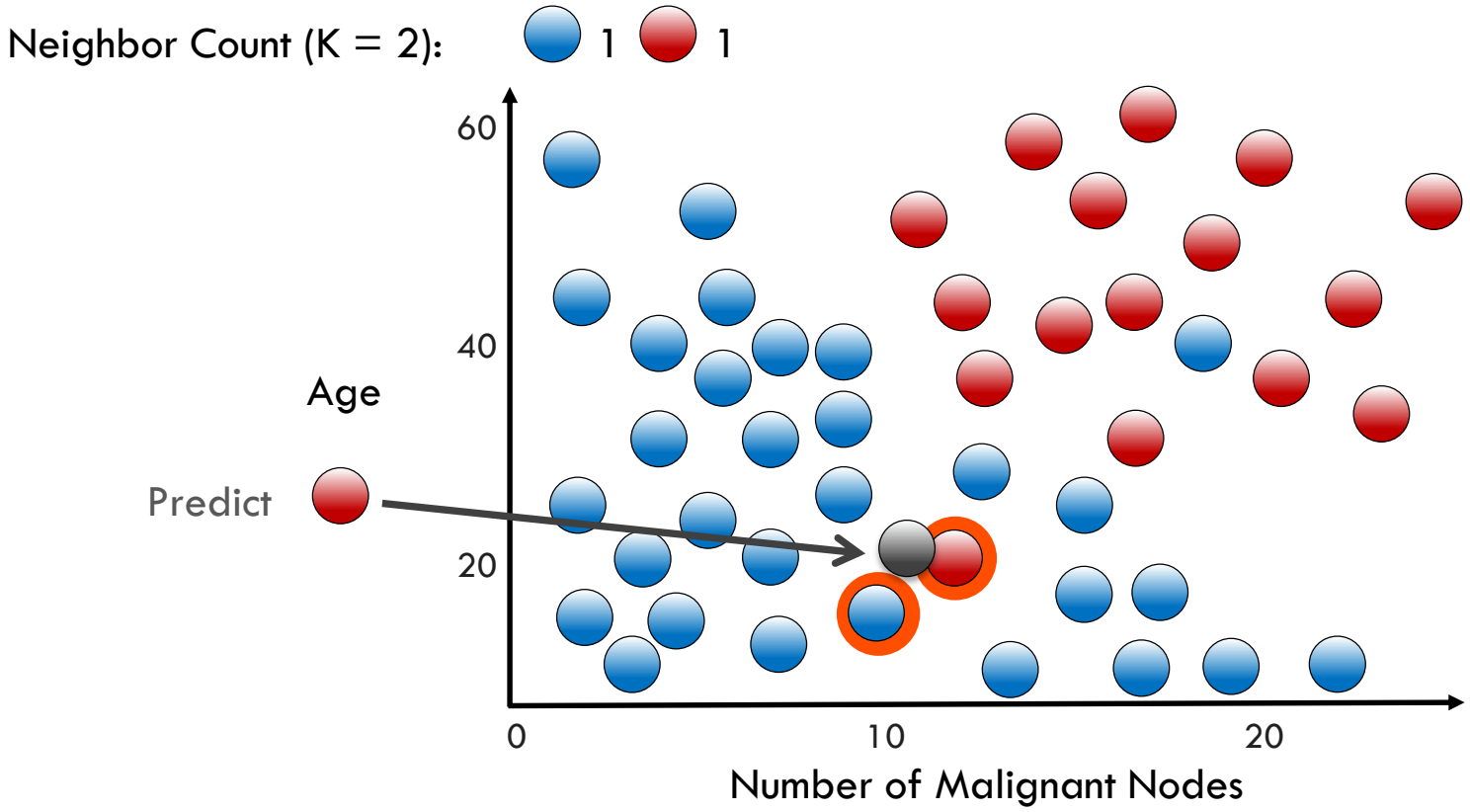
Your first classification algorithm (continued)

- K=1



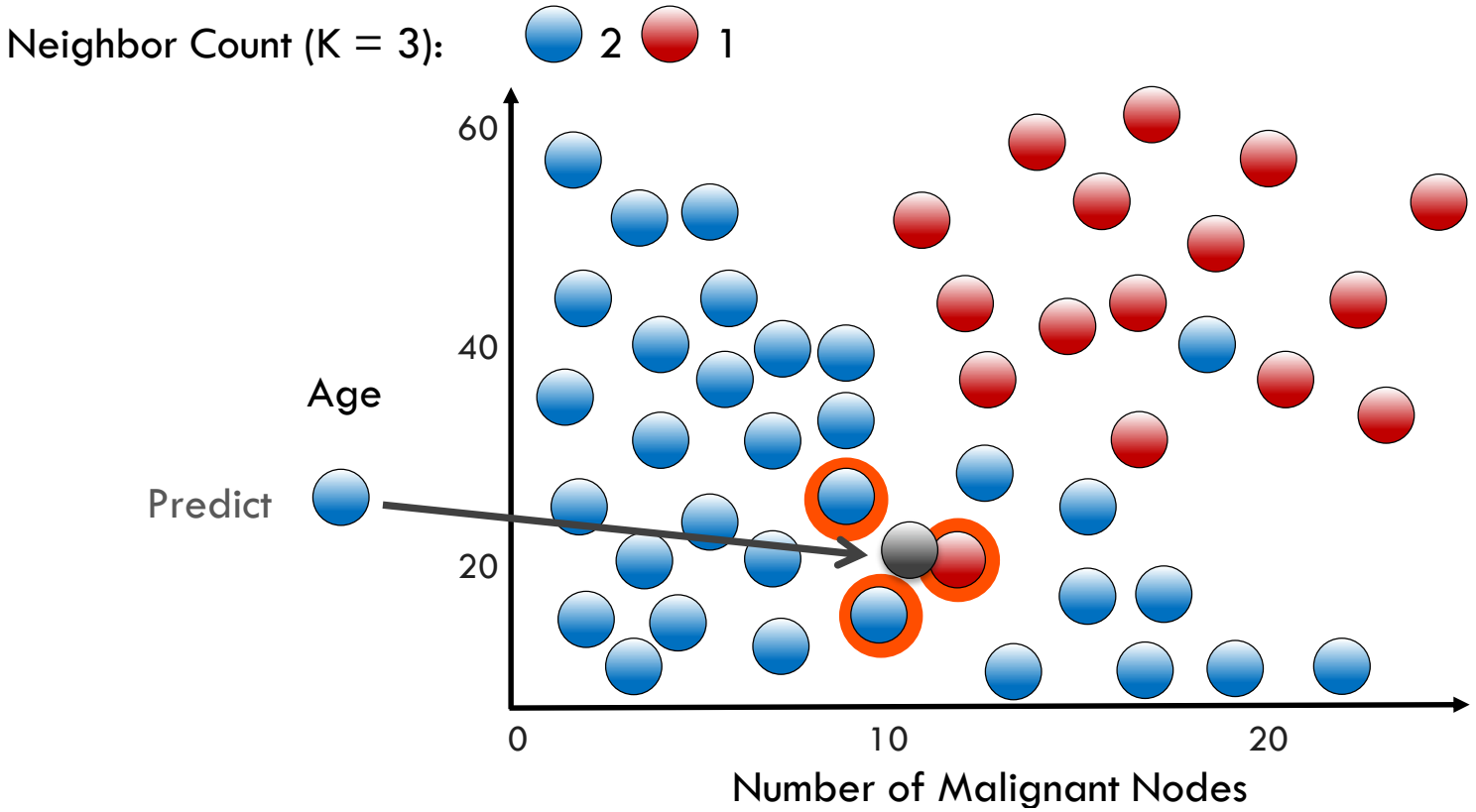
Your first classification algorithm (continued)

- K=2



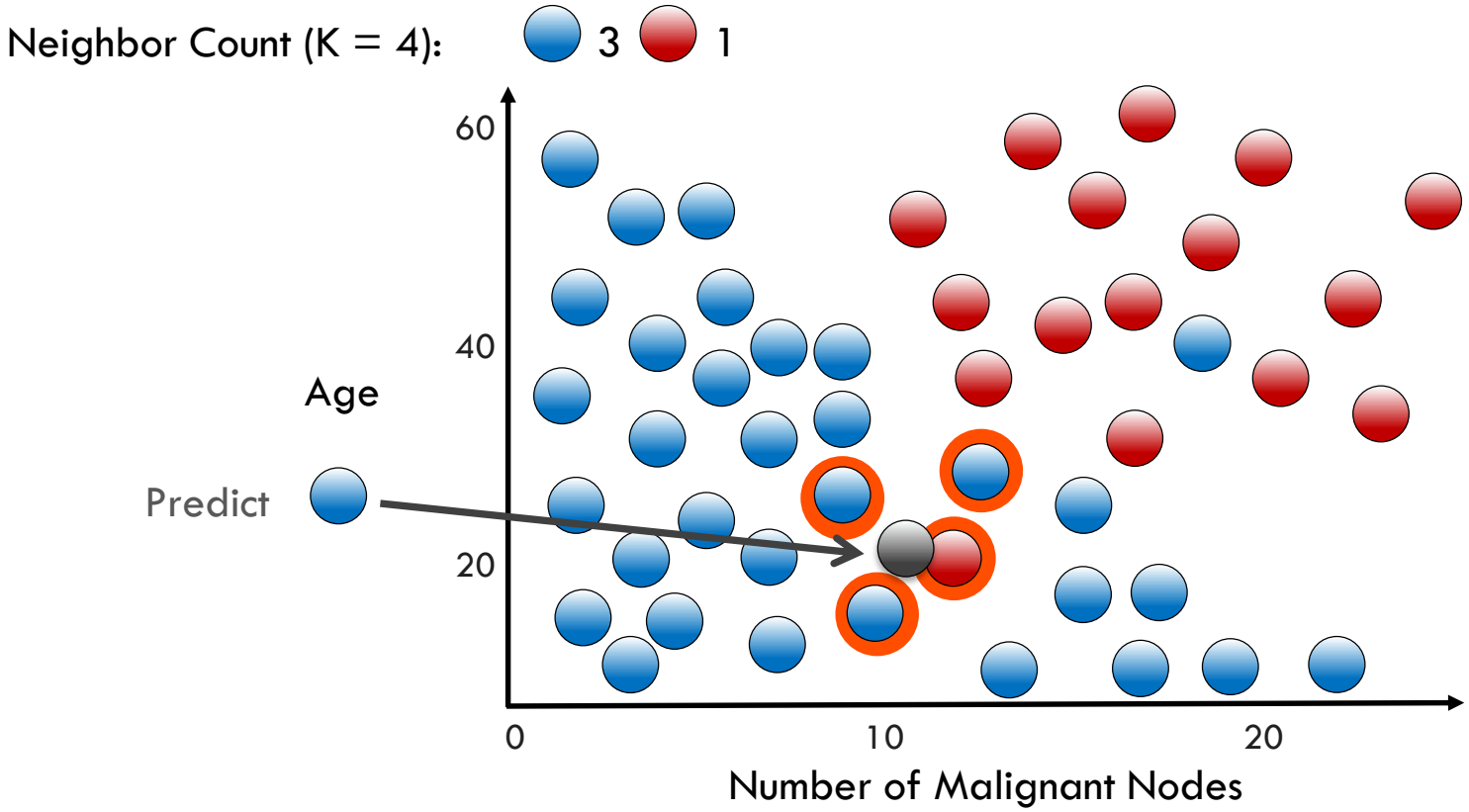
Your first classification algorithm (continued)

- K=3



Your first classification algorithm (continued)

- K=4

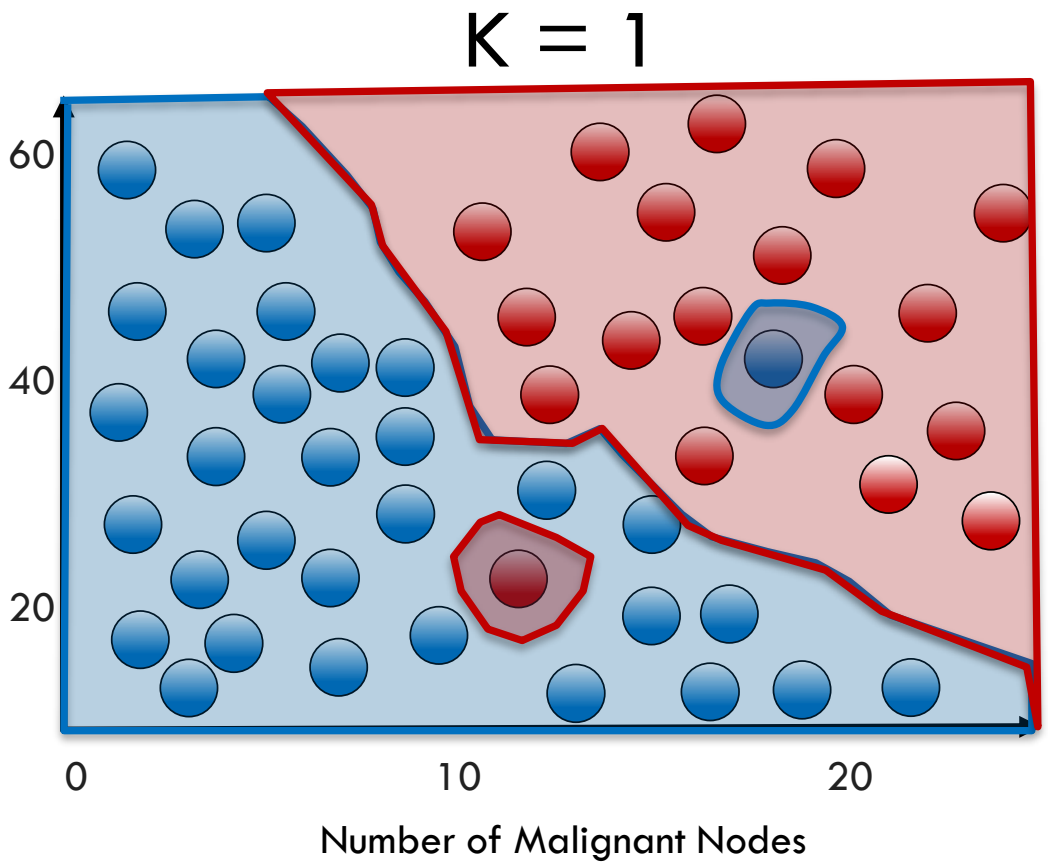


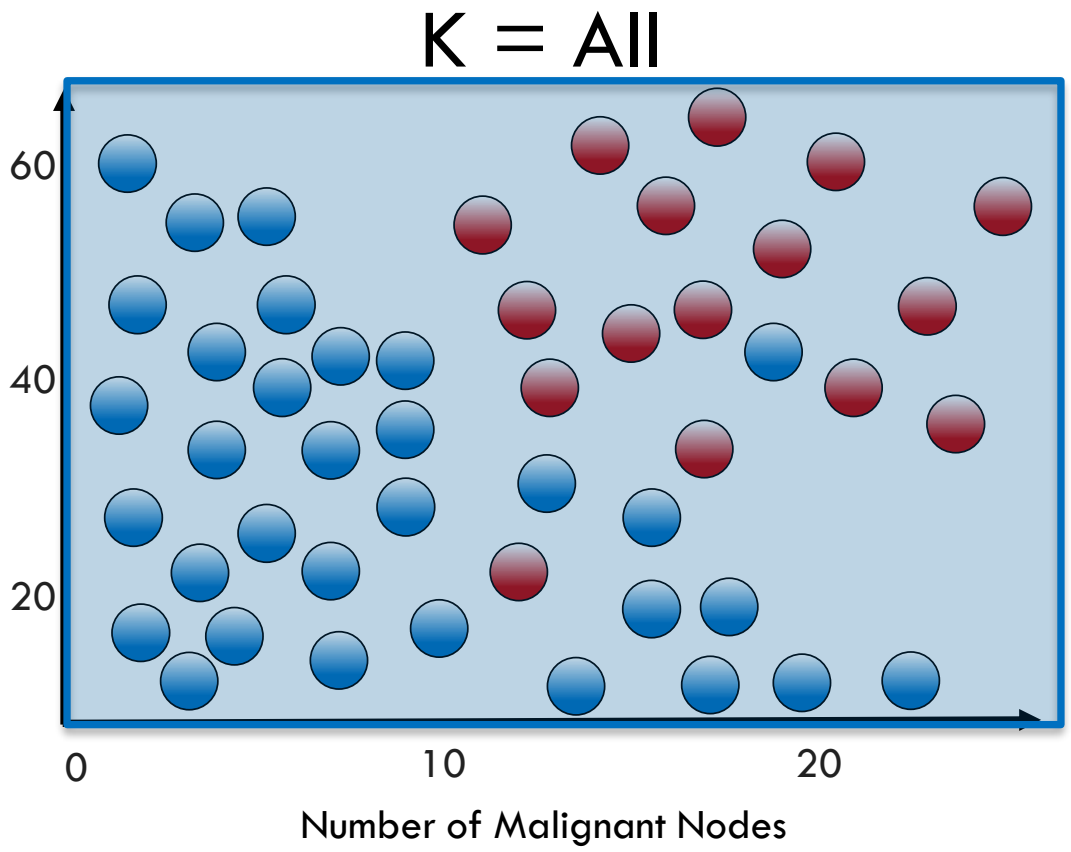
What is Needed to Select a KNN Model

- Correct value for ‘K’
- How to measure closeness of neighbors?
  - Euclidean distance (L2)

Decision boundary for varying K

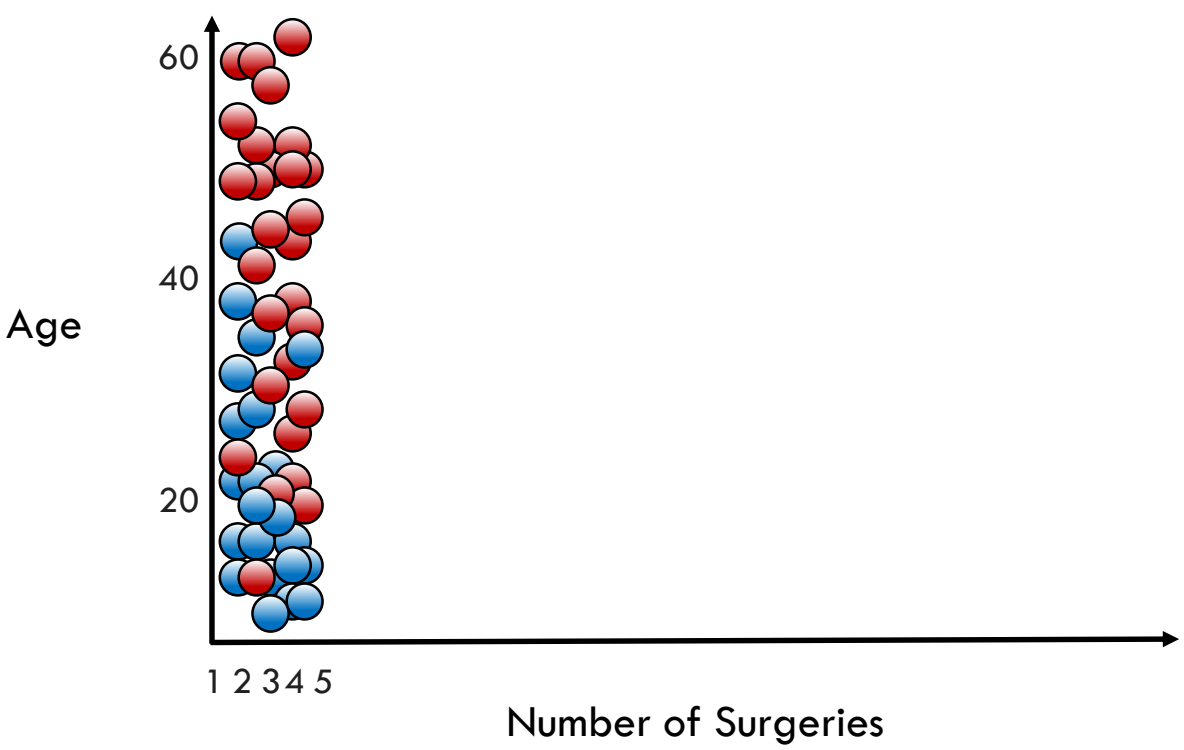
- K=1 or K=all

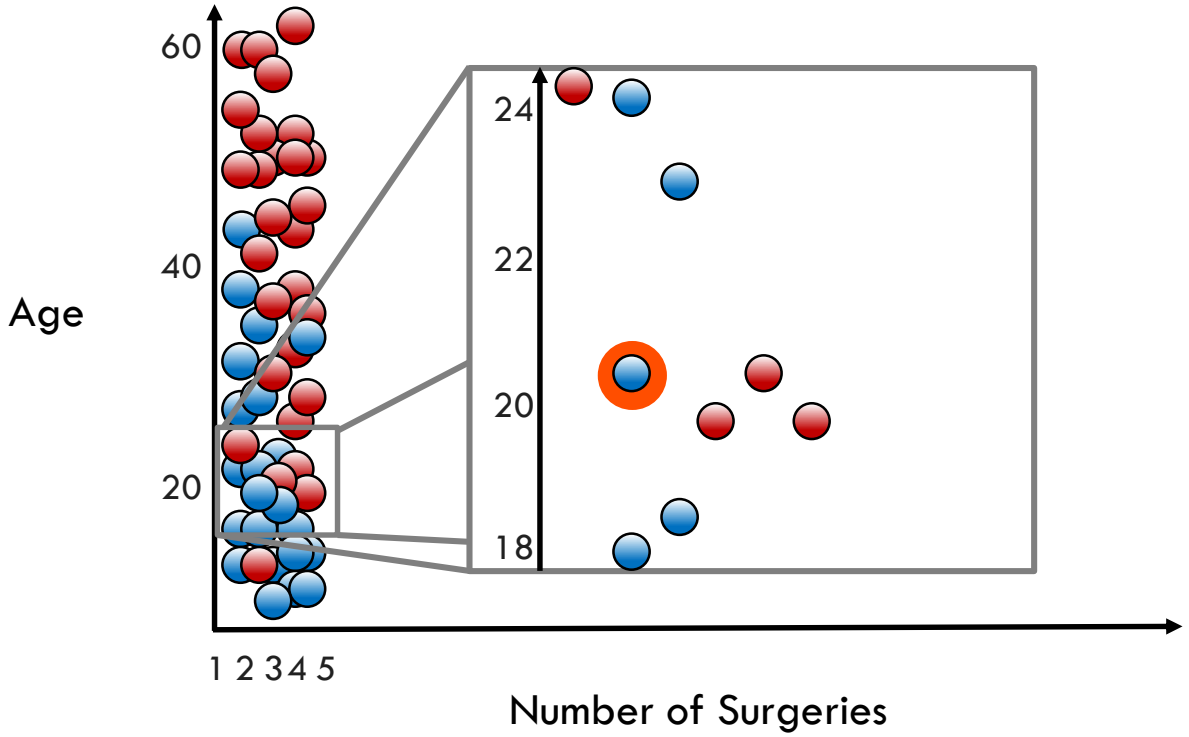




Feature scaling

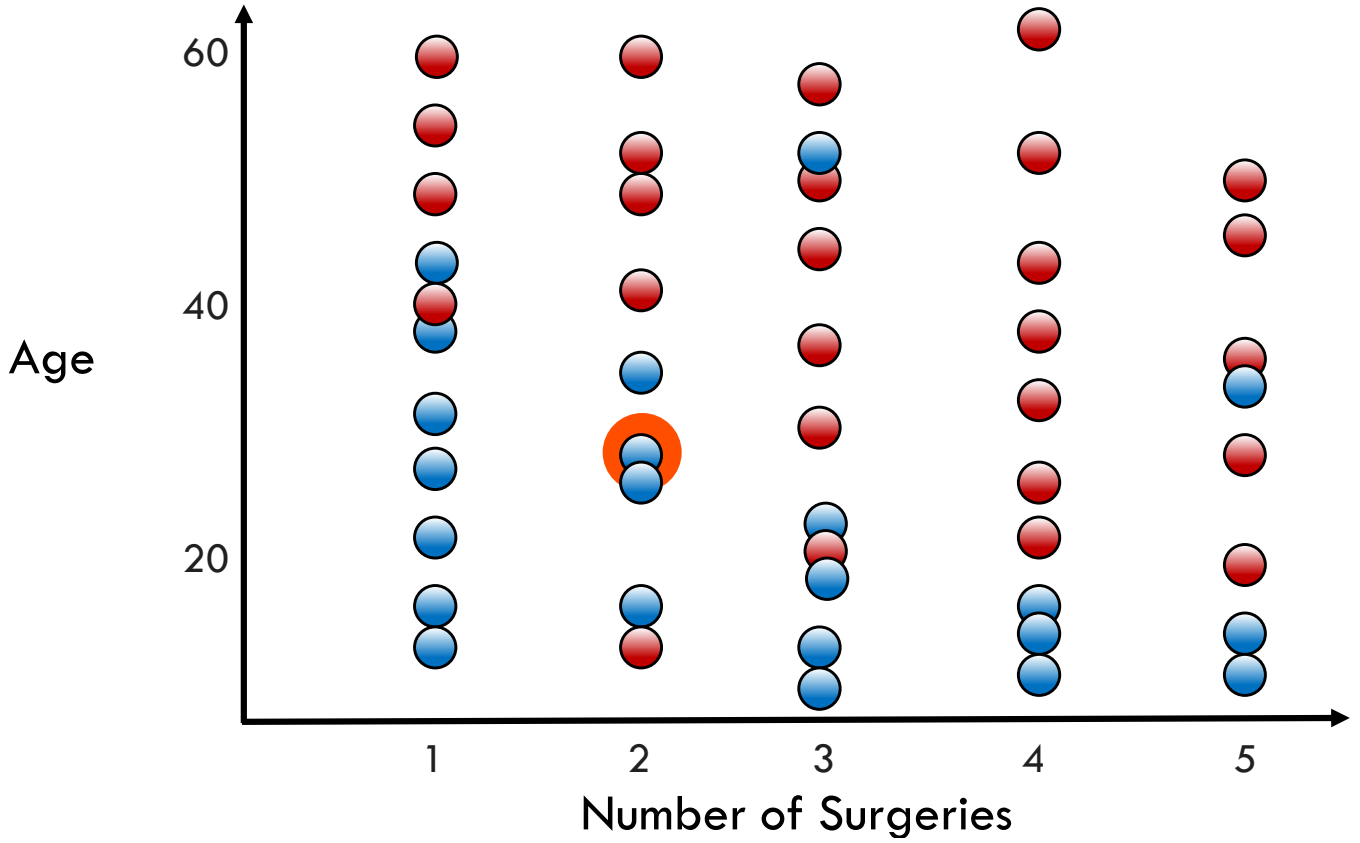
- Features with different numerical scales





Feature scaling (continued)

- After scaling

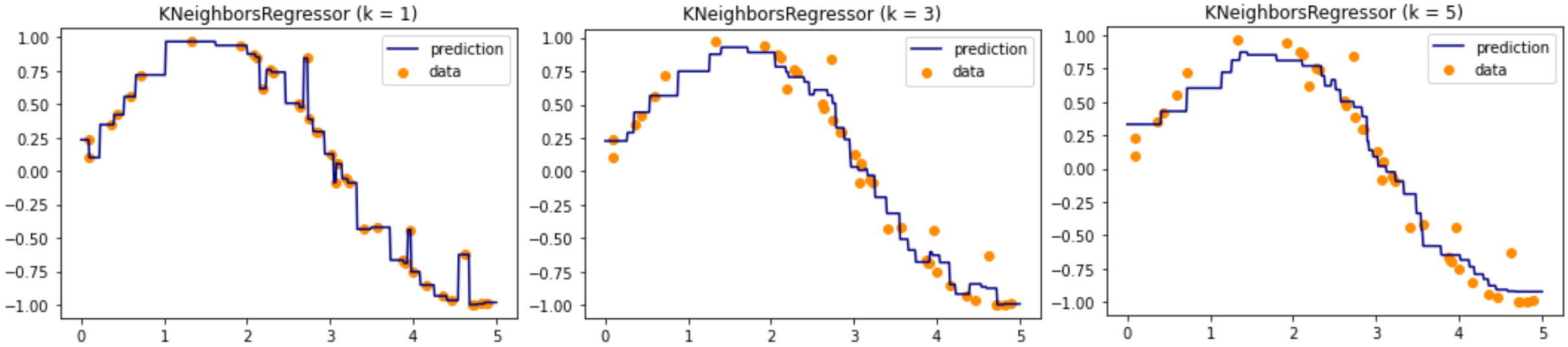


Feature scaling (continued)

- Standard scaler: mean center data and scale to unit variance
- Minimum-Maximum Scaler: scale data to fixed range (usually 0–1)
- Maximum Absolute Value Scaler: scale maximum absolute value

Regression with KNN

$$y = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$



Regression with KNN

$$y = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$