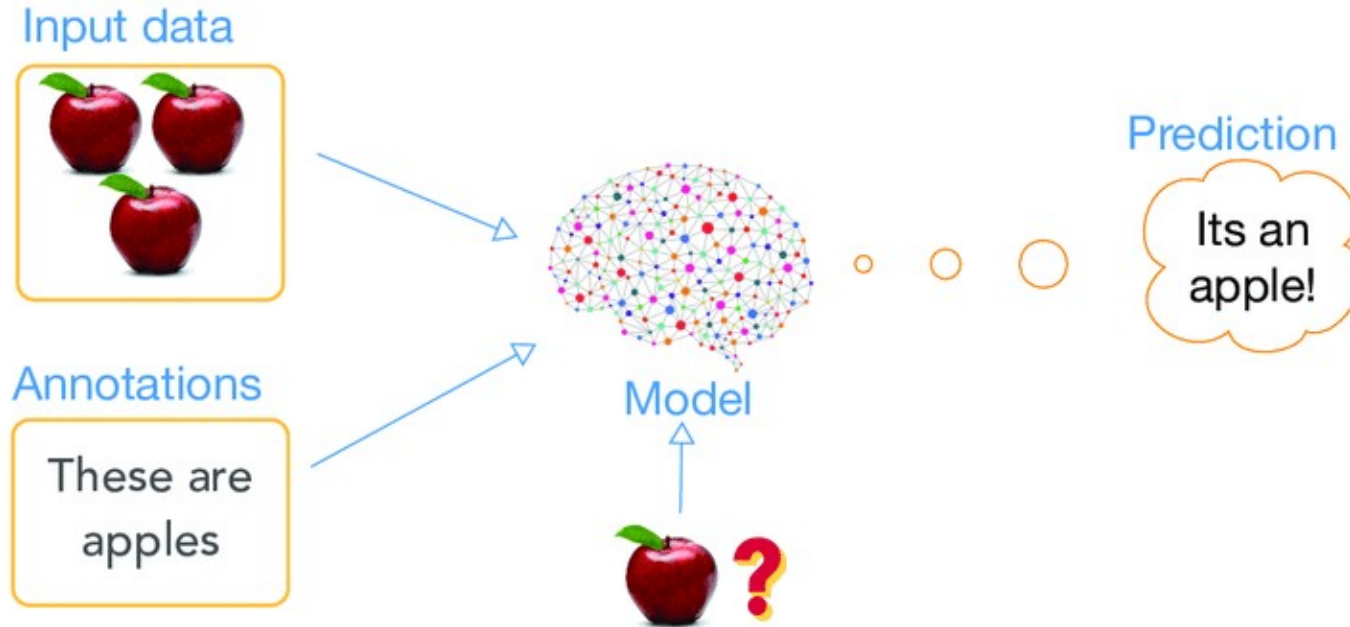


Supervised learning – continued

What is supervised learning?

supervised learning




Logistic regression

$$p(X) = P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 X}$$

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X$$



Link function: Links Y
to the linear predictor
Log-odds or logit

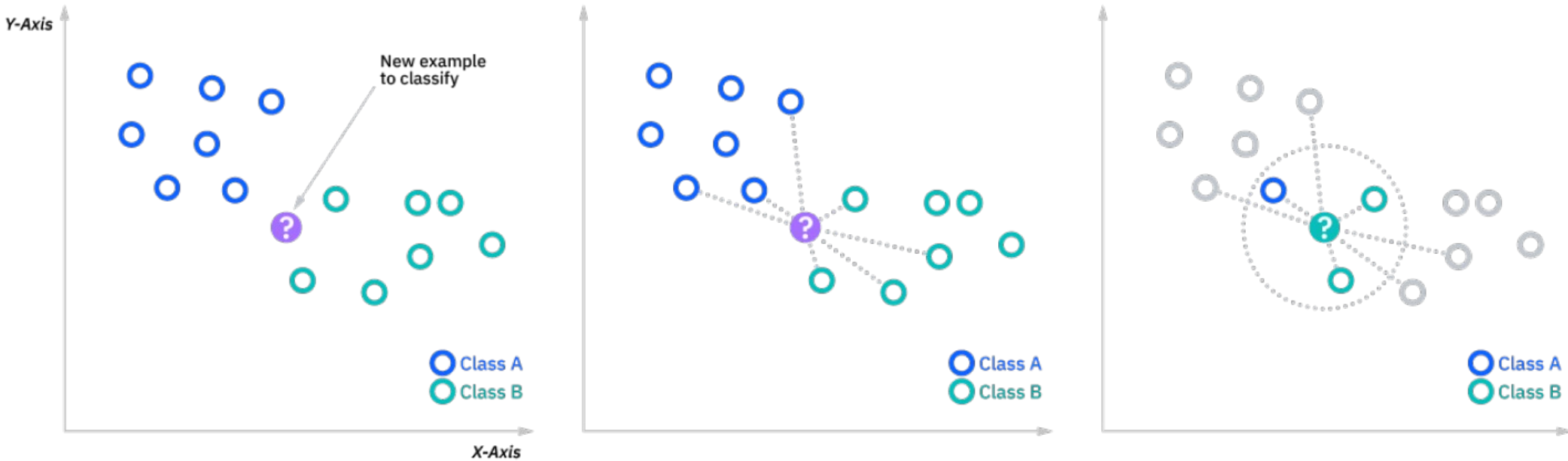


Linear predictor:
Linear function of X

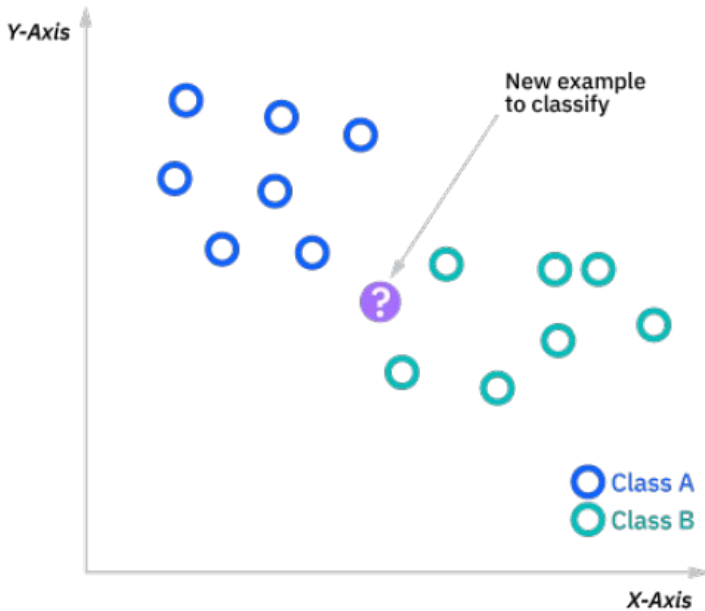
Other classification methods

- Ensemble learning
 - Boosting
 - Gradient boosting
 - AdaBoost
 - Bagging (**B**ootstrap **a**ggregating)
 - Random forest classification/regression
- Neighbors
 - K-nearest neighbors (KNN)

Neighbour based methods

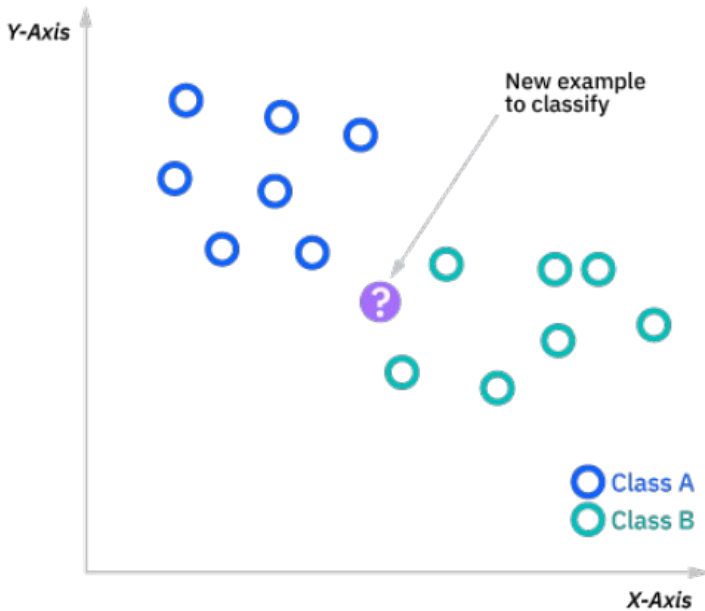


Neighbour based methods



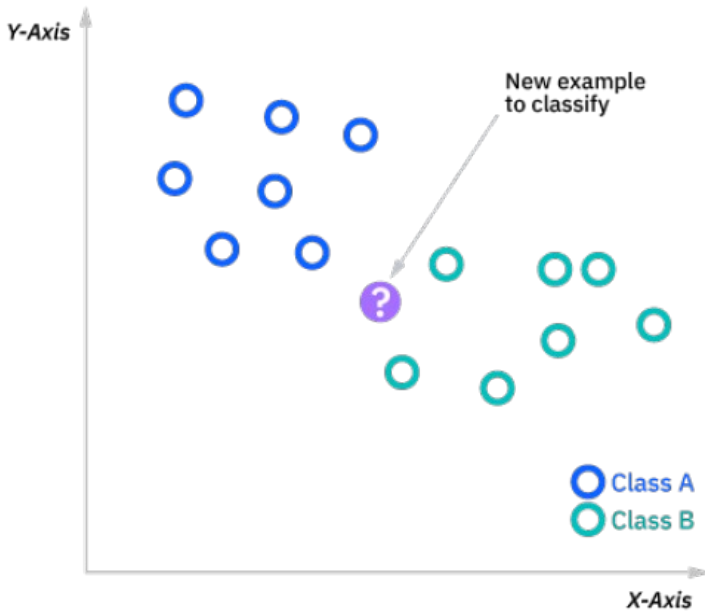
What class would you assign to the new point?

Neighbour based methods



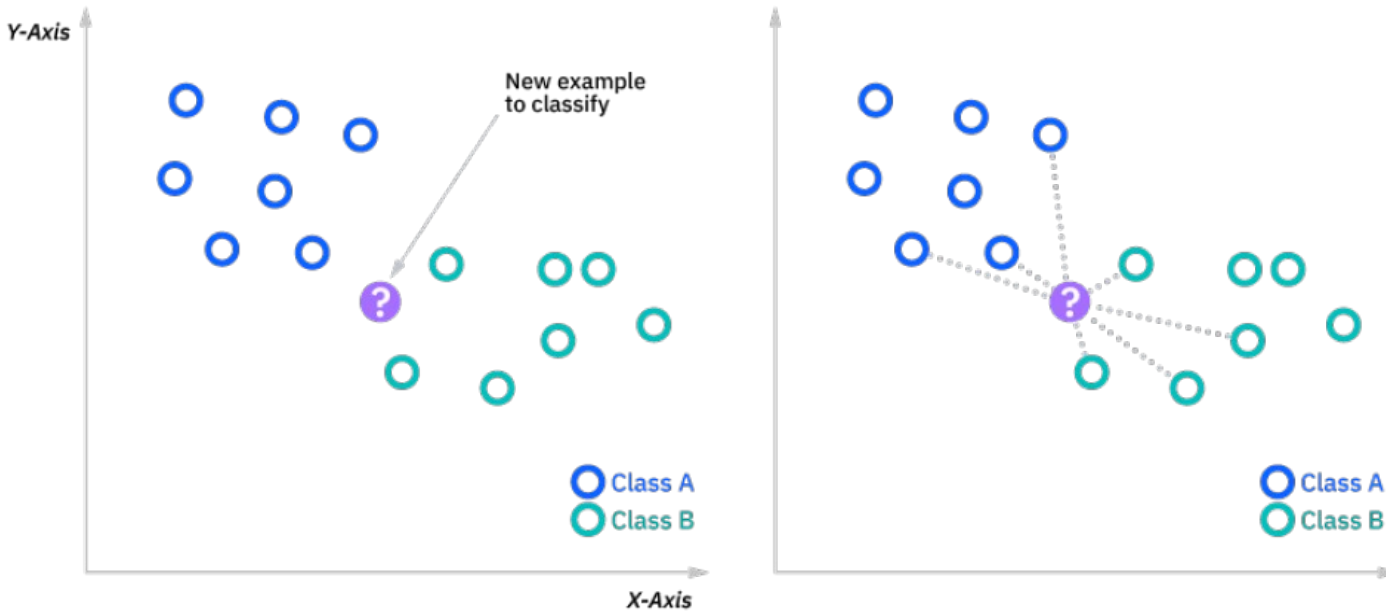
One way is to look at the neighbours of the new point

Neighbour based methods



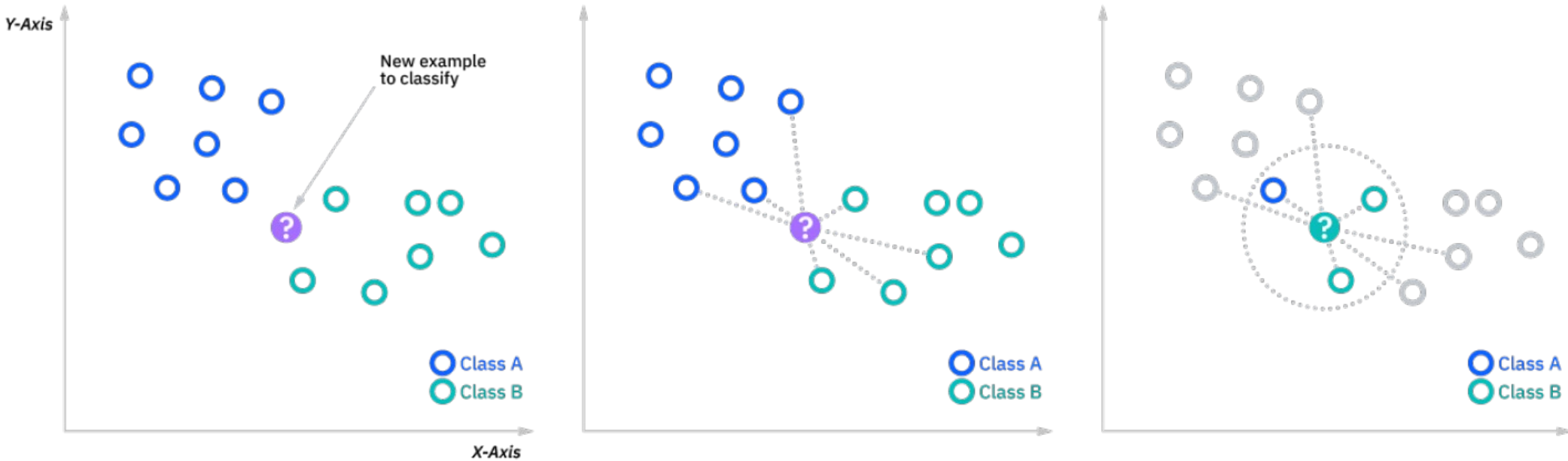
Ridiculous question time:
How do you choose
neighbours?

Neighbour based methods



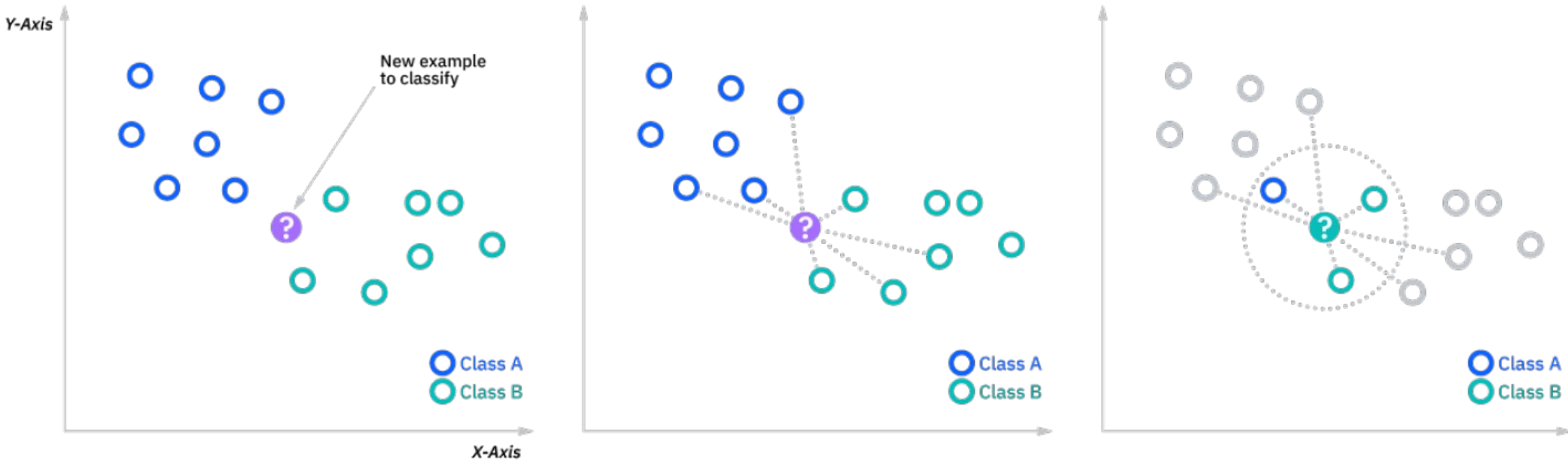
Choose k-nearest neighbours:
How do you pick k?

Neighbour based methods



Choose neighbours in a radius:
How do you pick radius?

Neighbour based methods



Once we have chosen neighbours, how do we predict?

Ensemble learning

- Wisdom of crowds

Ensemble learning

- Wisdom of crowds
 - Wikipedia
 - Democracy

Emsemble learning

- Aggregate weak models to get a much better one
 - Models must be very weakly correlated
 - Models must have some predictive power

Random forest classification

- Ensemble learning method
 - Based on decision trees

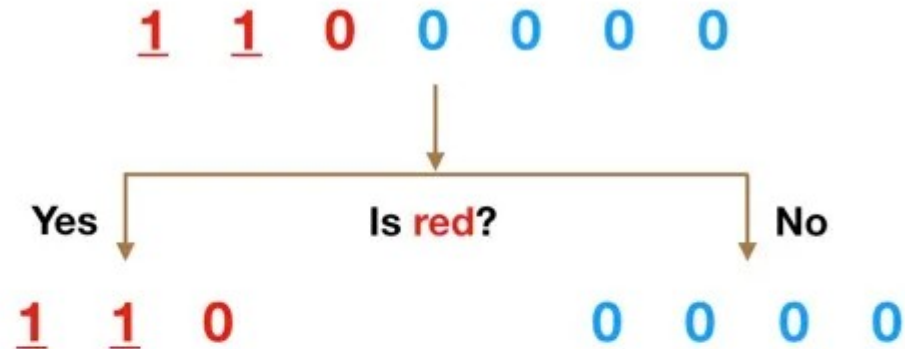
Quick digression: Decision trees

- Build trees that maximally separate our outcome of choice
 - Each node is decision based on a single variable (feature)

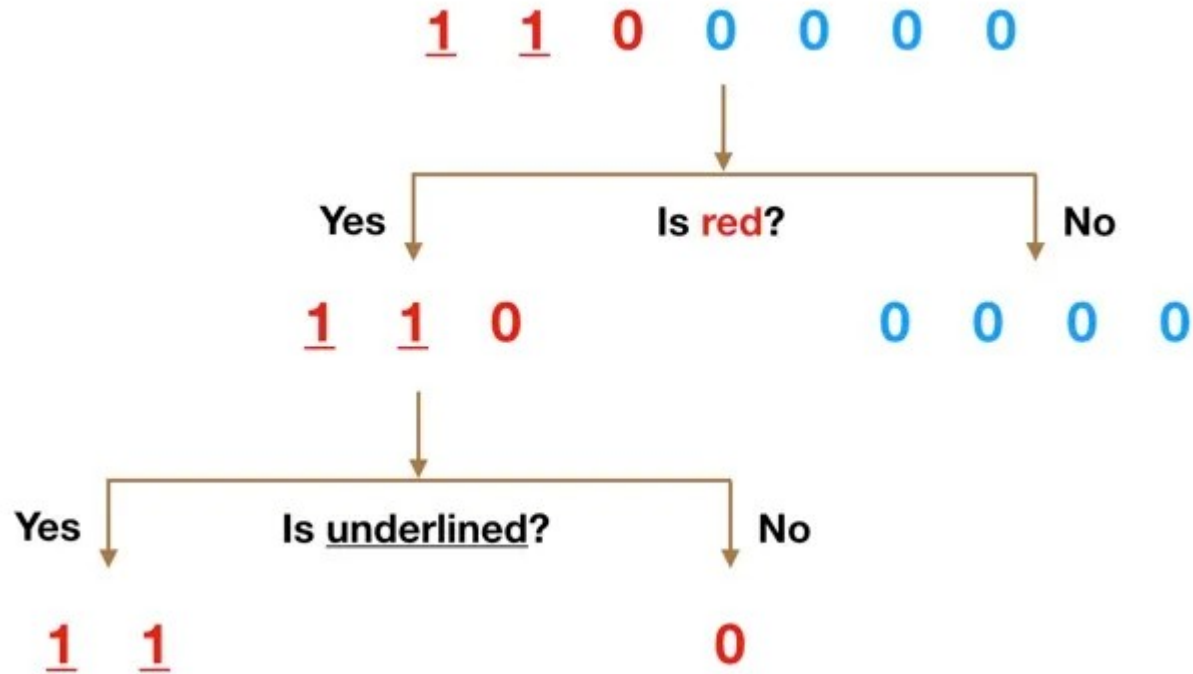
Quick digression: Decision trees

1 1 0 0 0 0 0

Quick digression: Decision trees



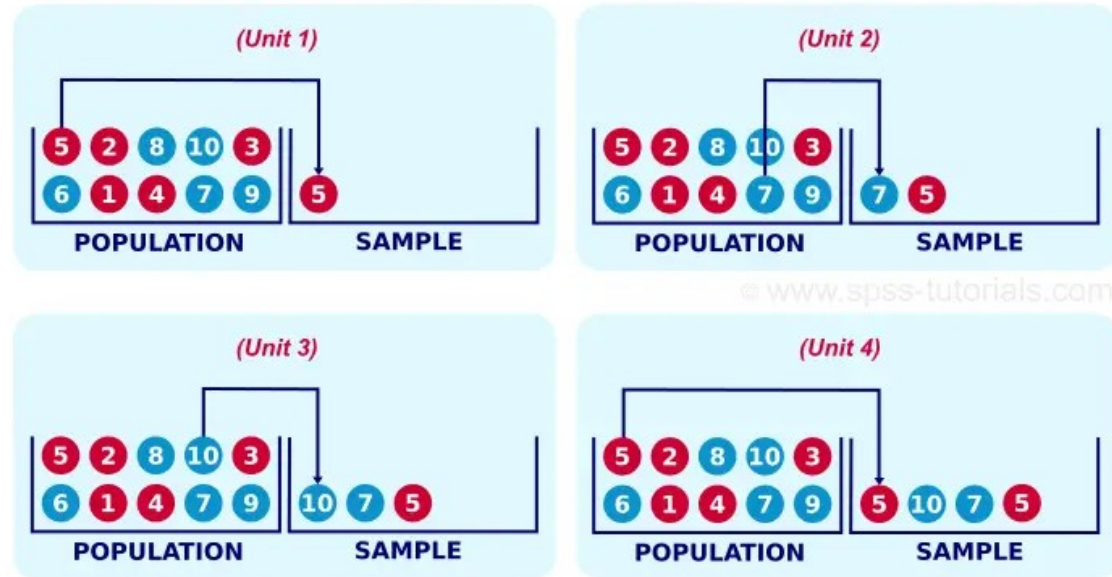
Quick digression: Decision trees



From Decision Trees to Random Forests

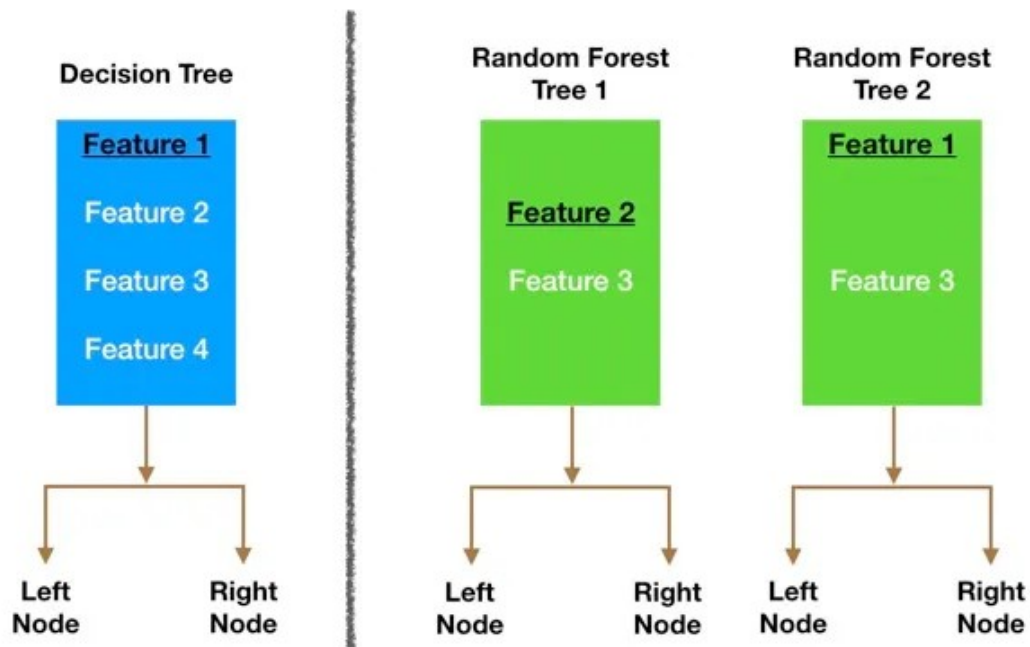
- Build many decision trees
 - Each tree based on random resampling of data (with replacement)

SIMPLE RANDOM SAMPLING WITH REPLACEMENT



From Decision Trees to Random Forests

- Forest with many decision trees
 - Each tree based on random resampling of data (with replacement)
 - Each tree built on a random set of features

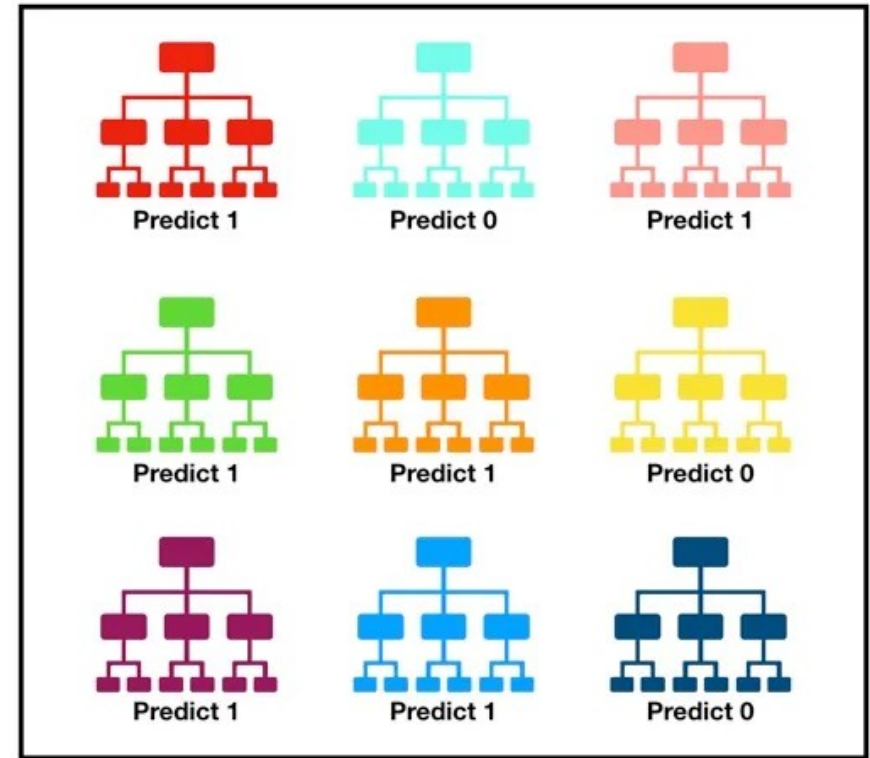


From Decision Trees to Random Forests

- Forest with many decision trees
 - Each tree based on random resampling of data (with replacement)
 - Each tree built on a random set of features
 - Ensures that the different trees are not highly correlated with each other

Random forest classifiers

- The final prediction is based on votes from individual trees
 - Each tree the RF gets one vote
 - Majority vote is the prediction



Tally: Six 1s and Three 0s

Prediction: 1

Advantages of random forest

- Highly flexible – can work for a variety of problems
- Very scalable – number of input features can be very high
- Reduces overfitting, while improving prediction
- Can be used for both classification and regression

Exercise time