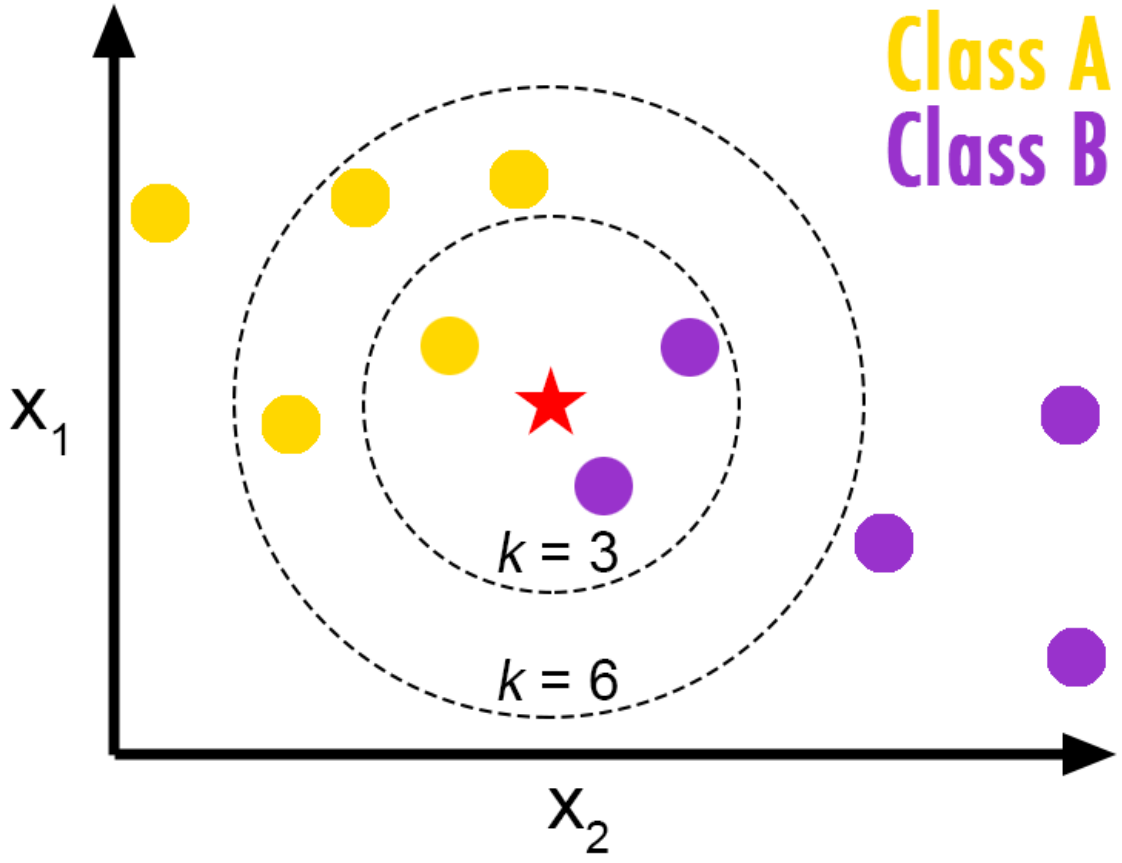
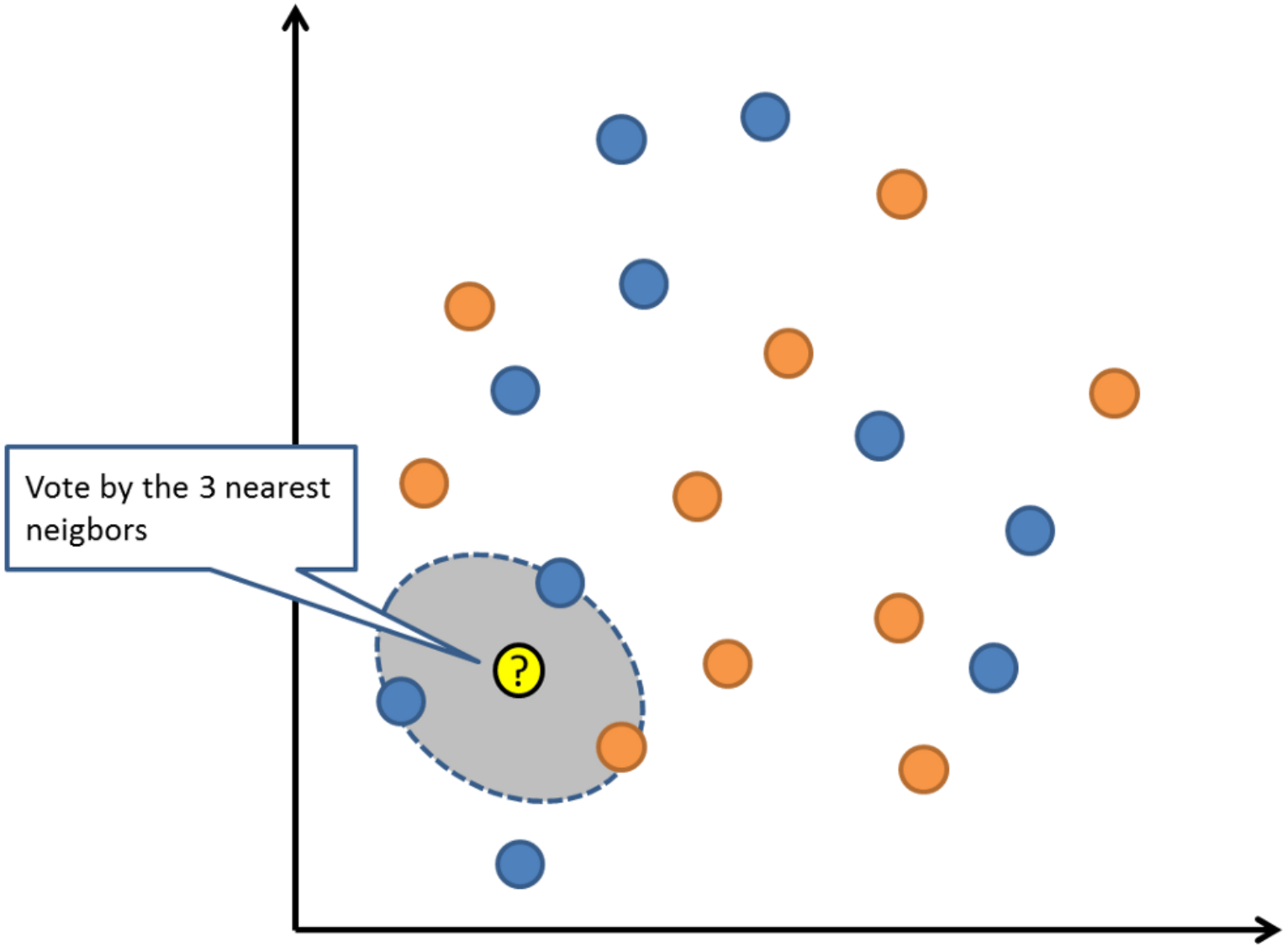
k nearest neighbors

Introduction

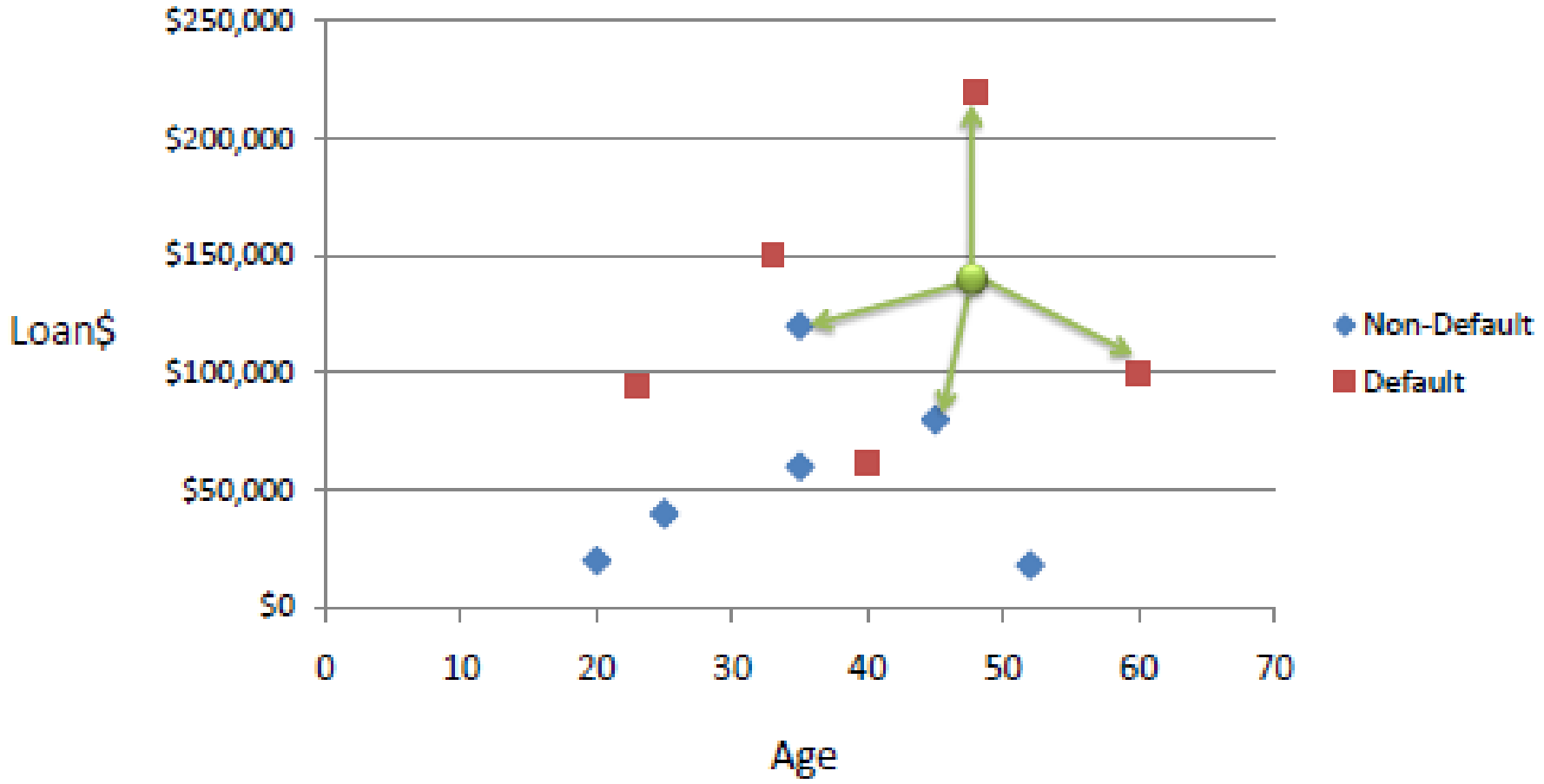
- An object is classified by a majority vote of its neighbors.



Introduction



Data concerning credit default



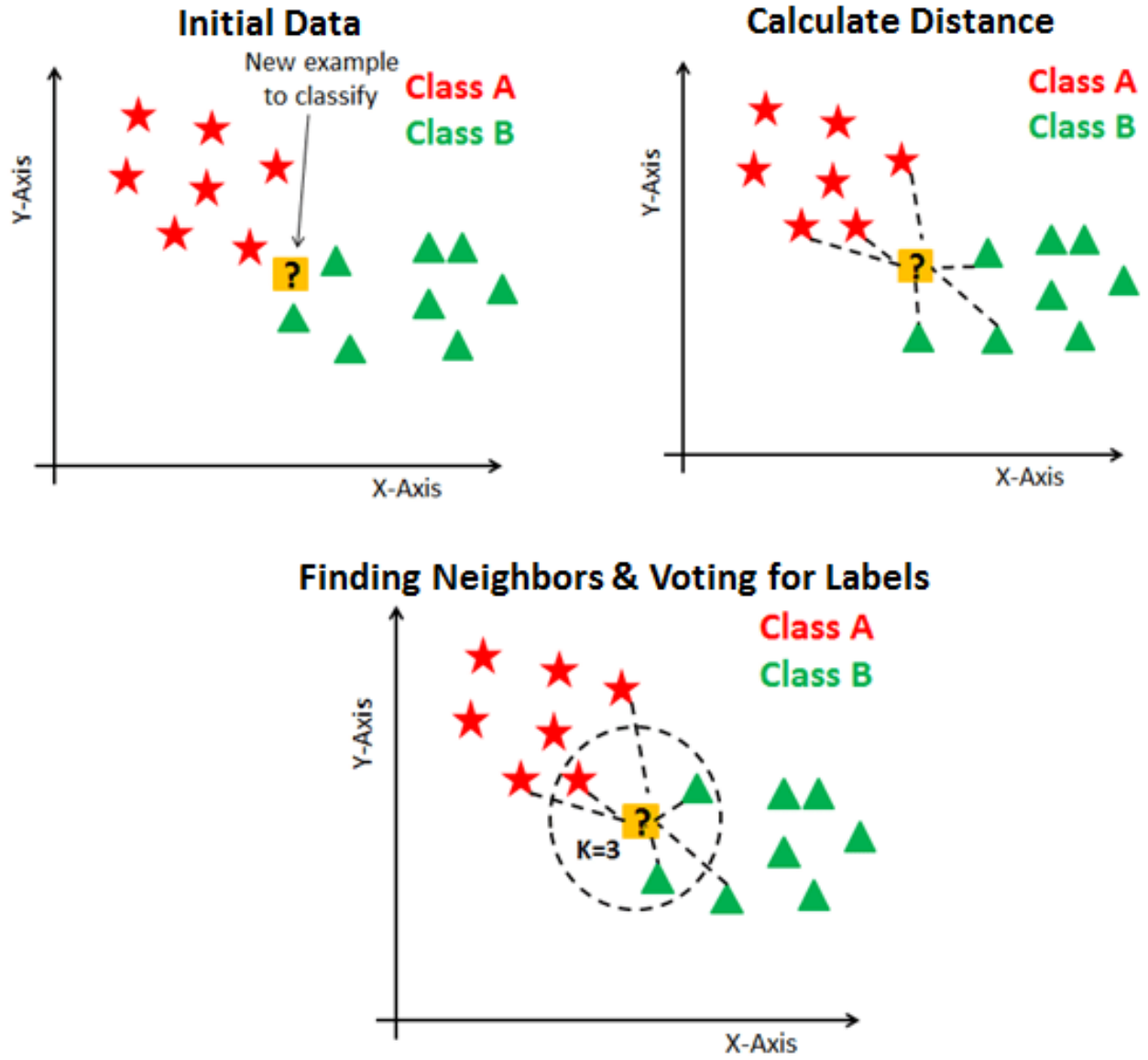
k-Nearest-Neighbors

- It is essentially classification by finding the most similar data points in the training data, and
- making an educated guess based on their classifications.
- In KNN, we must be able to keep the entire training set in memory.
- Performing classifications can be computationally expensive as the algorithm parse through all data points for each classification.
- For these reasons, kNN tends to work best on smaller data-sets that do not have many features.

How does the KNN algorithm work?

- In KNN, K is the number of nearest neighbors.
- Suppose P1 is the point, for which label needs to predict.
- First, you find the k closest point to P1 and then classify point by majority vote of its k neighbors.
- Each object votes for their class and the class with the most votes is taken as the prediction.
- For finding closest similar points, you find the distance between points.

How does the KNN algorithm work?



K Nearest Neighbours — Pseudocode

- 1. Load the training and test data
- 2. Choose the value of K
- 3. For each point in test data:
- - find the distance to all training data points
- - store the distances in a list and sort it
- - choose the first k points
- - assign a class to the test point based on the majority of classes present in the chosen points
- 4. End

Thanks