K-NEAREST NEIGHBORS

Thursday, January 27th

K-NEAREST NEIGHBORS

- A supervised learning method that aims to address classification and regression problems.
  - What is *supervised learning?*
    - A method that relies on **labeled input data** to learn a function that creates output (classifications) when given new unlabeled data.

What does the K-Nearest Neighbors algorithm actually do?

- Assumption: similar things exist in close proximity.
- For a new data point, we want to assign the object to the class most common among its *K* nearest neighbors.
  - Majority vote

K-NEAREST NEIGHBORS

For a new data point, we want to assign the object to the class most common among its *K* **nearest** neighbors.

- Euclidean distance

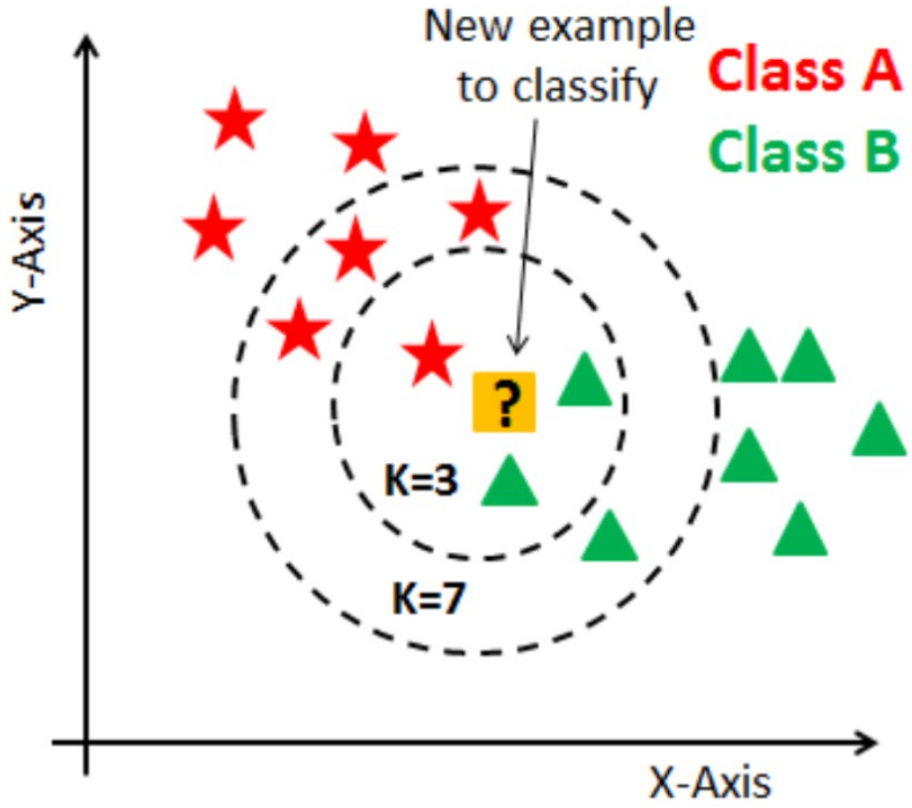

Photo by Sarang Anil Gotke on Kdnuggets

K-NEAREST NEIGHBORS

How to choose the best K?

- Usually choose odd numbers
- Can try multiple different values of K that reduce the number of errors
- K = 1 is sensitive
- Large K is more stable due to majority vote, however, too large of a K can result in increasing errors!

Advantages of KNN

- Simple model
- Do not need to adjust multiple parameters or assumptions
- Can be used for classification *or* regression.

K-NEAREST NEIGHBORS

Disadvantages of KNN

- Computational time
  - Dependent upon number of predictors or independent variables used.

K-NEAREST NEIGHBORS