K-NN CLASSIFIER

Shingchern D. You

K-NN intro

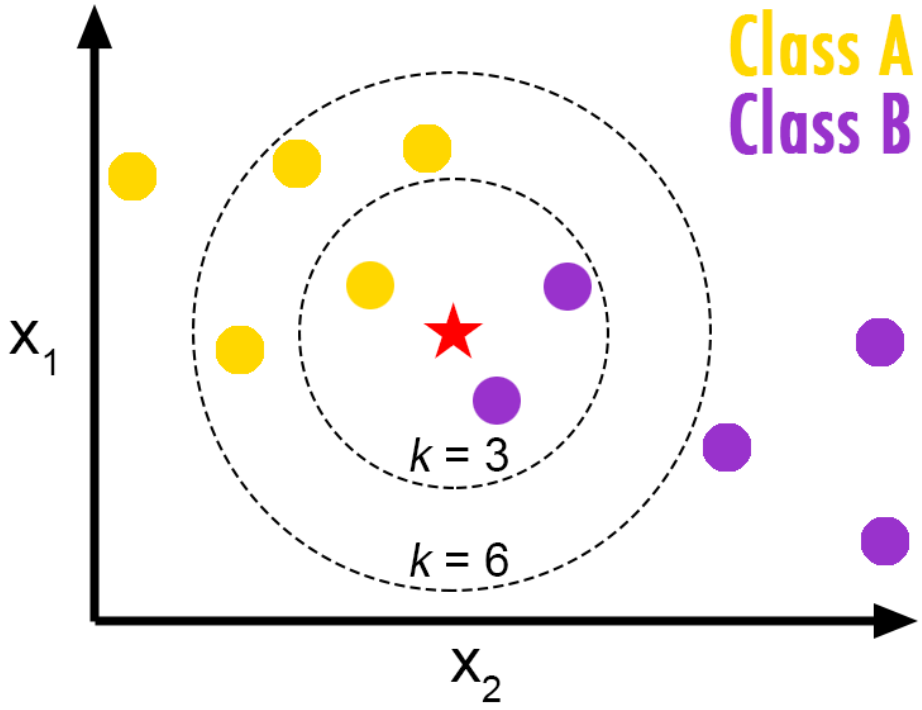
- The simplest case of supervised learning
- In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression – Wikipedia
- Instant-based learning (also called lazy learning)

K-NN algorithm for classification

- **Meta-parameter**: *k* is a user-defined constant
- Training phase: Store the feature vectors and class labels of the training samples
- Classification phase: Unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point

K-NN algorithm for classification (1)

- Simple illustration



K-NN algorithm for classification (2)

- X_i is a feature or attribute
- Classification result is a function of *k*
- In general, larger values of *k* reduces effect of the noise on the classification, but make boundaries between classes less distinct – remember the Occam's razor principle?

K-NN algorithm for classification (3)

- How to determine a suitable value of *k*
- Recall we have training, validation, and testing sets
- The validation set is usually used to tune the meta-parameters
- If the features are really "good," even k-NN produces satisfactory accuracy

K-NN for regression

- K-NN can also be used to perform regression
- You can find abundant documents about how to do it