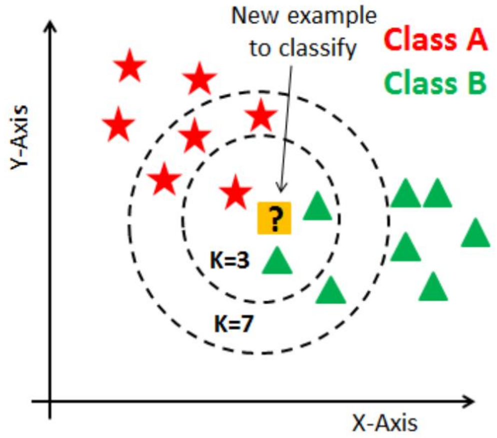
CLASSIFYING IRIS USING MACHINE LEARNING

Presented by: Julien Dib

Rida Diab

Lebanese University – Faculty of Engineering III

KNeighborsClassifier



- **K-nearest-neighbor (kNN)** classification is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data.
- The k-nearest-neighbor classifier is commonly based on the Euclidean distance between a test sample and the specified training samples. Let xi be an input sample with p features (xi1,xi2,…,xip) , n be the total number of input samples (i=1,2,…,n) and p the total number of features (j=1,2,…,p) . The Euclidean distance between sample xi and xl (l=1,2,…,n) is defined as
  $d(\mathbf{x}_i,\mathbf{x}_l)= \sqrt{(x_{i1}-x_{l1})^2+(x_{i2}-x_{l2})^2+\cdots+(x_{ip}-x_{lp})^2}$
- The k-nearest-neighbor classification rule is to assign to a test sample the majority category label of its k nearest training samples. In practice, k is usually chosen to be odd, so as to avoid ties.

Classification decision rule

- Classification typically involves partitioning samples into training and testing categories. Let $\mathbf{x}_i$ be a training sample and $\mathbf{x}$ be a test sample, and let ω be the true class of a training sample and $\hat{\omega}$ be the predicted class for a test sample.
- During the training process, we use only the true class ω of each training sample to train the classifier, while during testing we predict the class $\hat{\omega}$ of each test sample.
- With 1-nearest neighbor rule, the predicted class of test sample $\mathbf{x}$ is set equal to the true class ω of its nearest neighbor.
- For k-nearest neighbors, the predicted class of test sample $\mathbf{x}$ is set equal to the most frequent true class among k nearest training samples.

Confusion Matrix

- The confusion matrix used for tabulating test sample class predictions during testing is denoted as **C**
- During testing, if the predicted class of test sample **x** is correct, then the diagonal element of the confusion matrix is incremented by 1.
- If the predicted class is incorrect, then the off-diagonal element is incremented by 1.
- Once all the test samples have been classified, the classification accuracy is based on the ratio of the number of correctly classified samples to the total number of samples classified, given in the form

  Acc = $\frac{\sum_{\omega}^{\Omega} c_{\omega\omega}}{n_{total}}$

  where $C_{\omega\omega}$ is a diagonal element of **C** and n_{total} is the total number of samples classified.

Google Colab

Let’s go to Google Colab