CS 6316 : Machine Learning

In-class activity: K-Nearest Neighbor Algorithm

In this activity, you will further explore KNN algorithm through python code.

The data set we'll be using is the Iris Flower Dataset (IFD) which was first introduced in 1936 by the famous statistician Ronald Fisher and consists of 50 observations from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals. Our goal is to train the KNN algorithm to be able to distinguish the species from one another given the measurements of the 4 features.

Lets start by running an example code using scikit-learn to train a KNN classifier and evaluate its performance on the data set. There is 4 step modeling pattern:

1. Import the learning algorithm
2. Instantiate the model
3. Learn the model
4. Predict the response

The following steps will print out accuracy score for K=3: