**Assignment 4 (Due: 14 November 2019)**

Dataset for Classification Tasks

https://archive.ics.uci.edu/ml/datasets/Leaf    (Description)

https://archive.ics.uci.edu/ml/machine-learning-databases/00288/leaf.zip (Data File)

Column 1: Class Label (Total classes: 36)

Column 2: ID of sample for each class (tells you the total number of examples for each class)

Use attributes from column 3-16 for the Tasks below.

===========================

**Task 1: Performance Evaluation of k-nearest neighbor classifier using Cross-Validation.**

(a) Randomly pick 5 examples for each class for training/cross-va lidation (Total: 5 \* 36 = 180). The rest of the data should be used for testing.

(b) Perform leave-one-out cross-validation (LOOCV) for k = 1, 3, 5, 7, 9 to compute their mean accuracies.

(c) Use the training examples to build classifiers for k = 1, 3, 5, 7, 9 and compute their mean testing accuracies.

(d) Plot a graph for the results in (b) and (c) (x-axis: k and y-axis: mean accuracies (%))

(e) Based on LOOCV performance evaluation in (b), which k performs the best? Does it correspond to the testing error?

(f) Compare and Discuss the results in (b) and (c).

**Task 2: Study the effect of the number of training examples on the K-NN classification performance.**

(a) Randomly pick 6 examples for each class for training/cross-validation (Total: 6 \* 30 = 180). The rest of the data should be used as testing examples.

(b) Build a K-NN (K = 1) classifier using n=1 example from each class and compute the testing accuracy using the testing examples in (a).

(c) Repeat Step (b), for n = 2, 3, 4, 5, 6.

(d) Repeat Step (a), (b), (c) 20 times and compute their mean testing accuracies and the corresponding standard deviation.

(e) Plot a graph using the mean testing accuracies (y-axis) and the standard deviation (error bar for mean accuracies) against the number of training examples (x-axis)

(f) Discuss the results in (e).