ECE 760 Homework 2: kNN
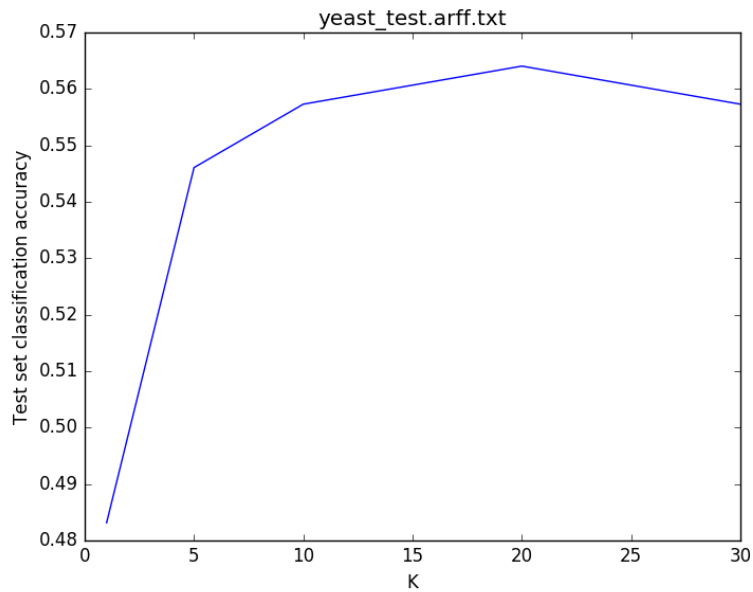
Qihong Lu

October 12, 2016

QUESTION 1

Code:

- kNN_alg.py: the main program that implements the standard kNN for multi-class classification and regression

- util.py: the definitions of some constants and helper functions

- kNN.py: run kNN algorithm, return test results

- kNN-select.py: run kNN algorithm, tune K by leave-one-out validation. Fit final model with the best K and return the test results
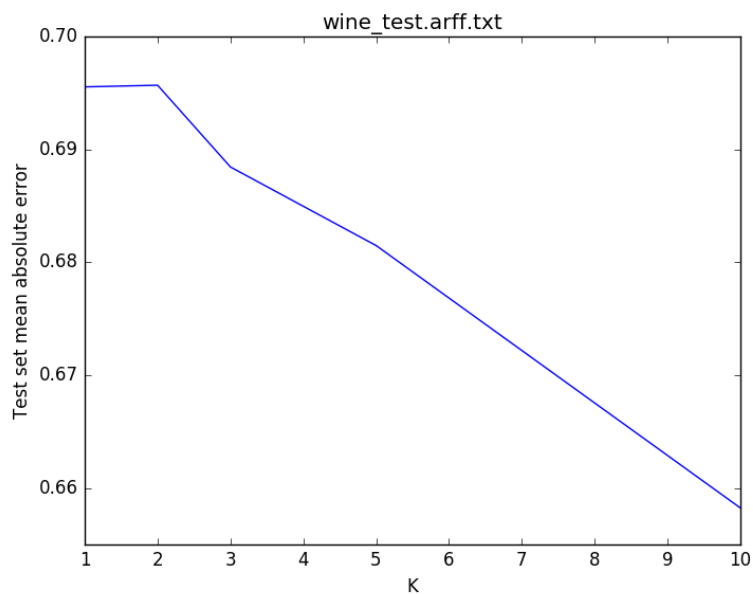
Dependencies:
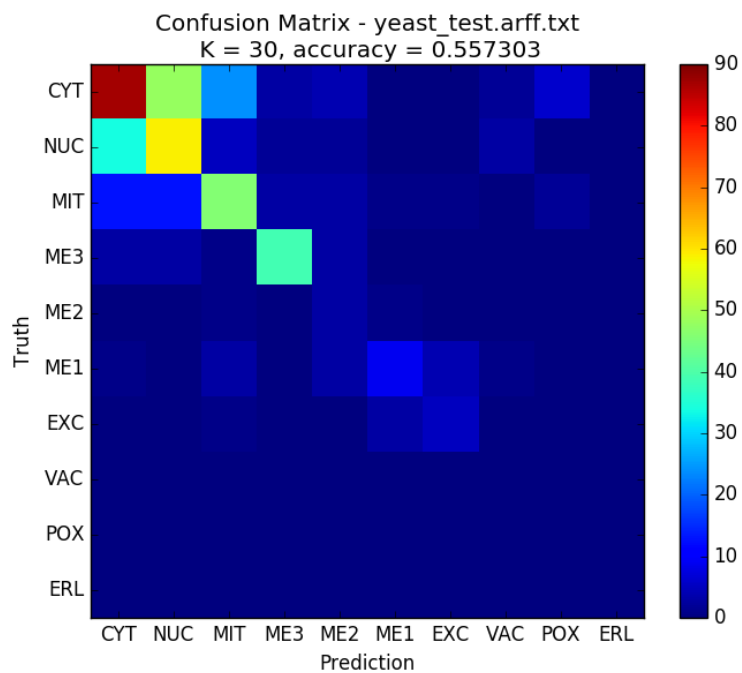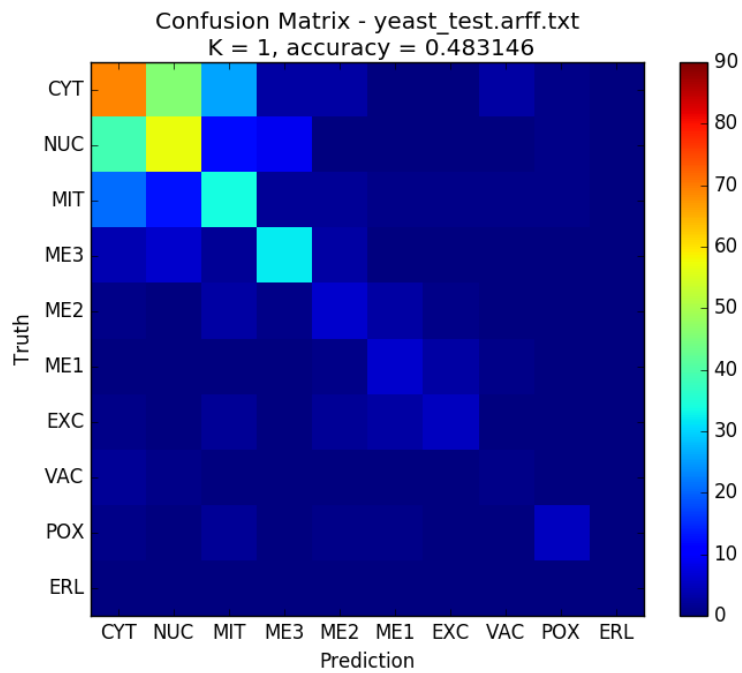
- pyhton 2.7

- numpy

- scipy

- sys

QUESTION2

For the yeast data set, draw a plot showing how test-set accuracy varies as a function of k. Your plot should show accuracy for k = 1, 5, 10, 20, 30.

For the wine data, draw a similar plot showing test-set mean absolute error as a function of k, for k = 1, 2, 3, 5, 10.
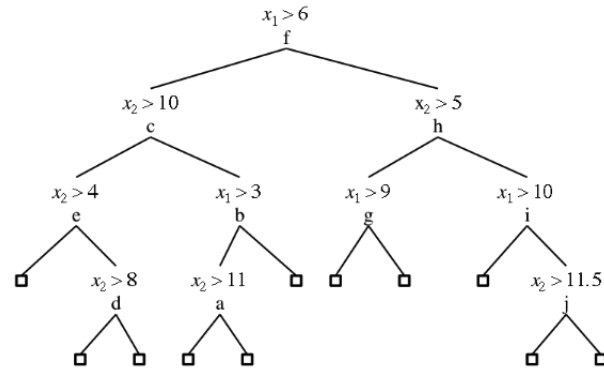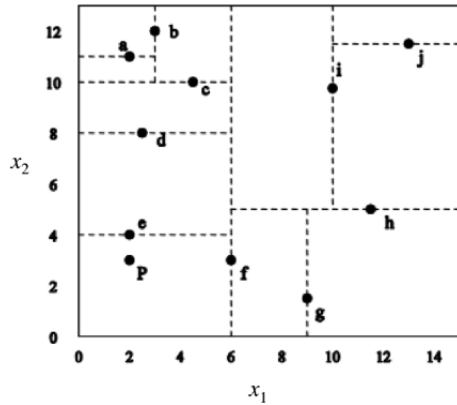
For the yeast data set, construct confusion matrices for the k = 1 and k = 30 test-set results. Show these confusion matrices and briefly discuss what the matrices tell you about the effect of k on the misclassifications.

Confusion Matrix - yeast_test.arff.txt
K = 1, accuracy = 0.483146

Confusion Matrix - yeast_test.arff.txt
K = 30, accuracy = 0.557303

QUESTION3

Using the k-d tree and the training set displayed in the figure below, show how the nearest neighbor for $x^{(q)} = (7, 10)$ would be found.

Distance	Best Distance	Best Node	Priority Queue
Monday	11C	22C	A clear day with lots of sunshine. However, the strong breeze will bring down the temperatures.
Tuesday	9C	19C	Cloudy with rain, across many northern regions. Clear spells across most of Scotland and Northern Ireland, but rain reaching the far northwest.
Wednesday	10C	21C	Rain will still linger for the morning. Conditions will improve by early afternoon and continue throughout the evening.