ECE 760 Homework 2: kNN

Qihong Lu

October 18, 2016

QUESTION1

Code:

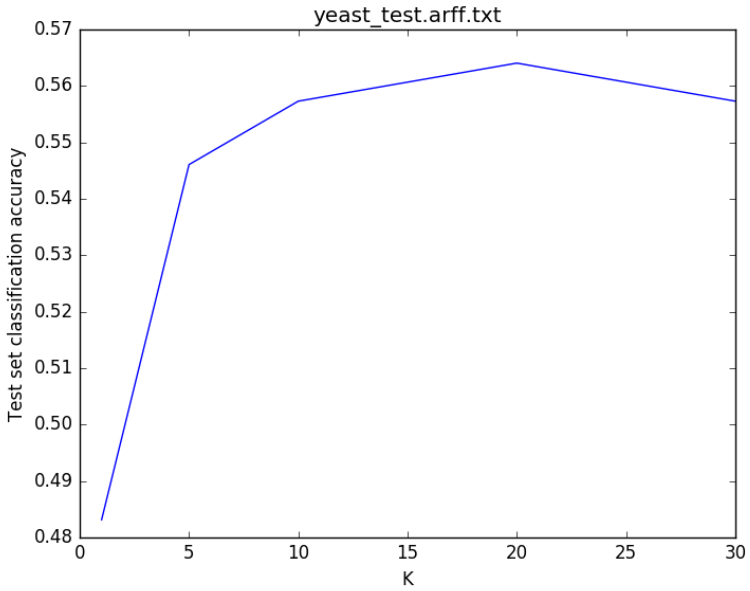
- kNN_alg.py: the main program that implements the standard kNN for multi-class classification and regression
- util.py: the definitions of some constants and helper functions
- kNN.py: run kNN algorithm, return test results
- kNN-select.py: run kNN algorithm, tune K by leave-one-out validation. Fit final model with the best K and return the test results

Dependencies:

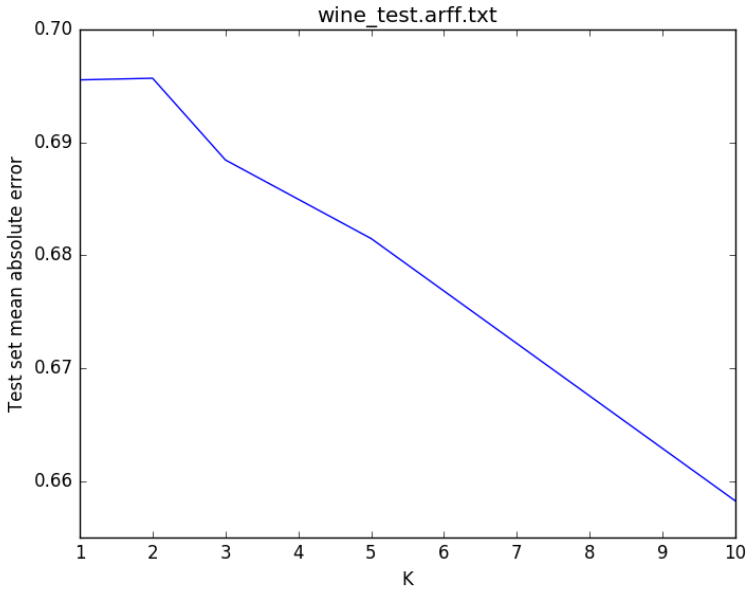
- pyhton 2.7
- numpy
- scipy
- sys

QUESTION2

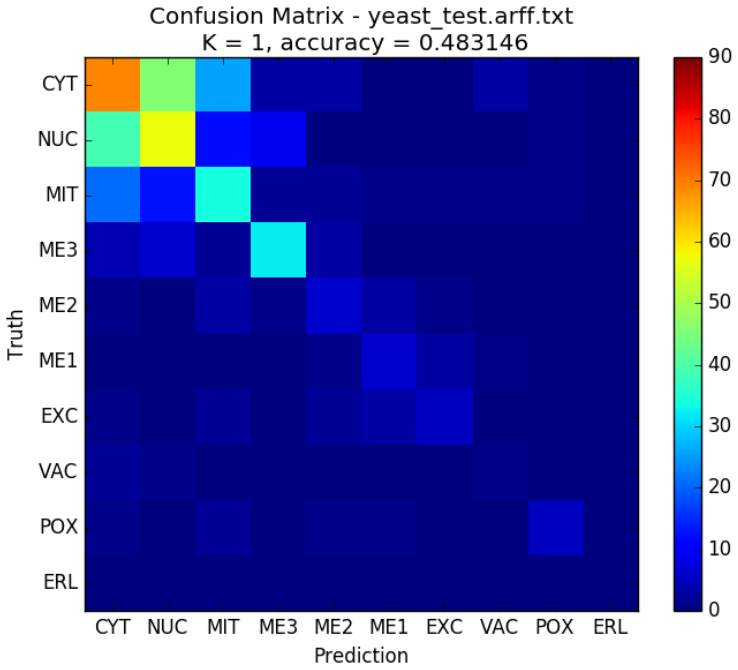
For the yeast data set, draw a plot showing how test-set accuracy varies as a function of k. Your plot should show accuracy for k = 1, 5, 10, 20, 30.

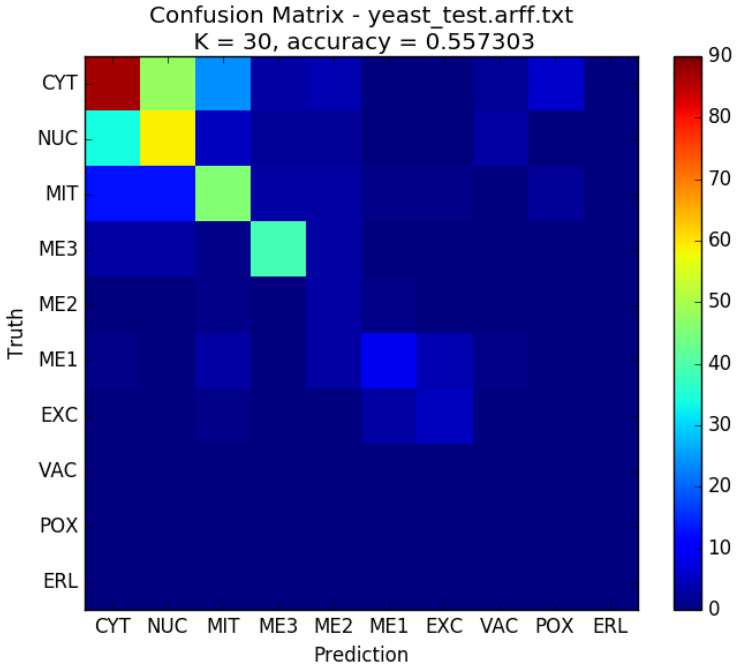


For the wine data, draw a similar plot showing test-set mean absolute error as a function of k, for k = 1, 2, 3, 5, 10.



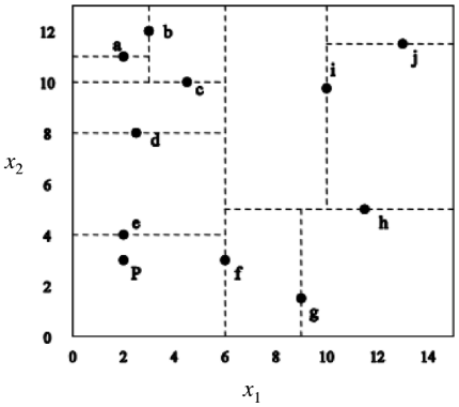
For the yeast data set, construct confusion matrices for the k = 1 and k = 30 test-set results. Show these confusion matrices and briefly discuss what the matrices tell you about the effect of k on the misclassifications.

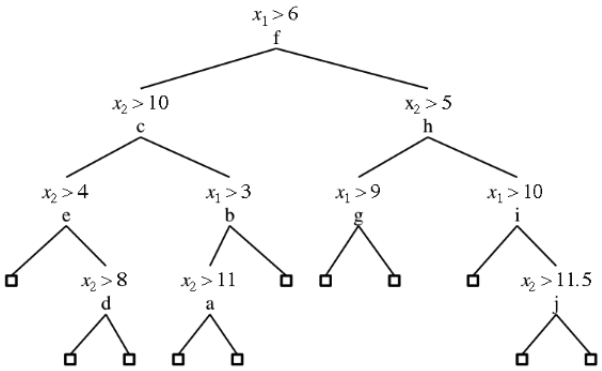




QUESTION3

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.





Action	Distance	Best Distance	Best Node	Priority Queue
NA	∞	∞	NA	(f, 0)
Pop f	8	8	f	(h, 0) (c, 1)
Pop h	8	10	f	(i, 0) (c, 1) (g, 5)
Pop i	3	3	i	(c, 1) (j, 3) (g, 5)
Pop c	2	2	c	(j, 3) (g, 5)

Therefore, the algorithm returns the node c.