

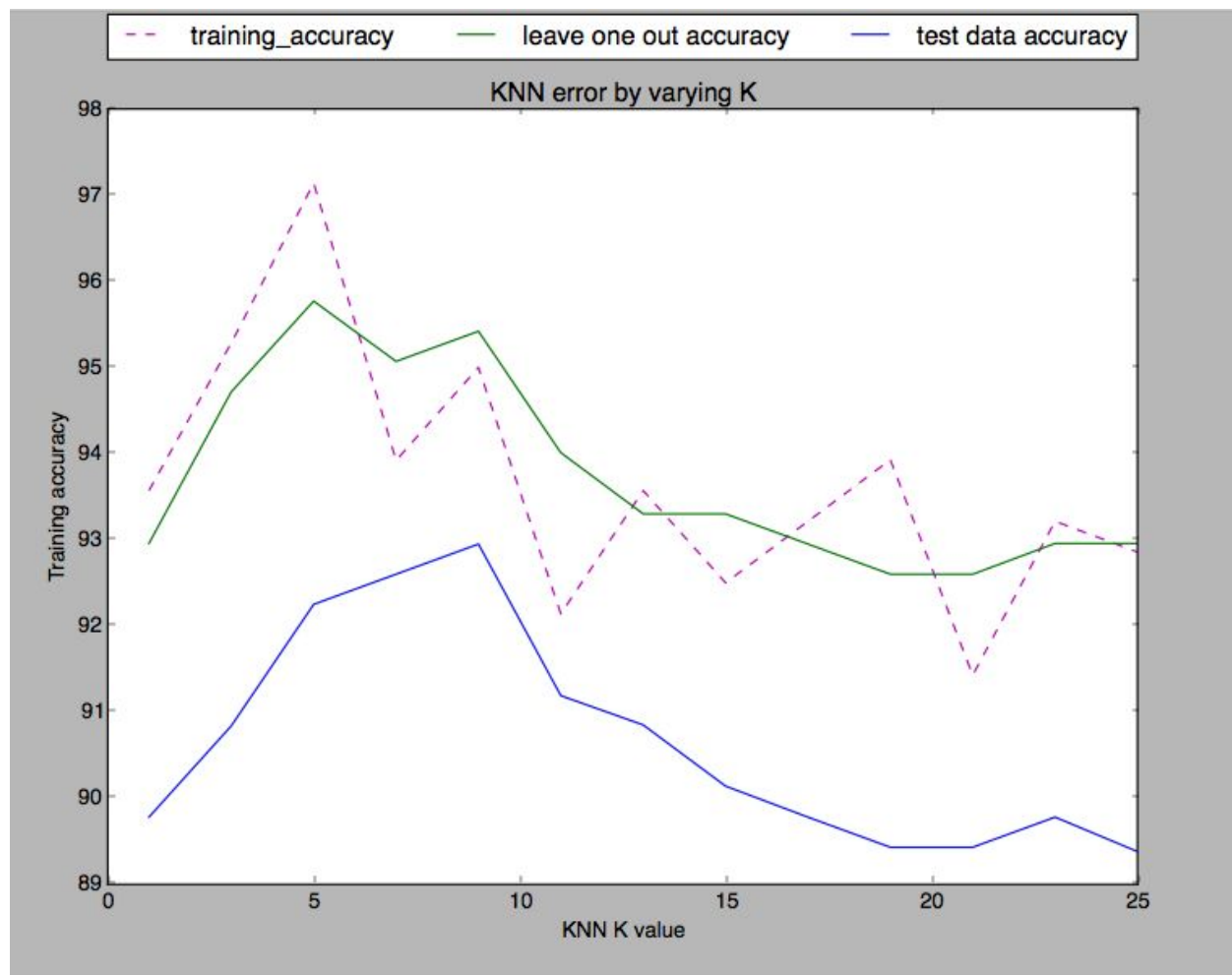
KNN and Decision Tree  
Neale Ratzlaff  
CS 434

1)

I performed leave-one-out cross validation and varied K from 1 to 25 in order to see which amount of decision boundaries yielded the best results. From this test, K = 5 gave the smallest error with an accuracy on the training set of >97%

2)

I computed the accuracy of the model on the training set, using leave-one-out cross validation, and using the test set. Not surprisingly, the testing set showed the smallest error, with the testing set giving the largest error, with around a 8% differential between the two.



3)

The test set showed the largest error, with a lower bound at about 89%. This is not surprising, as the test data will always have data that the model has not seen before. This makes the training set invalid for testing purposes, because the model has already been taught

specifically how to handle that data. This is shown above with the training accuracy being much higher than the other two. Leave one out validation is slightly lower than the training accuracy, because it still isn't dealing with completely new data.

## Part II

a)

b)

a)

b)

The decision tree will always be able to classify correctly all the training examples, because it can generate as many test nodes as it needs to, in order to create the necessary decision boundaries. Optimality may not be reached with a greedy algorithm, because the greatest information gain at any given step may not be the globally optimal step at that point.