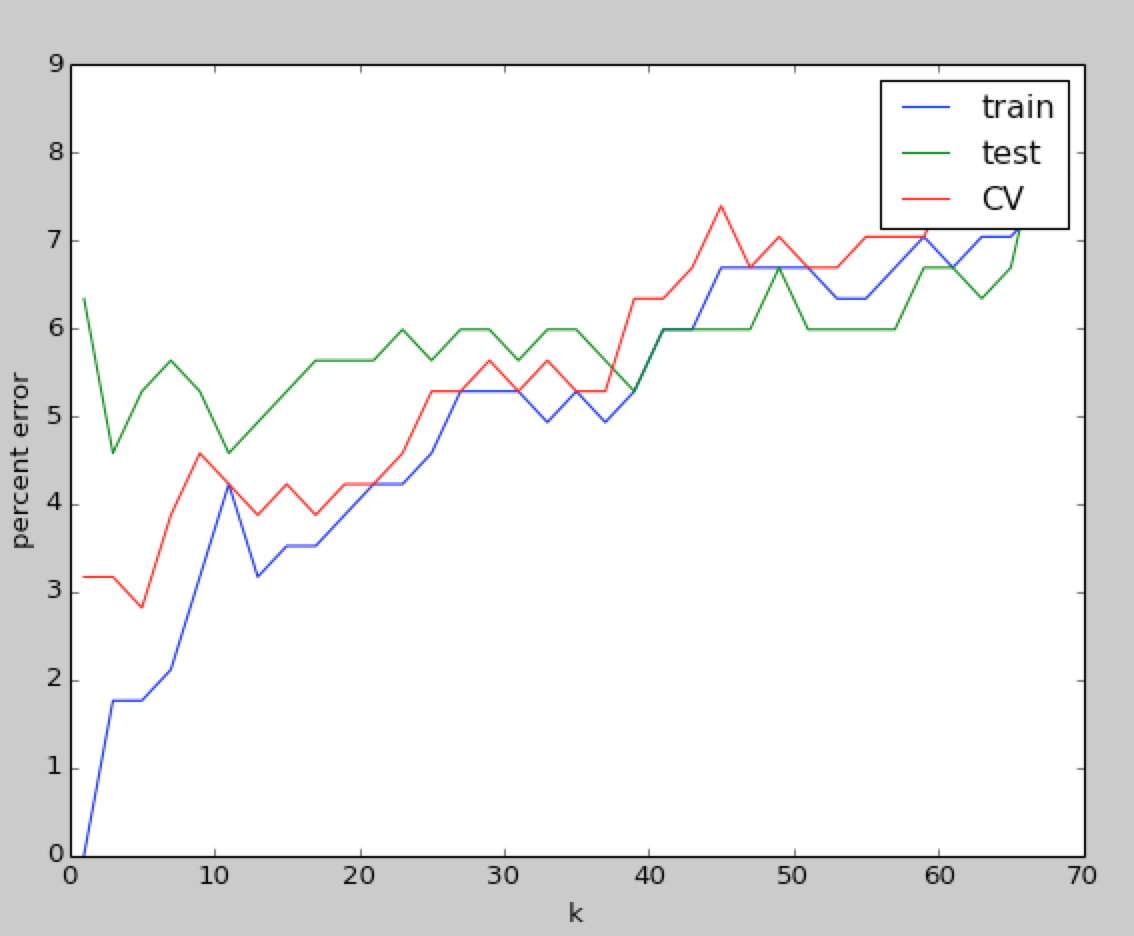
Alex Nguyen

Michael Lee

CS434

Assignment 2 - Logistic regression with L2 regularization

1. Part 1 – KNN
   1. Implemented algorithm – done.
   2. Graph of K vs. Error of Training, Test, Cross-validation error
   3. We observe that as K get larger that the error rate rises. Our training data is perfect when compared to itself and with a K of 1. The error rate rises pretty rapidly there after. Our test data's error rate stays approximately 6 percent error for the first 70 values of K. Our leave-one-out cross-validation error rate has a somewhat linear climb similar to the training set. Looking at the graph, a acceptable K value would be approximately 35 values of K.
2. Decision Trees
   1. Decision stump:
   2. Decision tree with depth 6:
      1. Tree with info gain per depth level is in tree.txt file
      2. Train error: 65.8450704225
      3. Test error: 71.4788732394
      4. Compared to the decision stump rates, the decision tree does not do much better. This could probably be the cause of over fitting to the training data or having too many levels in our tree.
3. Using decision tree rates to improve KNN
   1. By knowing which features splits the data according to the tree. We can use those features for the KNN algorithm to improve the division of neighbors. By improving neighbor hoods we can achieve a higher prediction result than using either KNN or decision trees alone.