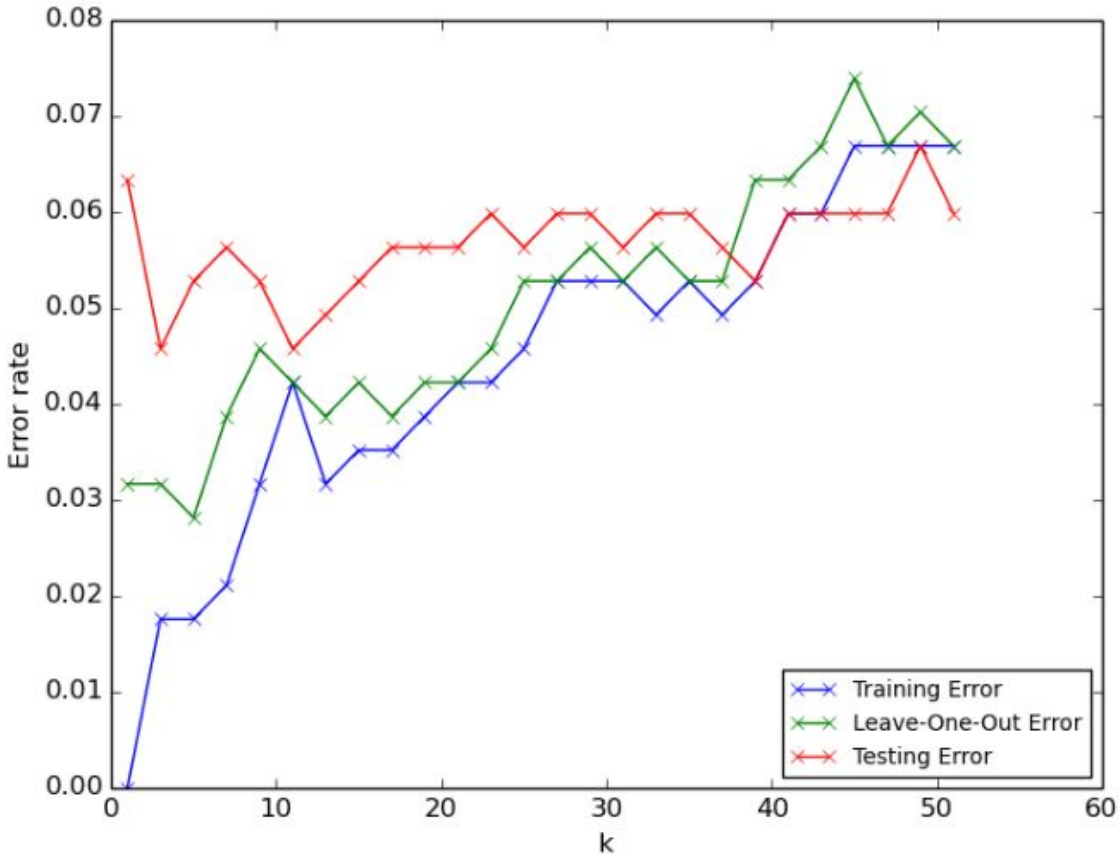
Monica Sek and Arthur Shing
CS 434
Implementation Assignment 2
4/30/2018

1. Model selection for KNN

1.1. Look at code.

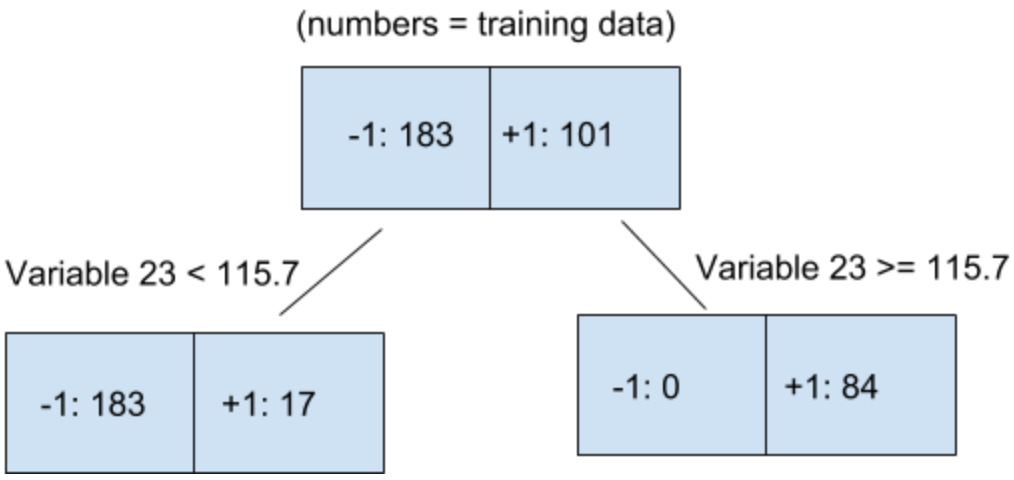
1.2.



1.3. Both the training error and the leave-one-out error increases in error rate as K increases while the testing error stayed at a consistent rate. In the end, all three of the testing errors share a similar end error rate at the highest K Our choice of K is K=5 because the leave-one-out error is the lowest at k=5, and that's when the three testing errors relatively have their lowest error rate.

2. Decision Tree

2.1.

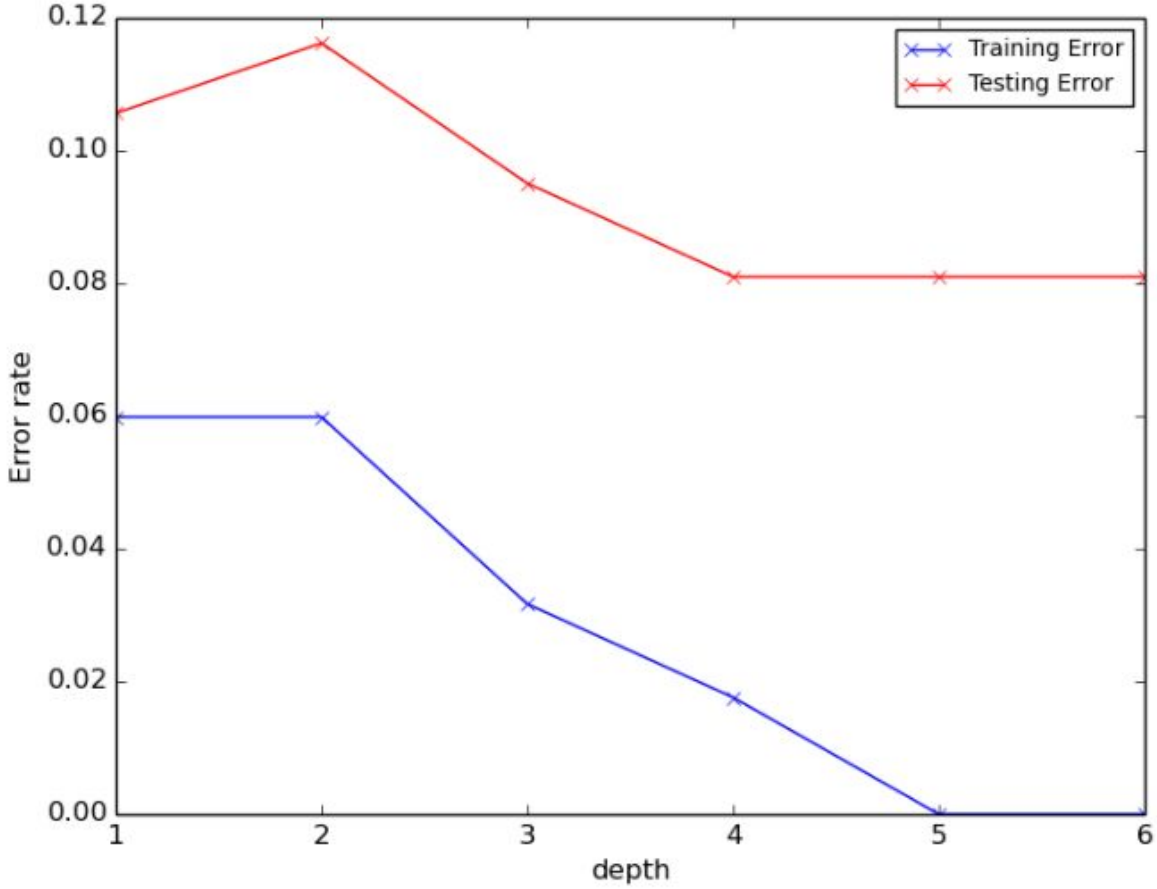


Computed information gain: 0.6435

Training error rate: 5.98%

Testing error rate: 7.04%

2.2.



The more depth we added to the tree the more accurate the results were, this would be ideal because as we teach the program, we would hope that it becomes more accurate. The training error eventually gains 100% accuracy, this is because we used that specific data set to create the thresholds for each depth which would be ideal for that data set. The testing error spikes at depth = 2 because of overfitting.

3. Extra Credit

3.1. KNN can used classifications given by decision trees to avoid neighbors that are in a different class. This will provide a more accurate prediction for KNN.