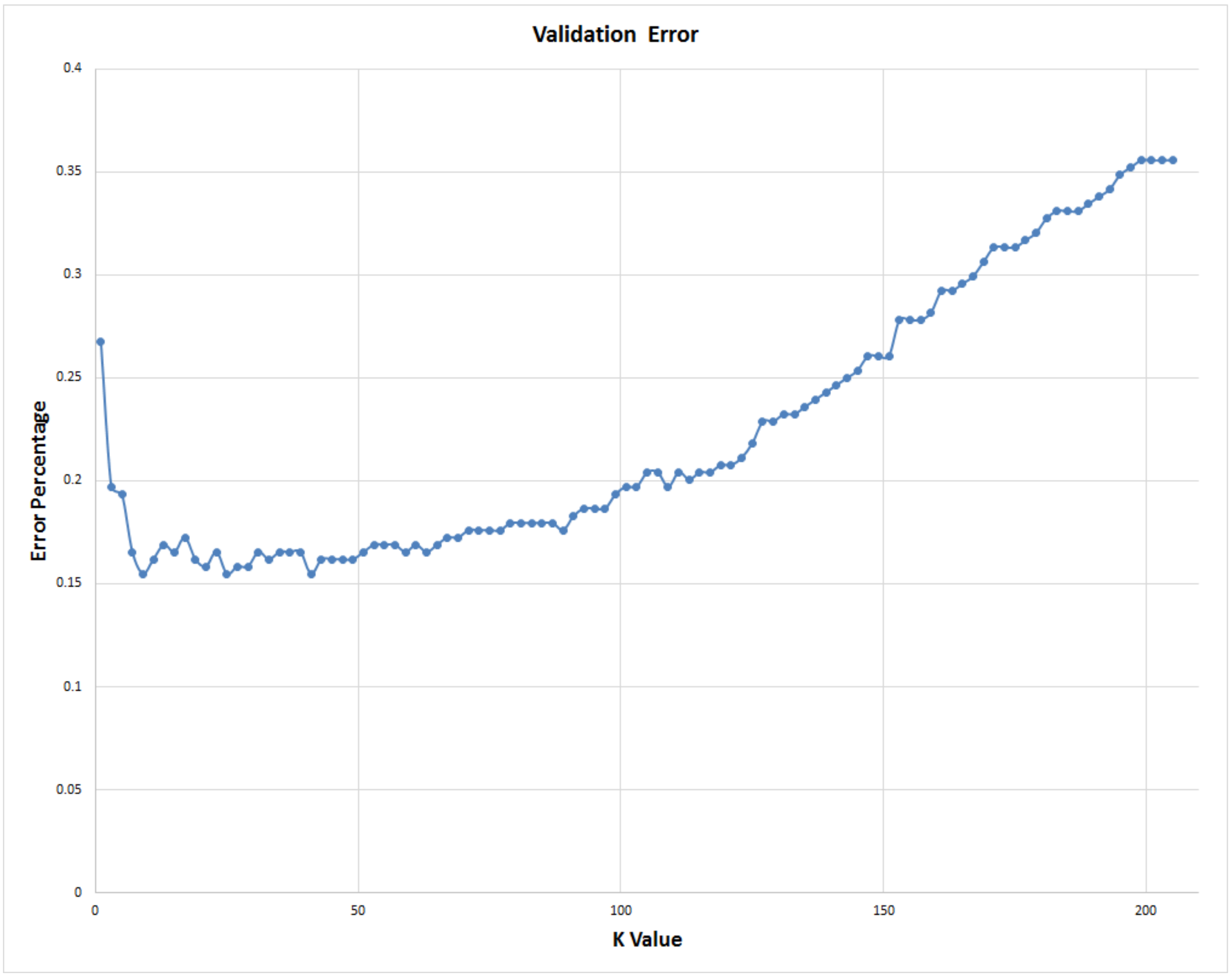
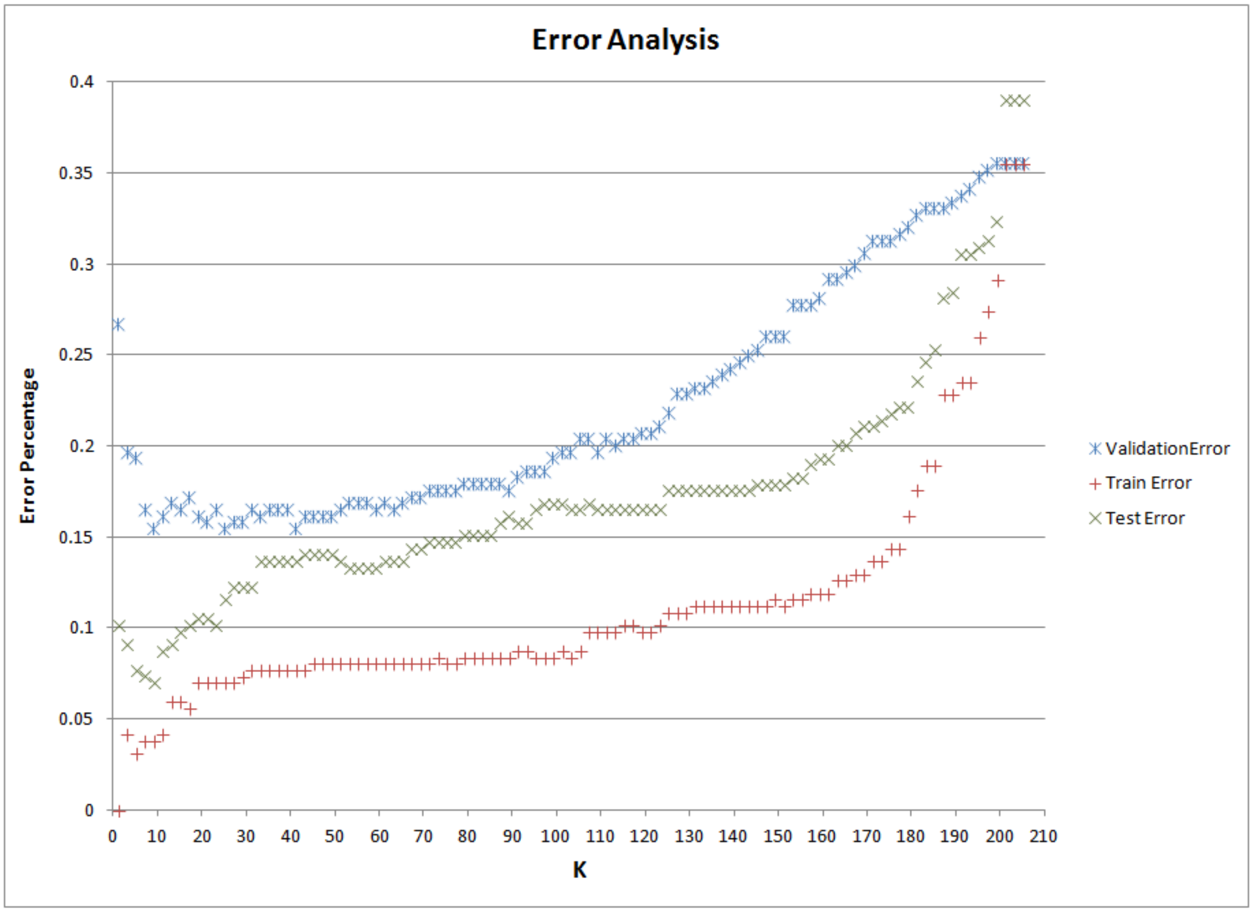
CS434 - Assignment 3

Nathan Woodworth
Sean Hammond
Stephanie Ison

Part 1

Using leave-one-out cross-validation to select the best K from tested values of 1, 3, 5, …, 205, the K value we chose is 9 with the smallest validation error.





The the relationship between the three different errors is that they all follow a similar path. Between K = 30 and K = 100 the errors increase by small amounts until each eventually begins to increase by larger amounts. From this we can conclude that when K is too large there is a greater chance for errors. And that as K gets larger the margin of those errors increases as well. It is also noticeable that the errors for the Testing data was higher than that of the Training data, which makes it seem like the Training data might not have been a good set for the Testing data.

Part 2

a) Both the learned stump and the learned decision tree. To help grading easier, please provide for each selected test its information gain.

Learned stump:

x5 = 1: gain 0.28620076122076754
| y = 1.0
x5 != 1
| y = 0.0

Decision Tree:

```
x5 = 1: gain 0.28620076122076754
|  y = 1.0
x5 != 1
|  x1 = 1: gain 0.04658020448553368
| |  x2 = 1: gain 0.7320666900931938
| | |  y = 1.0
| |  x2 != 1
| | |  y = 0.0
|  x1 != 1
| |  x2 = 1: gain 0.4208613886842806
| | |  y = 0.0
| |  x2 != 1
| | |  x5 = 3: gain 0.020501317658662366
| | | |  x2 = 2: gain 0.10434897109477148
| | | | |  x1 = 2: gain 0.7219280948873623
| | | | | |  y = 1.0
| | | | |  x1 != 2
| | | | | |  y = 0.0
| | | |  x2 != 2
| | | | |  x1 = 2: gain 0.9852281360342516
| | | | | |  y = 0.0
| | | | |  x1 != 2
| | | | | |  y = 1.0
| | |  x5 != 3
| | | |  x4 = 1: gain 0.048621352634209636
| | | | |  x1 = 2: gain 0.10218717094933333
| | | | | |  y = 1.0
| | | | |  x1 != 2
| | | | | |  x2 = 2: gain 0.7219280948873623
| | | | | | |  y = 0.0
| | | | | |  x2 != 2
| | | | | | |  y = 1.0
| | | |  x4 != 1
| | | | |  x6 = 1: gain 0.0597731301493174
| | | | | |  x3 = 1: gain 0.19087450462110944
| | | | | | |  y = 1.0
| | | | | |  x3 != 1
| | | | | | |  x1 = 2: gain 0.4199730940219749
| | | | | | | |  x2 = 2: gain 0.9182958340544896
| | | | | | | | |  y = 1.0
| | | | | | | |  x2 != 2
| | | | | | | | |  y = 0.0
| | | | | | |  x1 != 2
| | | | | | | |  y = 0.0
| | | | | x6 != 1
| | | | | | x5 = 2: gain 0.15200728380562722
| | | | | | | y = 1.0
| | | | | | x5 != 2
| | | | | | | x3 = 1: gain 0.109170338675599
| | | | | | | | x1 = 2: gain 0.01997309402197489
| | | | | | | | | x2 = 2: gain 1.0
| | | | | | | | | | y = 1.0
| | | | | | | | | x2 != 2
| | | | | | | | | | y = 0.0
| | | | | | | | x1 != 2
| | | | | | | | | x2 = 2: gain 0.9182958340544896
| | | | | | | | | | y = 0.0
| | | | | | | | | x2 != 2
| | | | | | | | | | y = 1.0
| | | | | | | x3 != 1
| | | | | | | | y = 1.0
```

b) The training and testing error rate of the learned stump and full decision tree.

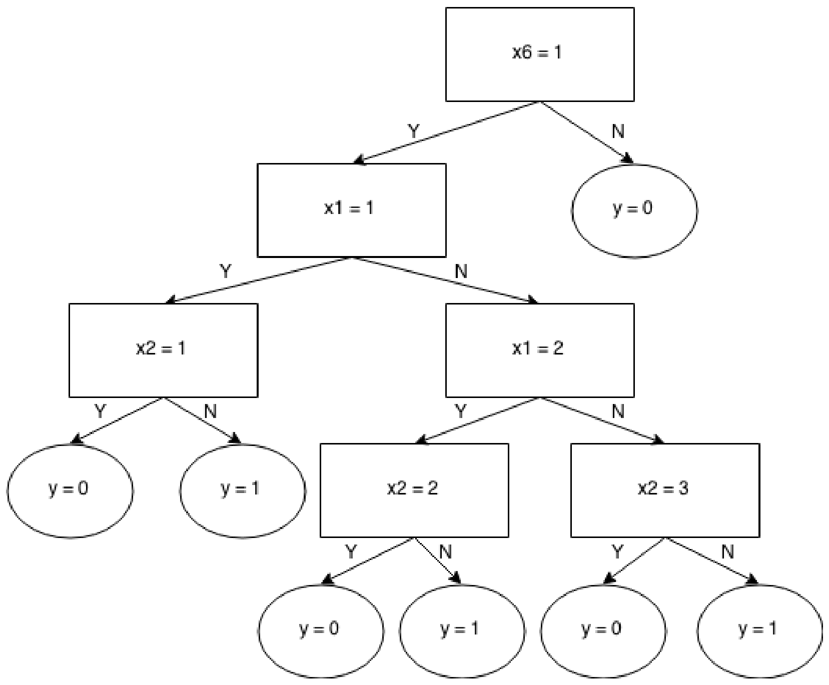
Stump Training Error: 0.2661290322580645%

Stump Testing Error: 0.25%

Tree Training Error: 0.0%

Tree Testing Error: 0.07407407407407407%

a) Given the formula for generating the class label, please provide a (as compact as possible) decision tree that will correctly classify all training examples



b) Do you expect the top-down greedy information algorithm to learn this optimal tree? Note that your answer should be general, not depending on the specific training set you use in this assignment.

The top-down greedy information algorithm's tree may not be the optimal tree. Greedy algorithms by nature pick the option that is locally optimal, the best choice at that given time, and in the case of the top-down greedy information algorithm this eventually leads to overfitting when the data split is not very large. If the tree is too specific to the training data, and it contains noise and outliers it is likely to perform poorly on testing data.