(1.3 Questions)

Question 1)

\*Note: Code for results are at testingStuff.py

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Tennis Data | | Sentiment Data | |
|  | Train | Test | Train | Test |
| AlwaysPredictOne | 0.642857 | 0.5 | 0.504167 | 0.5025 |
| AlwaysPredictMostFrequent | 0.642857 | 0.5 | 0.504167 | 0.5025 |
| FirstFeatureClassifier | 0.714286 | 0.666667 | 0.504167 | 0.5025 |

Question 2)

None, AlwaysPredictMostFrequent and AlwaysPredictOne has same performance for both Datasets

Question 3)

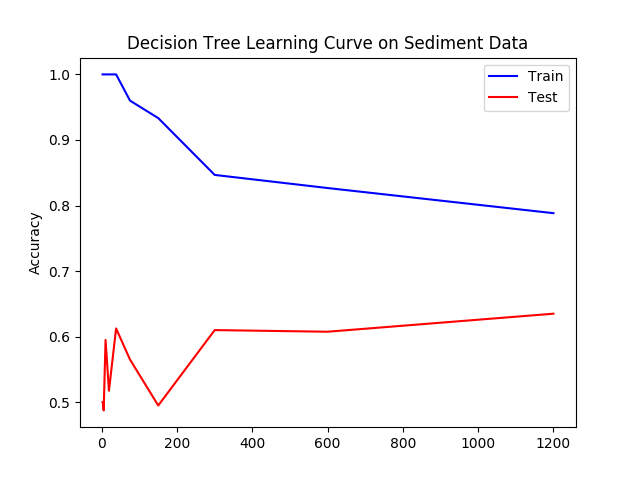
Tennis Data, is where FirstFeatureClassifier outperforms AlwaysPredictOne

Question 4)

The second line computes the training accuracy or average of the values of that are Y > 0 and X > 0;   
 because it gets the average of the points where the labels y are greater than 0 AND prediction of X is   
 greater then 0; then averages

Question 2 Decision Tree)

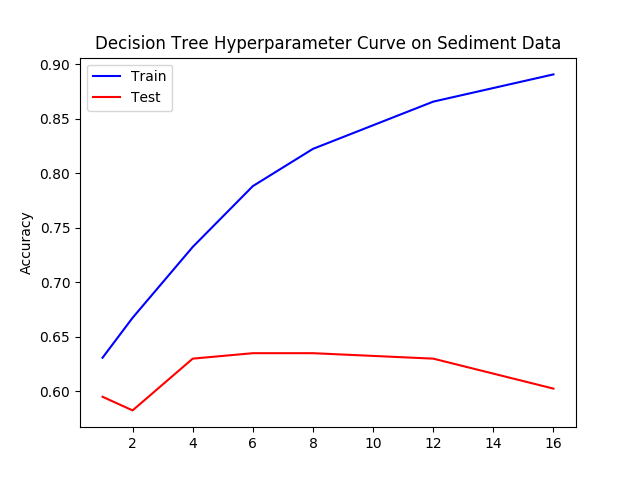
5)



\*Note: DT tested on testDecisonTree.py

6) There reason why the training data goes down is because we are limited to a certain depts. With lower data do test on, we can see that the predictions have a high accuracy, but as we increase the amount of data, it starts to decrease, and at one point, it starts to decrease at a steady rate. The test data starts to increase as we get as we get more data. It is not monotonically because there are times when the accuracy decreases and then decreases.

7)



8) There reason why the Training set is monotonically increasing is because we have more depth to train our data, and give higher accuracy. The reason we it is monotonically increase is because we start to overfit the training set. We can see that around depth 12 is when we start starts to show how overfit the data is because as training accuracy increases, the test data accuracy starts decreasing and is no longer able to generalize.

3 Nearest Neighbor)