Yuechen Zhao
1126125
zzyc1993@uw.edu
CSE 446 Assignment 3

1. SVM

(1.1)

These two points are (1, 0, 0) and (1, 2, 2) in 3D space respectively.
One vector that is parallel to w is (0, 2, 2).

(1.2)

The distance between the two points = $\sqrt{8}$
So margin = $\sqrt{8} / 2 = \sqrt{2}$

(1.3)

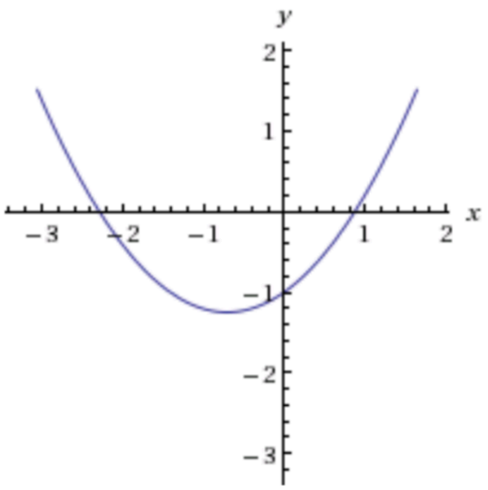
Suppose w = (a, b, c), then a = 0, and b = c, and also $\sqrt{2} = 1 / ||w||$.
Therefore, w = (0, ½, ½).

(1.4)

Since the inequalities are tight, then we have:
$w \cdot (1, 0, 0) + w0 = -1$
$w \cdot (1, 2, 2) + w0 = 1$
Therefore, w0 = -1

(1.5)

$f(x) = -1 + (0, ½, ½) \cdot (1, \sqrt{2}x, x^2)$
$f(x) = ½ * x^2 + \sqrt{2}/2 * x - 1$



2. Ensemble Methods

(2.1)

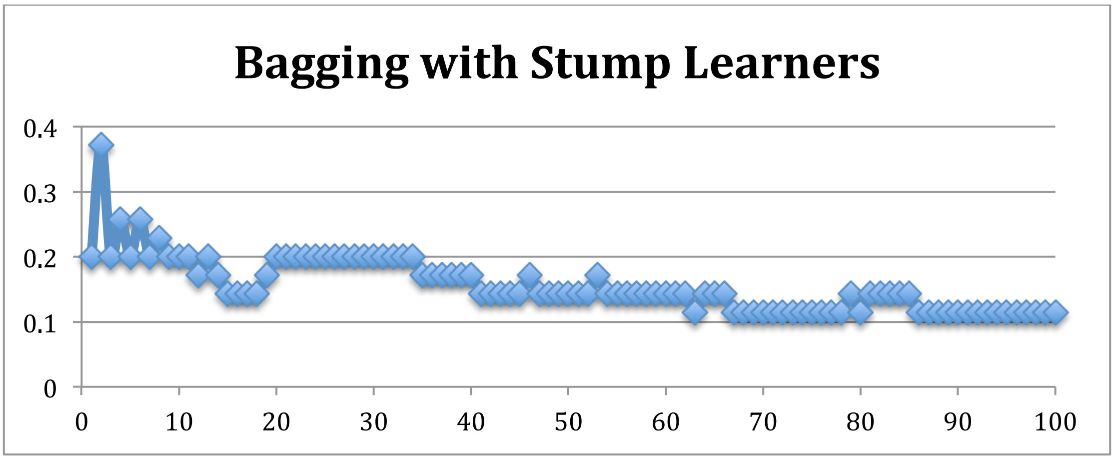
1. I will do the learning by using scikit-learn's DecisionTreeClassifier, and by specifying the max_depth parameter I can distinguish stump and depth 2 tree. Thus I can use the classifier object to fit my X_train and Y_train, and then predict Y_test by passing in X_test, and then I can get the error of the prediction.

2. Previously when we have uniform weights, we calculate the entropy as
$$H(Y) = -\sum_{i=1}^{k} P(Y = y_i) \log_2 P(Y = y_i)$$
But now instead of using the first P(Y = yi), we use weights (Dt) instead.

   Also for the conditional entropy
$$H(Y \mid X) = -\sum_{j=1}^{v} P(X = x_j) \sum_{i=1}^{\kappa} P(Y = y_i \mid X = x_j) \log_2 P(Y = y_i \mid X = x_j)$$
We also use the weights (Dt) instead.

(2.2)

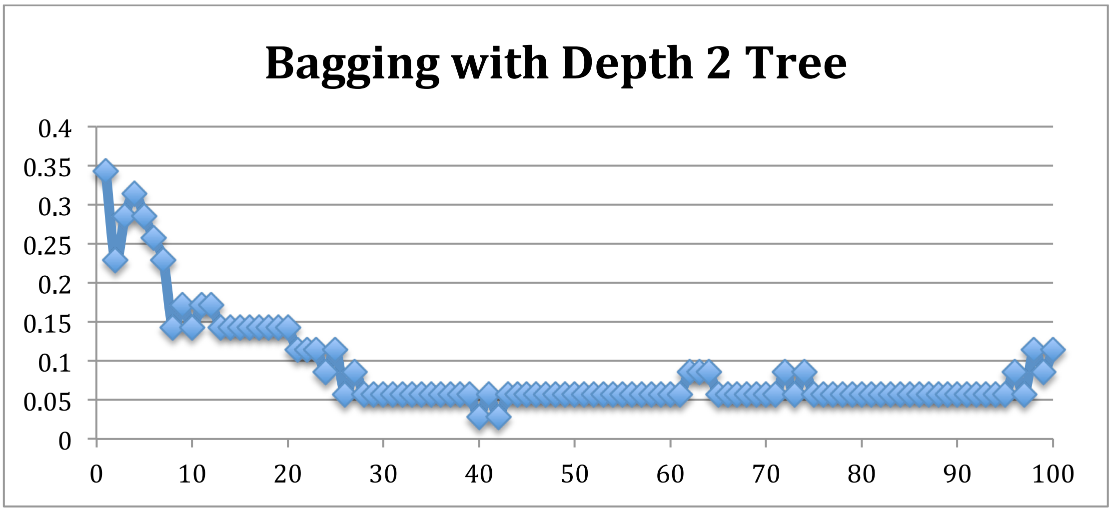
1.



In this algorithm, I used 100 iterations. In each iteration, I randomly select 50 training data, train a classfier, predict my Y_test, add to the previously predicted data, and then calculate the average. If the average prediction > 0.5, I predict Y=1, otherwise predict Y=0. Then I can calculate the error for this iteration.

This algorithm's advantage is it randomly select data points out of the whole set of training data, so it is likely it will leave out the outliers, and predict more accurately. However, also because of its randomization, it has fluctuating test errors, as shown above. Though test error has some fluctuation, it goes down generally.

2.



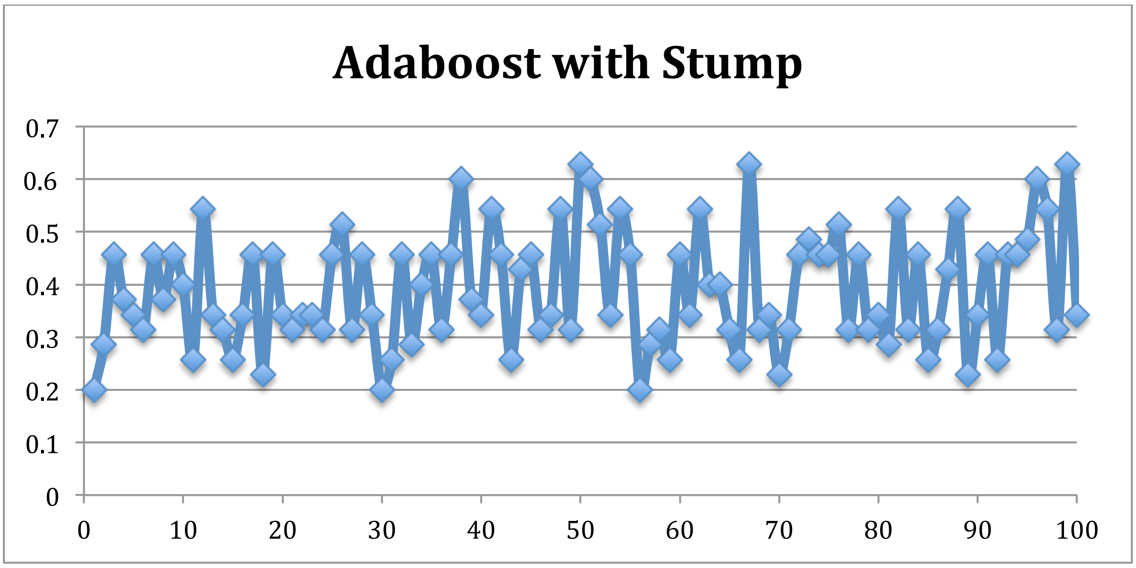
In this algorithm, I did almost the same thing, except I used depth 2 tree instead of stump. This algorithm generates slightly lower test errors than using stumps, and its result tends to vary slightly more than using stumps. However, I think it still needs some significance tests to determine whether this algorithm is significantly better than the previous one, or the other way around.

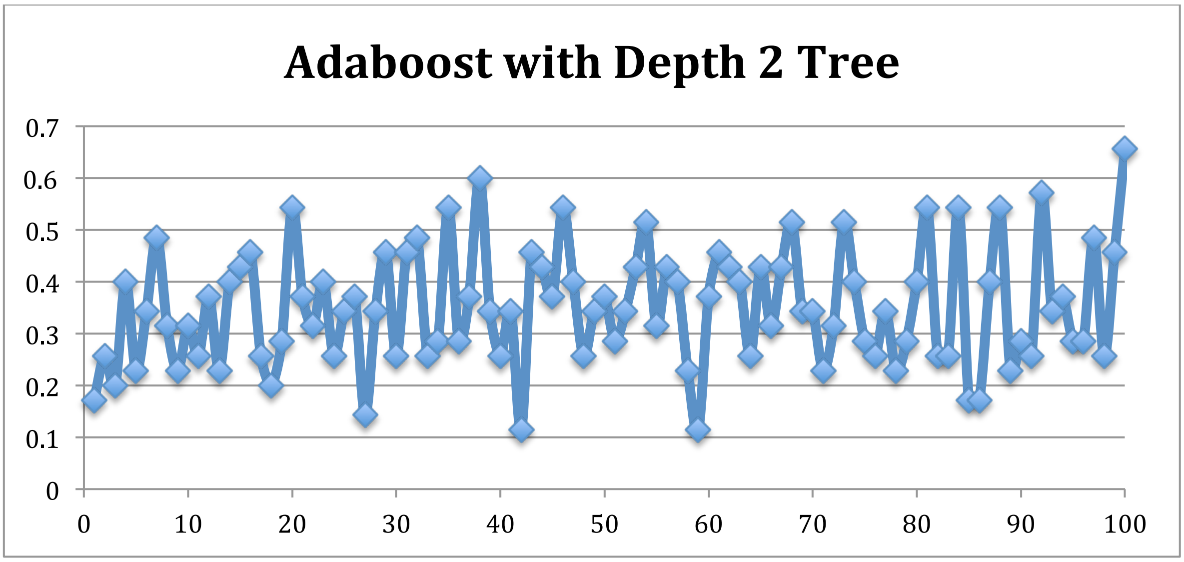
3.

Instead of using M = 50 as I did previously, I changed to M = 71 (the total number of training data), calculate the unique number of data in each iteration, and then repeat for 100 iterations. The average result = 63.7887323944%, which is very close to the theoretical value.

(2.3)

1.





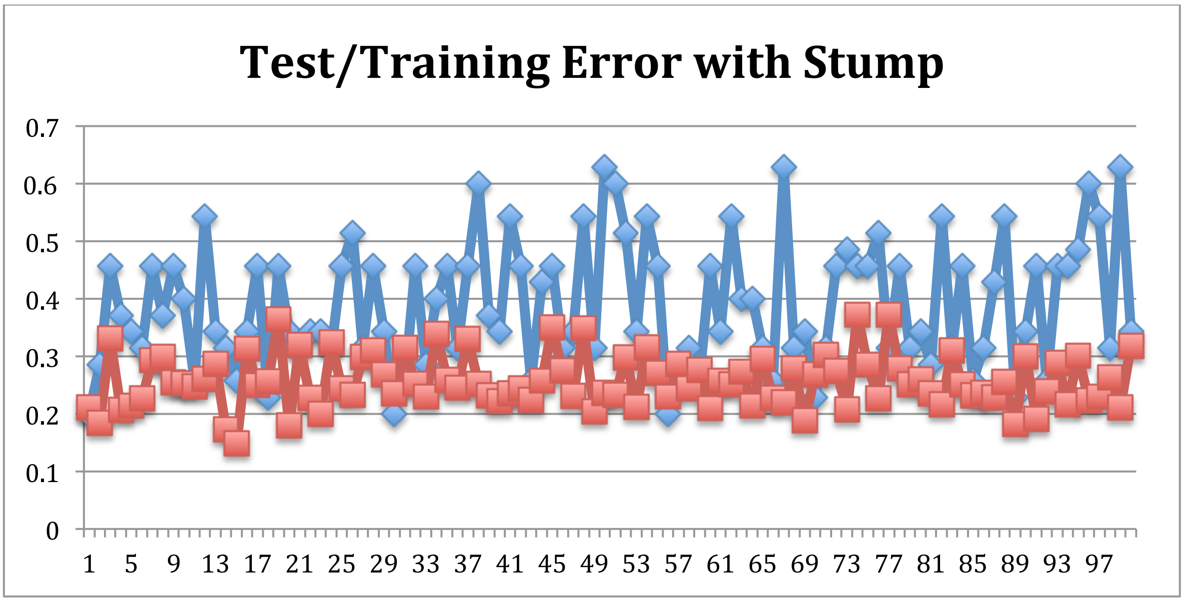
2.

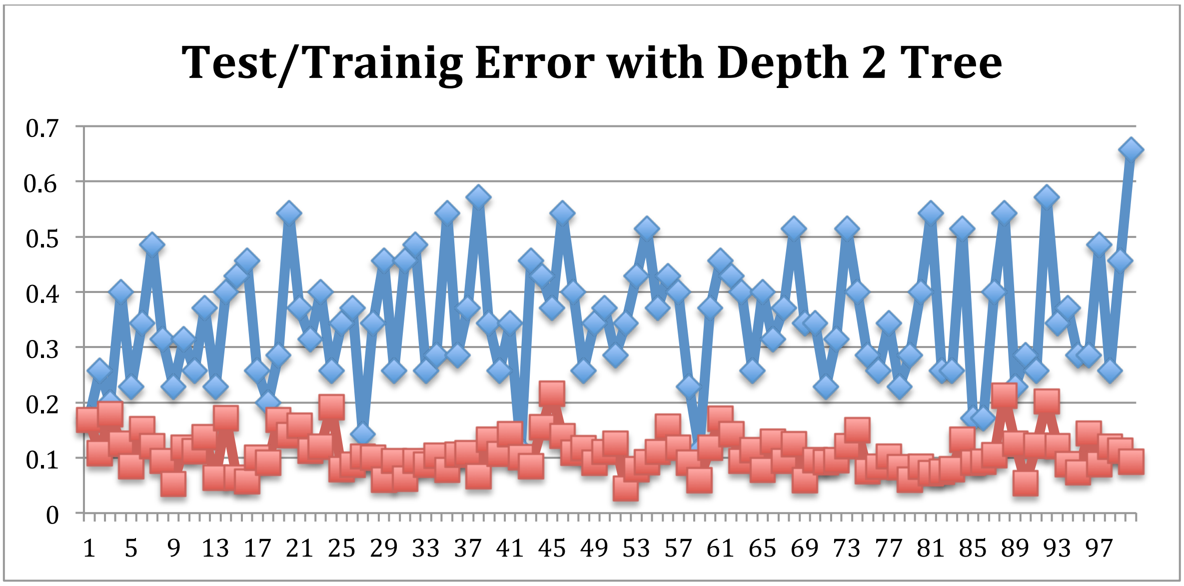
These two base learners tend to have roughly equal performance. It is a tie because in each iteration, we weigh the data, then learn it using stump or depth 2 tree, thus the weights would dominate the learning and prediction. Therefore, the two base learners does not have much difference in this case, since the weights play a more important role.

3.

To test for overfitting, I did 1000 iterations using both base learners, and drew the results in charts, as shown below the first 100 iterations. The blue (upper) ones are test errors, and the red (lower) ones are training errors. Although errors tend to fluctuate more vigorously with more iterations, there is no significant sign of overfitting in both cases, because there is not a time when test error goes up significantly while training error goes down significantly.

One significant difference between these two algorithms is that, depth 2 tree has lower training error than stump. Other than that, these two base learners have very similar results.


Test/Training Error with Stump
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97


Test/Trainig Error with Depth 2 Tree
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97