Jiaming Chen
CSE 446
HW3

1.1

$\langle 0,2,2 \rangle = w$ parallel to W

1.2

margin = $\mathrm{sqrt}(2^2+2^2) = \mathrm{sqrt}(2)$

1.3

$\mathrm{sqrt}(2) = 1/\mathrm{norm}(W)$
$W = \langle 0,2a,2a \rangle$
$1/2 = 8a^2$
$1/16 = a^2$
$a = 1/4$

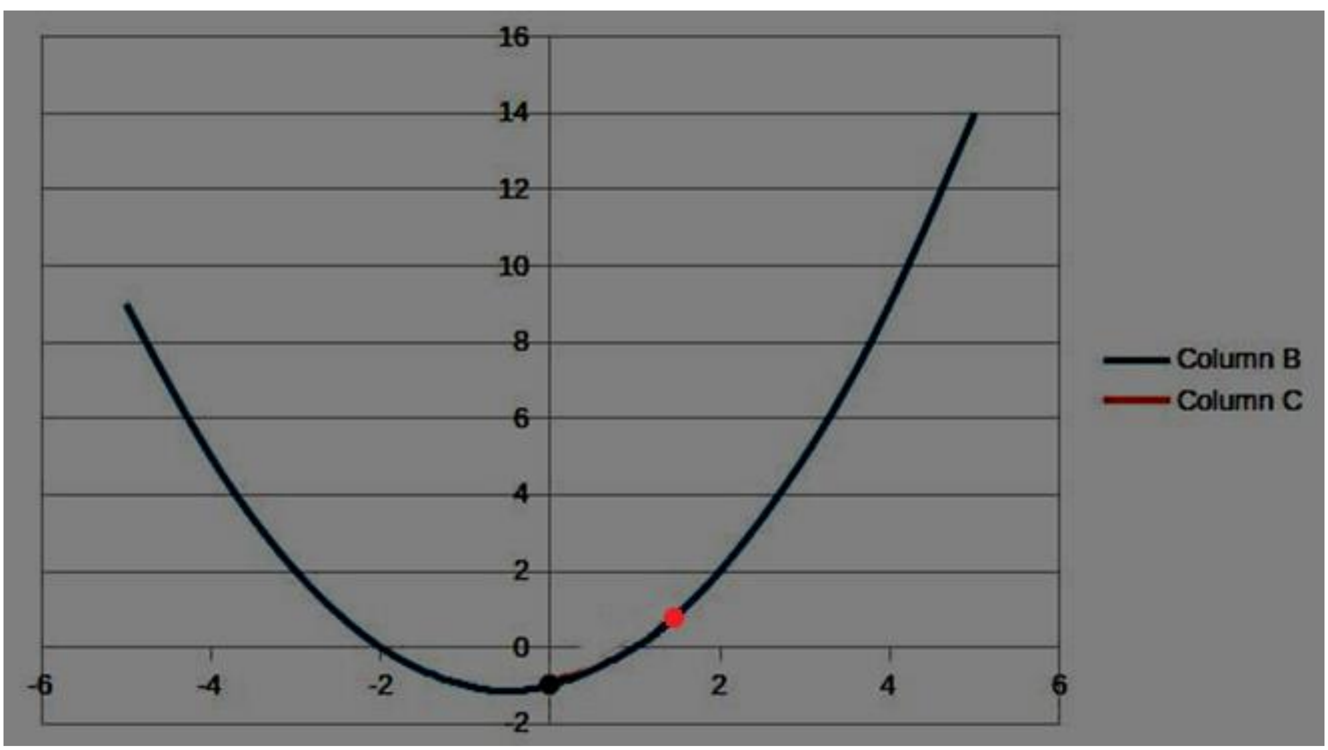
$W = \langle 0,1/2,1/2 \rangle$

1.4

$y_1(W * \Phi(x_1) + w_0) >= 1$
$y_2(W * \Phi(x_2) + w_0) >= 1$
$W*\Phi(x_1)+w_0 <= -1$
$W*\Phi(x_2)+w_0 >= 1$
$w_0 <= -1$
$2+w_0 >= 1$
$w_0 = -1$

1.5

$f(x) = -1 + \mathrm{sqrt}(2)*x/2 + x^2/2$



2.1.1

I will use DecisionTreeClassifier in the python sklearn library. For stump, I will set max_depth = 1; for depth two tree, I will set max_depth = 2. Both classifier will use entropy as criterion. This criterion will select split that gives the best information gain. And both of them are continuous because I set the feature to have value 1 to 4.

2.1.2

IG = H(Y) – H(Y|X)
H(Y) = $-\sum_{i=1}^{k} P(Y = yi)log2P(Y = yi)$
H(Y|X)= $-\sum_{j=1}^{v} P(X = xj)\sum_{i=1}^{k} P(Y = yi|X = xj)log2P(Y = yi|X = xj)$

P(Y=yi) in the above equation is $\sum_{j=1}^{n} D(j)\delta(Yj = yi)$

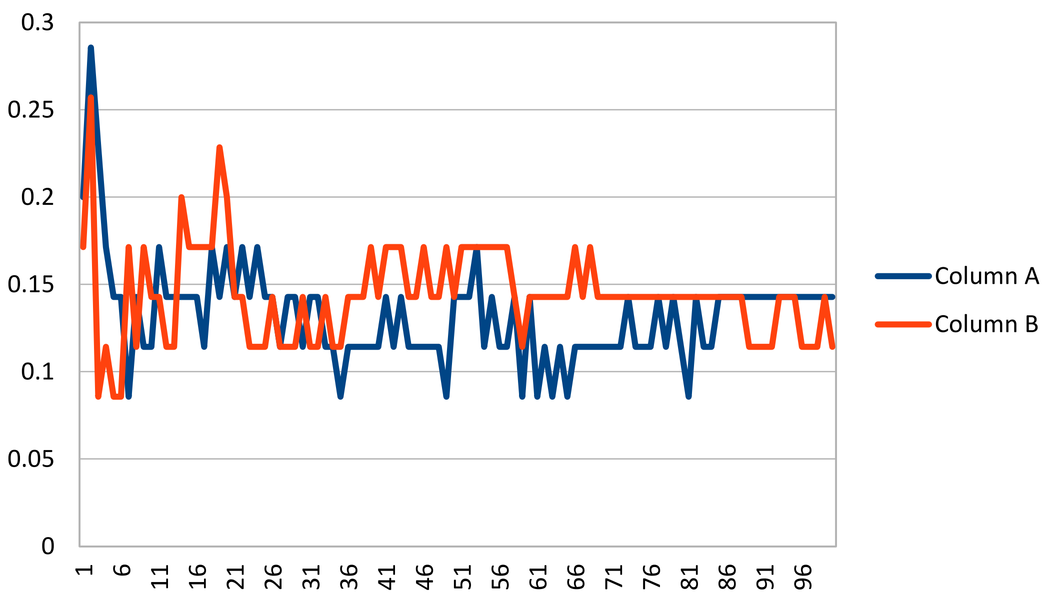
P(Y=yi|X=xj) in the above equation is P(Y=yi,X=xj)/P(X=xj)
P(Y=yi,X=xj) = $\sum_{k=1}^{n} D(j)\delta(Yk = yi,$Xk=xj)
P(X=xj) = $\sum_{k=1}^{n} D(j)\delta(X = xj)$

2.2.1

Column A is for decision stump. The error converge toward 15%, and the best case is about 8% error. The error jumps up to 30 % at the beginning because the stump is a simple model, but the prediction gets more accurate as it gets more votes.

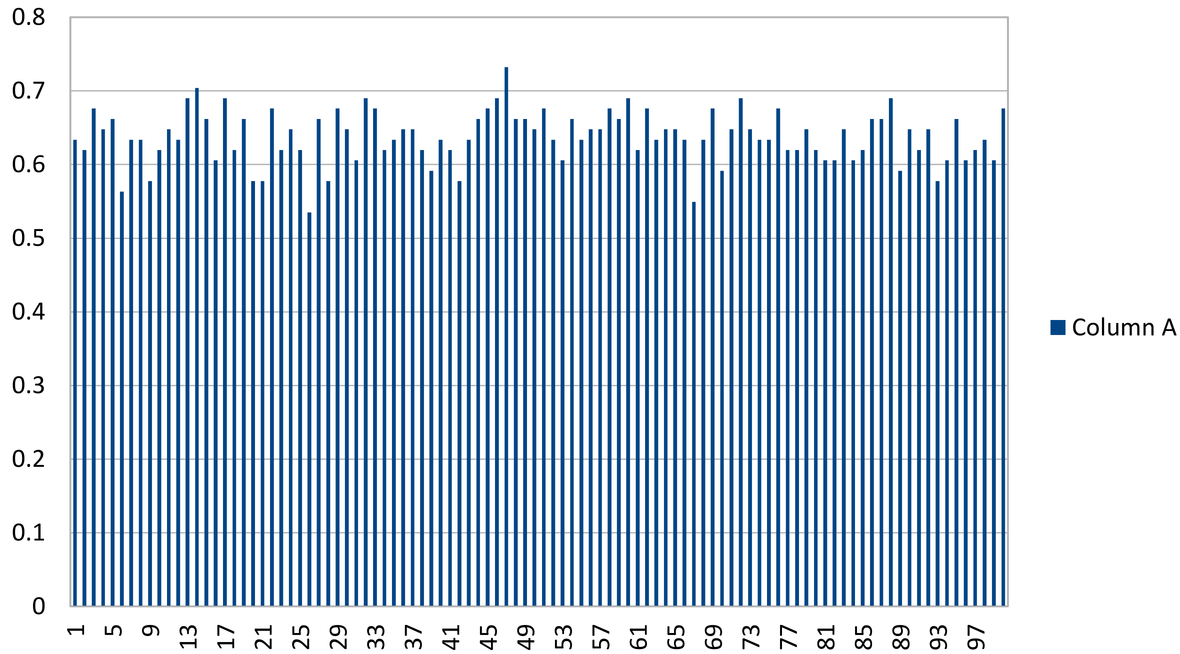
2.2.2

Column B is for Depth 2 tree. It is almost the same as decision stump, converging toward 11% - 15%. And the best case is also around 8 %. The error doesn't jump as high at the beginning because depth two tree is more complicated, but in the long term, both stump and depth two have similar predictions.



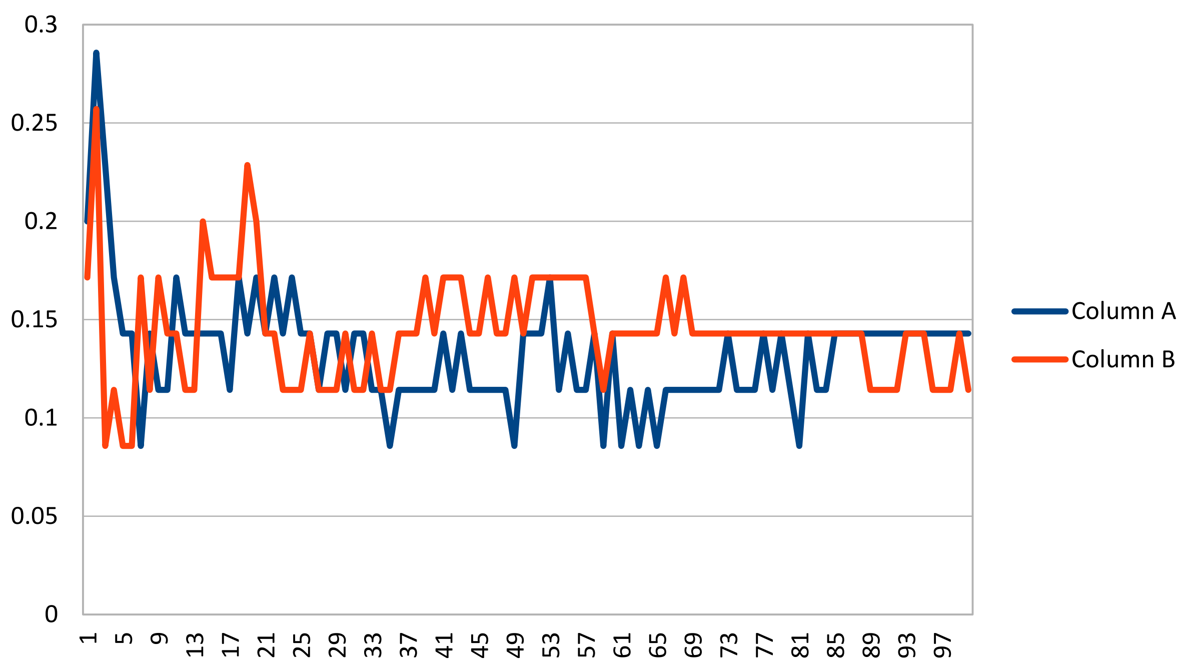
2.2.3

In this experiment, I ran randint to sample from N numbers range from 0 to N-1, and used a set to record all the numbers. The uniqueness is calculated by size of set / N. In 100 iteration, the mean is 63.8% unique. The worst case is 53.5% unique, and the best case is 73.2% unique. The result is very close to the expected uniqueness 63.2%.



Column A is stump, B is Depth 2.

2.3.1



2.3.2

It is a tie because their predictions have similar accuracy. This result happens because the solution for the data set is not very complex, which can be represented easily by stump or depth two tree. Beyond that, we can see that depth two tree converges faster than stump at 0 to 6 iterations, which suggests depth two learner might be better if we have more complex data. But considering how quickly both classifiers converge, there is no obvious reason to claim one to be better than another.

2.3.3

I ran adaboost and looked for an iteration when training error decreases and testing error increases. (Errors after the 40th iteration stays the same, so I only plot the first 40 iterations.) Even though I can find some iterations that have that pattern, the pattern doesn't last long.

The stump gives a smoother error graph unlike the error in depth 2 that jumps a lot more, but the best prediction with stump classifier(8% error) is worse than the best prediction with depth 2 tree(6% error). For depth 2 classifier, the algorithm makes good prediction at iteration 4 to 6, and the error increase quickly afterward. This could be the sign of overfitting. However, the error rate comes down eventually. Therefore, there is no obvious sign of overfitting in stump or depth 2 tree.

Series 1 is Training
Series 2 is Testing