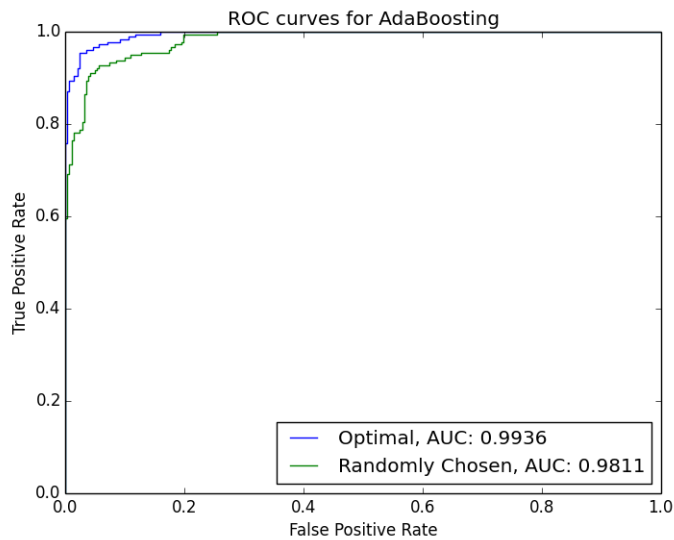# CS6140 Machine Learning Fall 2014 Homework 4, Wei Luo

PROBLEM 1

Weak Learning via "Optimal" Decision Stumps: accuracy is 0.9609
Weak Learning via "Randomly Chosen" Decision Stumps: accuracy is 0.9370
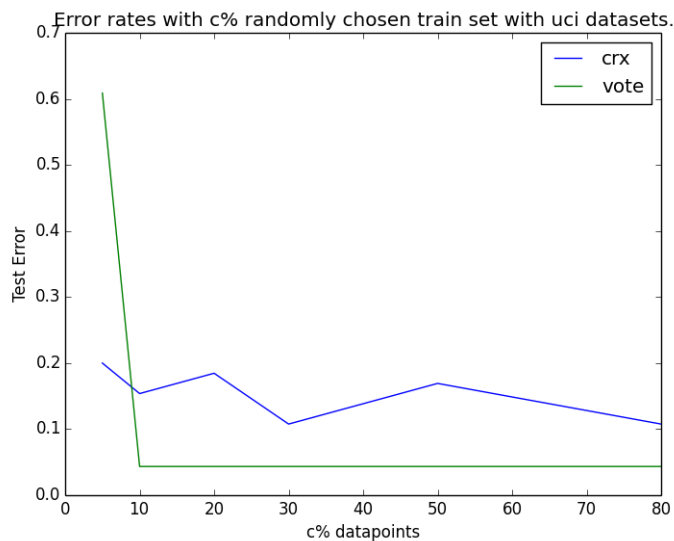The ROC curves for these result:

PROBLEM 2

Run adaboosting with 10 folds cross validation on crx dataset, average accuracy is 0.866923
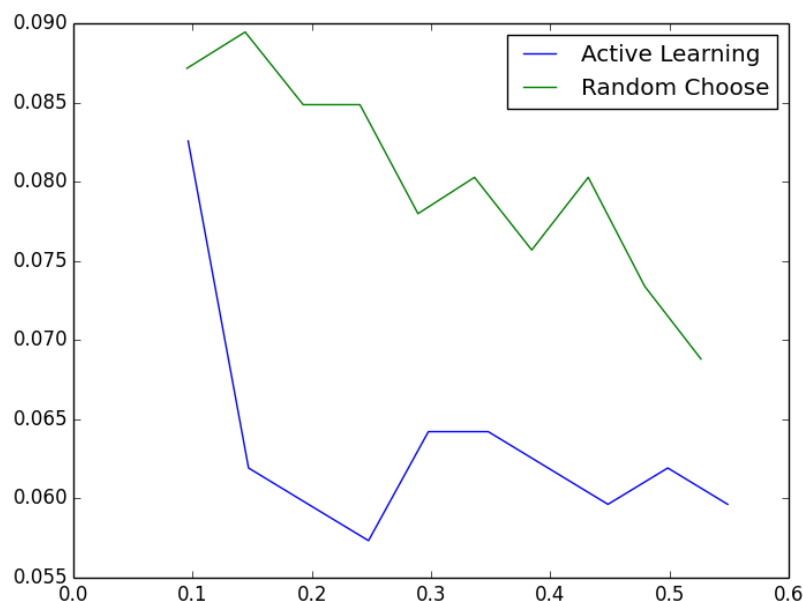Run adaboosting with 10 folds cross validation on vote dataset, average accuracy is 0.961594
Try adaboosting with different size of sample training dataset on crx and vote, err rates are:

PROBLEM 3

The results for active learning and randomly built on spambase dataset:

PROBLEM 4

Train the data with 20 random selected functions, 200 rounds for each adaboosting iteration, I obtained a result of accuracy 0.756107275624 .

PROBLEM 5

a) VC dimension for unions of two rectangles is 17.
Consider the similar case in c), for triangles, since it has 3 sides, it can cut through 7 points on a circle which can have at most 3 continuous blocks clockwise. Similarly, for rectangles, the 4 sides can shatter 9 points which can have at most 4 continuous blocks clockwise. For two rectangles, since they have 8 sides, the maximum number of points that can be shattered is 17.
b) VC dimension for circles is 3.
Since any 3 points can be placed such that they can form a triangle. And for any labels assignment on that triangle vertices, we can always find a circle to shatter them.
For 4 points, if there is a point inside the convex hull of the others. We can not shatter them with a circle if it is labeled $+$ and all others labeled $-$. Otherwise, we cannot shatter them if they are labeled $+,-,+,-$ clockwise.
c) VC dimension for triangles is 7.
Given 7 points on a circle, for any labeling, each class can have at most 3 continuous blocks clockwise. With a triangle, we can always use the three sides to cut those blocks.
For 8 points, if there is a point inside the convex hull of the others. We can not shatter them with a triangle if it is labeled $+$ and all others labeled $-$. Otherwise, we cannot shatter them if they are labeled $+,-,+,-,+,-,+$ clockwise.
d) VC dimension for multidimensional "sphere" $f(x) = sing[(x - c)(x - c) - b]$
in m dimensions space $\mathbb{R}$ is $m + 1$.
Similarly to the circles, on a 2 dimensional space, 3 points determine a circle, this m-dimensional sphere could be determined by $m + 1$ points.
For $m + 1$ points, we can arrange them on a m-dimensional sphere determined by them. Once

we do so, we can always find a sphere to shatter them no matter how they are labeled, just like the case when m=2.

For $m+2$ points, by Radon's theorem, any set of $m+2$ points can be partitioned into 2 disjoints whose convex hulls interest. Thus, for any set of points, if they are labeled that way, no sphere can shatter them.

PROBLEM 6

The accuracy of bagging with 50 decision trees (depth 5) is 0.9369.
The accuracy of bagging with 50 decision trees (depth 7) is 0.9435.