Assignment 2: Aug 18

Instructions: *You are free to code in any language of your choice. Your code should ask to input training set and output a classifier upon execution. Discussion among the class participants is highly encouraged. But please write your own code to make sure that you understand the algorithms. Submit the code and report by th 6th September 2017*

Question 1 (Classification problem) *Apply the following algorithms on the given classification dataset*

- *CART*

- *C4.5*

- *Random Forest: Set number of trees to $N = 10, 20, 30, 50, 100$. To generate each tree in the random forest, use random subsamples as follows: generate a number m' selected uniformly at random in $\{1, 2, \ldots, |S|\}$. Then, select m' number of data points (with replacement) uniformly at random for training. The attribute set can remain the same for each tree. Plot the test error as a function of number of trees in each random forest.*

HRA Dataset	training error	test error
CART		
C4.5		
Random Forests		

Note: You are expected code each algorithm. Do not use any readily available packages.

Each dataset is divided into training data and testing data. Train using training data and test on test data.

For the random forest, report the best test error you obtain over different choices of N in the table.