Read Me File for Nishant Subramani and Jeanette Pranin’s code.

There are two code files that you must run in a specific order. The first builds the tree(s) and outputs almost all of the relevant information. The second reads in two csvs produced by our first function and plots a graph.

Step 1) Install numpy library in python

Step 2) Install R, the statistical programming language gui from <http://cran.r-project.org/doc/manuals/r-release/R-admin.html>

Step 3) Put the train, validation, and test files into the same directory as the .py file and .R file I have submitted. The names of these files must be put into the function.

The first function PS2-Nishant-Subramani.py can be called by issuing two commands.

execfile(“PS2-Nishant-Subramani.py”)

decision\_tree(‘btrain.csv’ , ‘bvalidate.csv’ , ‘btest.csv’)

Note: The file names must be put in the proper order. The first argument is the string name of the training file, the second argument is the string name of the validation file, and the third is the string name of the test file. The last line of the .py file has a commented line showing exactly how to call that last line for the data given. That has been reiterated above.

Note: This function takes 4023 seconds on my computer as plotting the learning curve takes a long time. The learning curve portion of the function starts when "Learning Curve Data for Full Trees is being Generated, and put into a csv" is printed. At this point, if pressed for time, you can break out of the function and cancel the build and use the data provided in two csv files: noprunelc.csv and prunelc.csv. These files contain the csv files utilized by the .R code provided and plot the learning curves for a specific random subset trial done on my computer earlier.

Note: The output should look like the output attached in output.txt. It will not be a file, but rather a print out on the console. I put this printout into a text file for your convenience.

The second function can be called by using 4 commands and pause between the 3rd and 4th commands on the Rgui directly.

setwd("C:\\Users\\Nishant\\Documents\\Nishant\\Northwestern\\Spring Quarter 2015\\EECS 349\\Homework 2")

source(“PS2-Nishant-Subramani.R”)

learningcurve(“noprunelc.csv”)

learningcurve(“prunelc.csv”)

Note: The string in the setwd command should be the directory that my learningcurve.R file is in. On my computer that is the working directory.

Note: Pause between the 3rd and 4th commands as the function plots the learning curve for each of the trees, with pruning and without.