# CS 6375: Machine Learning (Spring ’19)

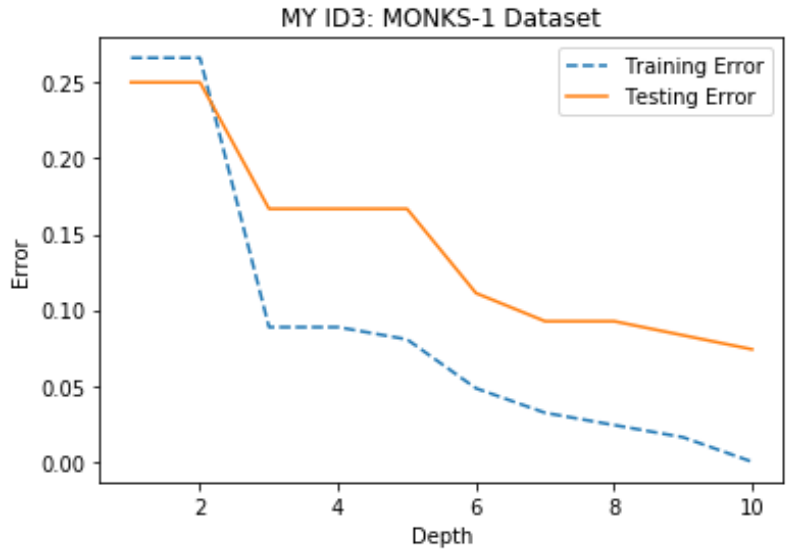
# PROGRAMMING ASSIGNMENT – 1

Submitted by:

skt180001 - Safal Tyagi

vxv170013 - Vamshider Reddy Voncha

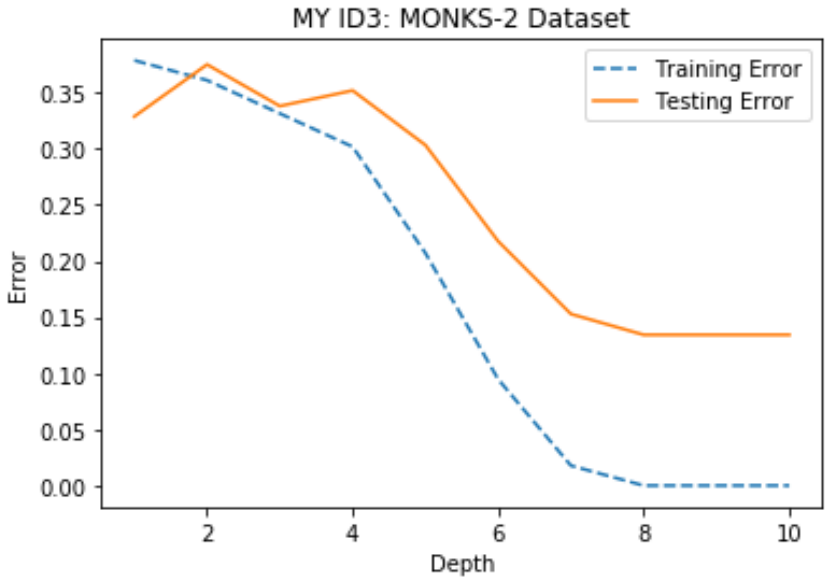
a. For depth = 1, ……., 10, learn decision trees and compute the average training and test errors on each of the three MONK's problems. Make three plots, one for each of the MONK's problem sets, plotting training and testing error curves together for each problem, with tree depth on the x-axis and error on the y-axis.




MY TREE: monks-1 AVERAGE ERRORS

**Train Error = 9.11%.**

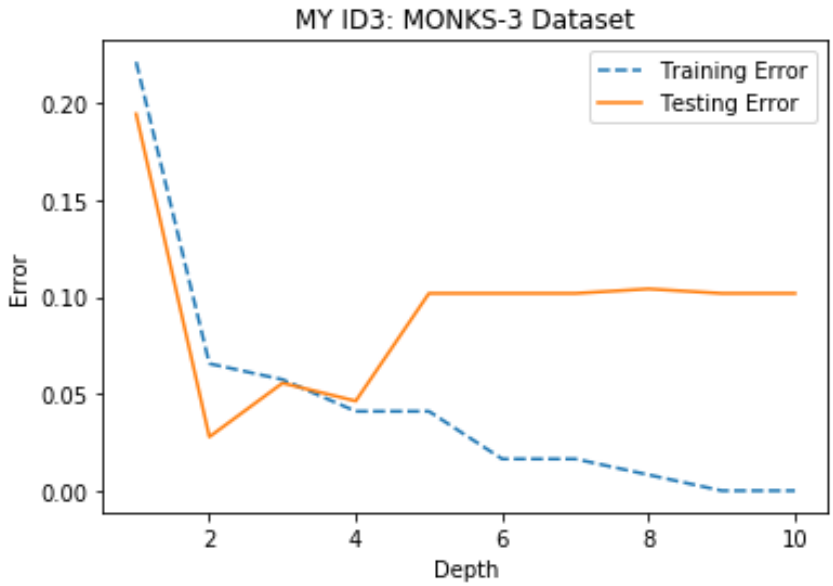
**Test Error = 14.54%.**




MY TREE: monks-2 AVERAGE ERRORS

Train Error = 16.92%.

Test Error = 24.70%.




MY TREE: monks-3 AVERAGE ERRORS

**Train Error = 4.67%.**

**Test Error = 9.38%.**

b. For monks-1, report the learned decision tree and the confusion matrix on the test set for depth=1 and depth=2.

FOR Depth – 1 of the Decision Tree on MONKS-1 Dataset

TREE
+-- [SPLIT: x4 = 1]
|  +-- [LABEL = 0]
+-- [SPLIT: x4 = 1]
|  +-- [LABEL = 1]

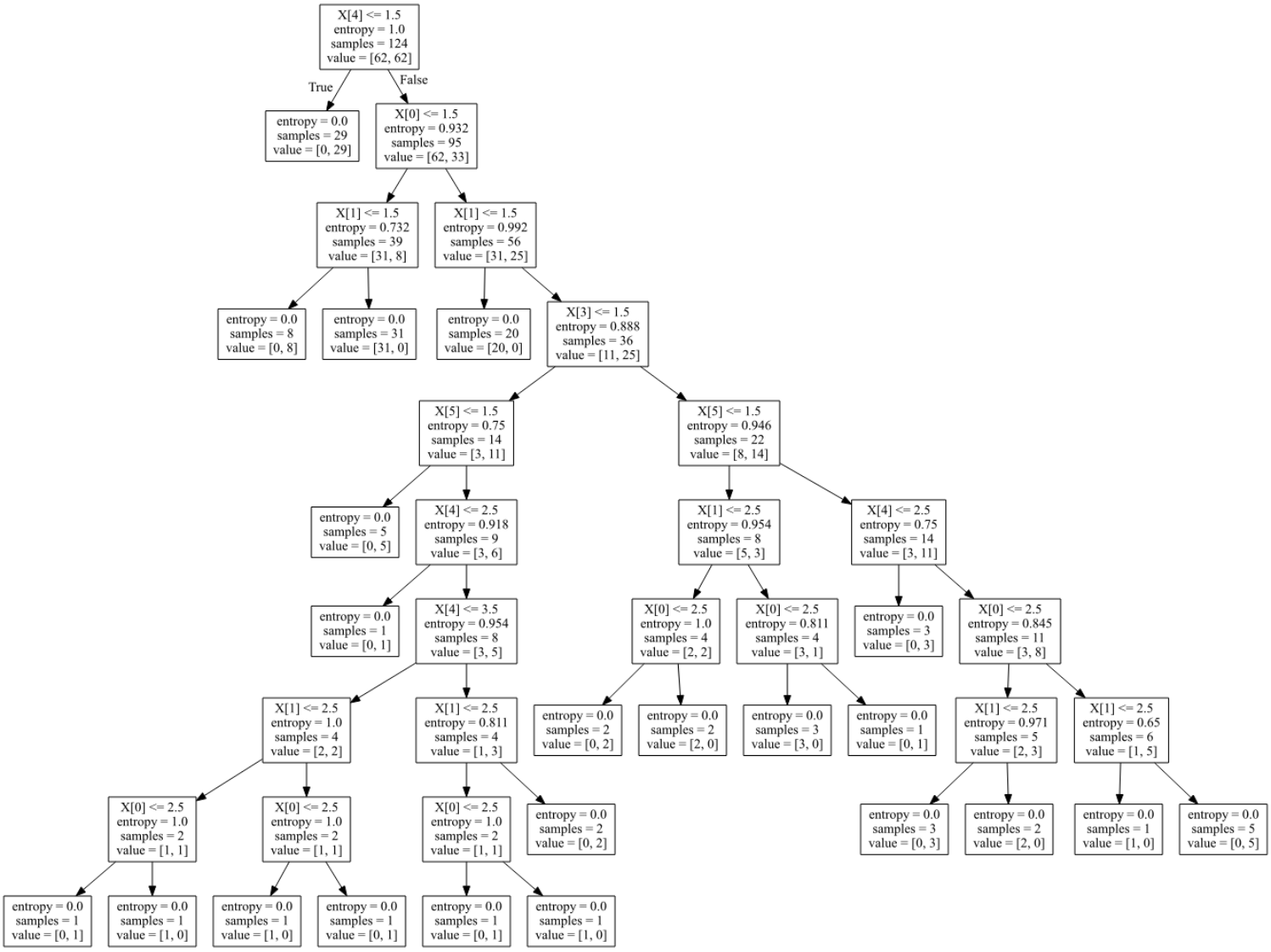
Confusion matrix: [[108, 108], [0, 216]]

FOR Depth – 2 of the Decision Tree on MONKS-1 Dataset

TREE
+-- [SPLIT: x4 = 1]
|  +-- [SPLIT: x0 = 1]
|  |  +-- [LABEL = 0]
|  +-- [SPLIT: x0 = 1]
|  |  +-- [LABEL = 0]
+-- [SPLIT: x4 = 1]
|  +-- [LABEL = 1]

Confusion matrix: [[108, 108], [0, 216]]

c. For monks-1, use scikit-learns default decision tree algorithm to learn a decision tree. Visualize the learned decision tree using graphviz. Report the visualized decision tree and the confusion matrix on the test set.




SKLEARN: monks-1 CONFUSION MATRIX

| | Predicted TRUE | Predicted FALSE |
|---|---|---|
| Actual TRUE | 205 | 11 |
| Actual FALSE | 29 | 187 |

d. Repeat steps 2 and 3 with your "own" data set and report the confusion matrices. You can use other data sets in the UCI repository.

MY TREE: **breast-cancer** CONFUSION MATRIX

| | Predicted TRUE | Predicted FALSE |
|---|---|---|
| Actual TRUE | 5 | 17 |
| Actual FALSE | 14 | 37 |