Machine Learning: Programming Assignment #1

Due on February 28, 2019 at 11:59pm

Professor Sriraam Natarajan Section 005

Viraj Mavani - VDM180000

Part A

Part A directs to complete the functions for Decision Tree learning using the ID3 algorithm and then visualizing the Train and Test errors for each tree with different maximum depths from 0 to 10. The ID3 algorithm was implemented using the NumPy library, The other part of plotting was programmed using the Matplotlib Library. Additionally, the same process was implemented for the Monks 2 and Monks 3 dataset as well.
The initial tree with depth 3 on the Monks 1 dataset can be seen as given in the figure below.

```
TREE
+-- [SPLIT: x4 = 1]
|       +-- [SPLIT: x0 = 1]
|       |       +-- [SPLIT: x1 = 1]
|       |       |       +-- [LABEL = 1]
|       |       +-- [SPLIT: x1 = 1]
|       |       |       +-- [LABEL = 0]
|       +-- [SPLIT: x0 = 1]
|       |       +-- [SPLIT: x1 = 1]
|       |       |       +-- [LABEL = 0]
|       |       +-- [SPLIT: x1 = 1]
|       |       |       +-- [LABEL = 1]
+-- [SPLIT: x4 = 1]
|       +-- [LABEL = 1]
Test Error = 16.67%.
```

Figure 1: Decision Tree of depth 3

The output plots for Part A can be found in the figure below.

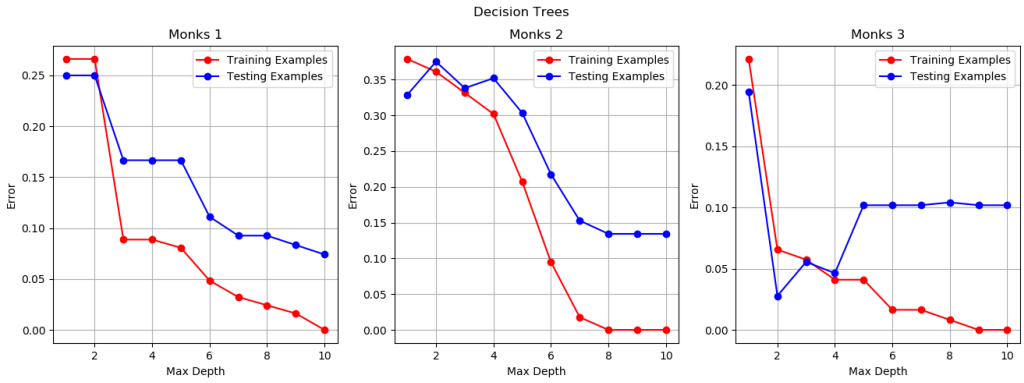

Figure 2: Depth-Error Curves

Part B

Part B of the assignment directs us to learn the decision trees of depth 1 and depth 2 and then compute the confusion matrices for the same on the Monks 1 dataset. The decision trees are given in Figure 3 below.

```
Depth 1 Decision Tree on Monks 1 Dataset
TREE
+-- [SPLIT: x4 = 1]
|       +-- [LABEL = 0]
+-- [SPLIT: x4 = 1]
|       +-- [LABEL = 1]
Depth 2 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]
```

Figure 3: Decision Trees of Depth 1 and 2

The confusion matrix for Depth 1 Decision Tree on the test set is given in Figure 4.

	Classifier Positive	Classifier Negative
Actual Positive	108.0	108.0
Actual Negative	0.0	216.0

Figure 4: Depth 1 Decision Tree Confusion Matrix

The confusion matrix for Depth 2 Decision Tree on the test set is given in Figure 5.

	Classifier Positive	Classifier Negative
Actual Positive	108.0	108.0
Actual Negative	0.0	216.0

Figure 5: Depth 2 Decision Tree Confusion Matrix

Part C

This section of the assignments directs us to learn a decision tree using the Scikit-Learn implementation and visualizing it using the GraphViz library. The sklearn decision tree by default uses the Gini algorithm and so I had to set it to compute on the basis of entropy or in other words, ID3. Figure 6 shows the decision tree and Figure 7 shows the Confusion Matrix on the test set.

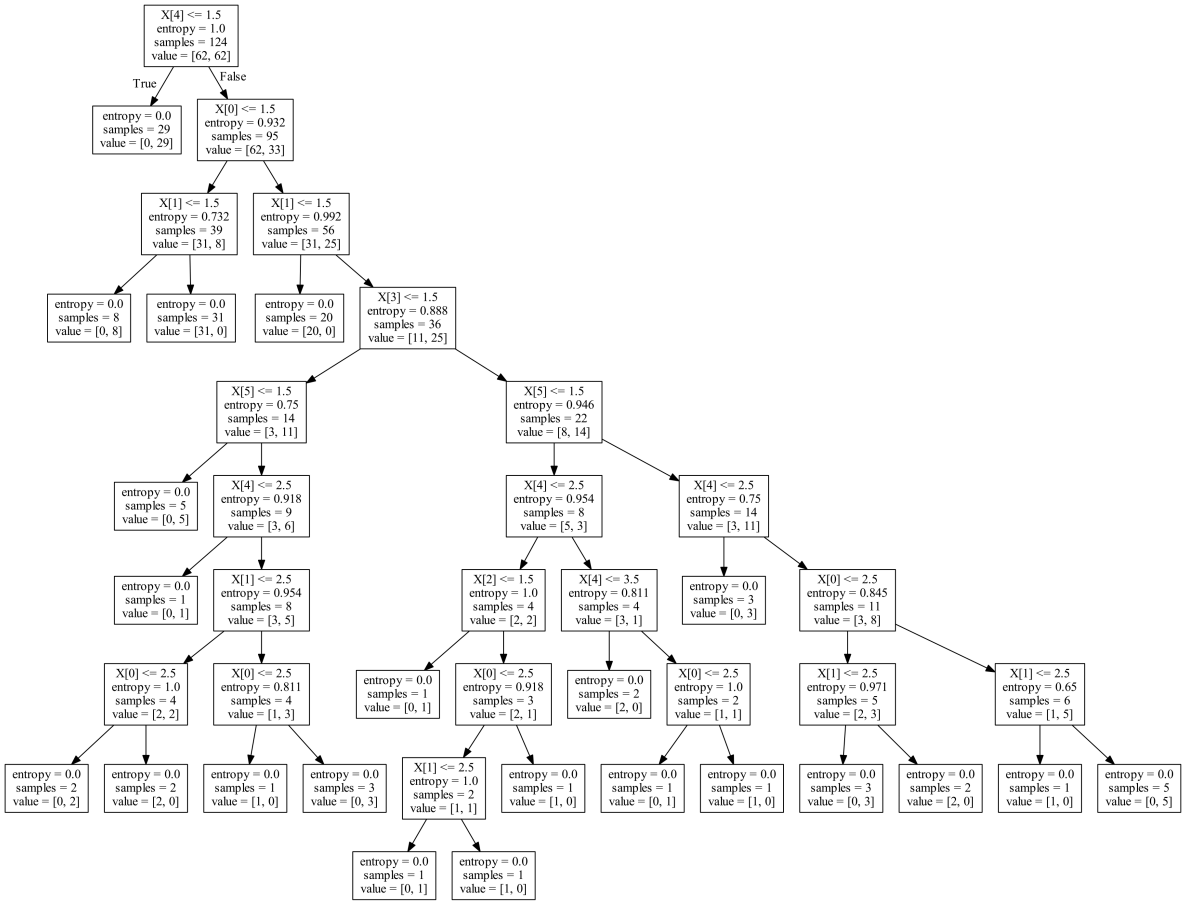

Figure 6: GraphViz visualization of Sklearn Decision Tree

	Classifier Positive	Classifier Negative
Actual Positive	205.0	11.0
Actual Negative	31.0	185.0

Figure 7: Sklearn Decision Tree Confusion Matrix

Part D

Part D directs us to implement Part B and Part C for a custom dataset. Here, for this project, I have used the Car Evaluation Dataset from the UCI Machine Learning Repository. I used a 80-20 split for training and testing with 1382 samples in training set and 345 samples in test set.
The required figures are given below.

```
Depth 1 Decision Tree on Car Evaluation Dataset
TREE
+-- [SPLIT: x3 = 0]
|       +-- [LABEL = 2]
+-- [SPLIT: x3 = 0]
|       +-- [LABEL = 2]
Depth 2 Decision Tree on Car Evaluation Dataset
TREE
+-- [SPLIT: x3 = 0]
|       +-- [SPLIT: x5 = 1]
|       |       +-- [LABEL = 0]
|       +-- [SPLIT: x5 = 1]
|       |       +-- [LABEL = 2]
+-- [SPLIT: x3 = 0]
|       +-- [LABEL = 2]
```

Figure 8: Depth 1 and Depth 2 Decision Trees on Car Evaluation Dataset

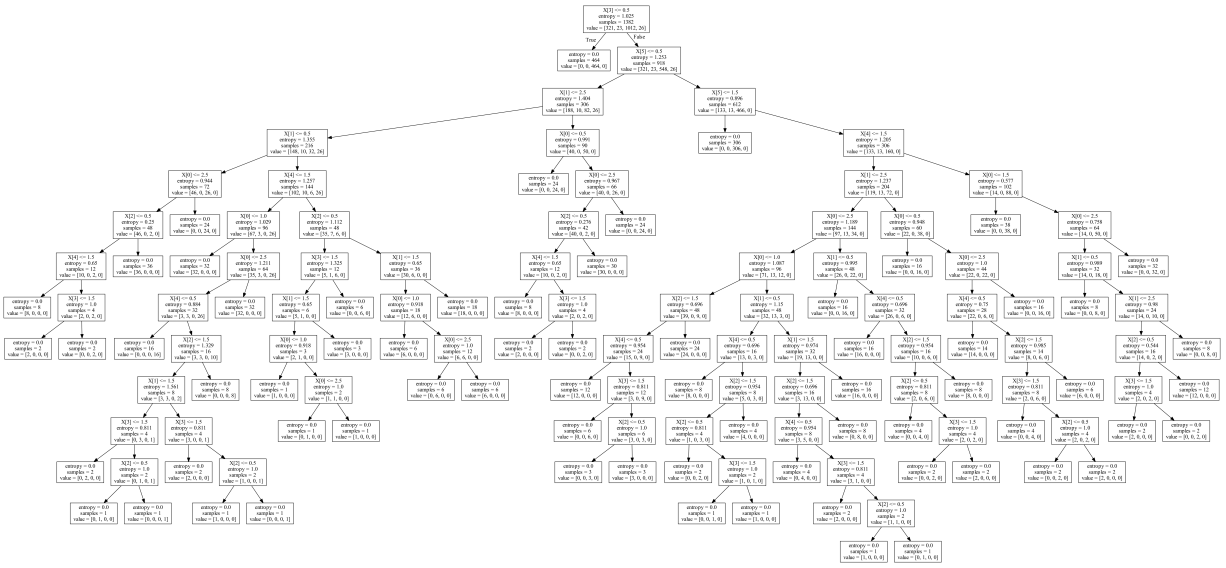

Figure 9: GraphViz visualization of Sklearn Decision Tree on Car Evaluation Dataset

	Predicted acc	Predicted good	Predicted unacc	Predicted vgood
Actual acc	0.0	0.0	63.0	0.0
Actual good	0.0	0.0	46.0	0.0
Actual unacc	0.0	0.0	198.0	0.0
Actual vgood	0.0	0.0	38.0	0.0

Figure 10: Depth 1 Decision Tree Confusion Matrix on Car Evaluation Dataset

	Predicted acc	Predicted good	Predicted unacc	Predicted vgood
Actual acc	63.0	0.0	0.0	0.0
Actual good	46.0	0.0	0.0	0.0
Actual unacc	8.0	0.0	190.0	0.0
Actual vgood	38.0	0.0	0.0	0.0

Figure 11: Depth 2 Decision Tree Confusion Matrix on Car Evaluation Dataset

	Predicted acc	Predicted good	Predicted unacc	Predicted vgood
Actual acc	33.0	0.0	30.0	0.0
Actual good	46.0	0.0	0.0	0.0
Actual unacc	0.0	0.0	198.0	0.0
Actual vgood	38.0	0.0	0.0	0.0

Figure 12: Sklearn Decision Tree Confusion Matrix on Car Evaluation Dataset