

Decision Tree Implementation

For Iris dataset your task is to print steps for every split in the decision tree.
Your Project is divided in 2 parts -

1. Print the Decision tree steps as specified in the example below. This is of 80 marks in total.
2. Decision Tree Implementation: Building actual decision tree for the dataset and then printing it. This is of 20 marks.

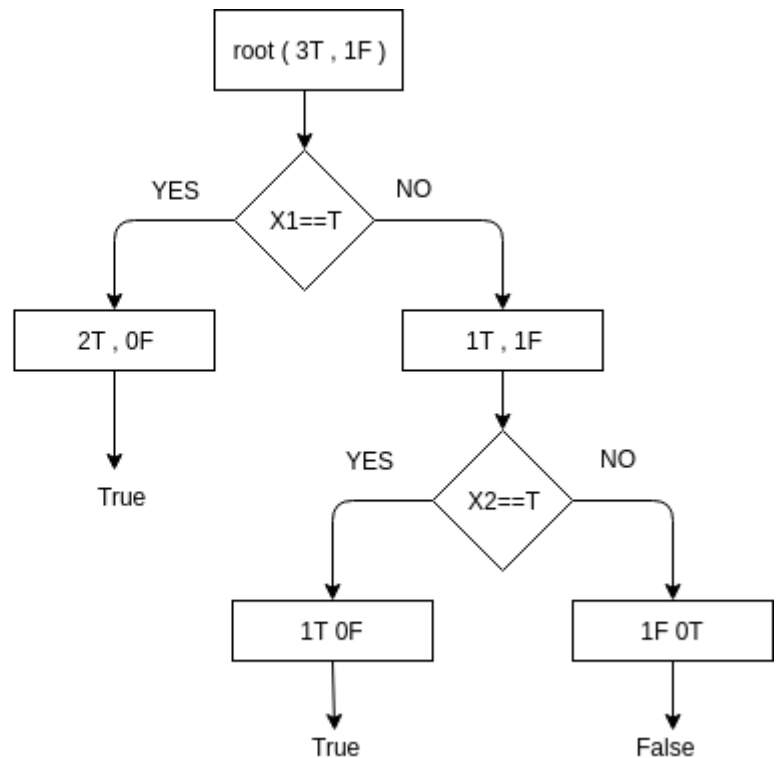
Note: You can submit multiple times before deadline/evaluation. Your final submission will be evaluated. No submission is allowed after the project is evaluated by TA.

Dataset: Iris dataset.

Implementation: Consider the decision tree for OR below:

X1	X2	Y (X1 OR X2)
True	True	True
False	True	True
True	False	True
False	False	False

Truth Table for OR



Expected Output:

Level 0

Count of 0(False) = 1

Count of 1(True) = 3

Current Entropy is = 0.811278124459

Splitting on feature X1 with gain ratio 0.311278124459

Level 1

Count of 0 = 1

Count of 1 = 1

Current Entropy is = 1.0

Splitting on feature X2 with gain ratio 1.0

Level 2

Count of 0 = 1

Current Entropy is = 0.0

Reached leaf Node

Level 2

Count of 1 = 1

Current Entropy is = 0.0

Reached leaf Node

Level 1

Count of 1 = 2

Current Entropy is = 0.0

Reached leaf Node

Comments : Your code must have proper comments for better understanding.

Extra features : You can add extra images etc for clarity and better presentation of your results.

Score : Score will be given by the TA based on your submission.

Submission : You have to upload zipped file which has python notebook with implementation.

Your project will be evaluated on following parameters -

- Correctness (Max Score 60)
- Quality of Code (Max Score 10)
- Proper Commenting (Max Score 10)
- Bonus: Actual Tree Implementation instead of just printing steps. (Max Score 20)