Assignment 1 - Decision Tree Classifier

MACHINE INTELLIGENCE - PESU UE18CS303

[Graded for 5 Points]

Decision Trees are one of the easiest and popular classification algorithms to understand and interpret. The goal of using a Decision Tree is to create a training model that can be used to predict the class or value of the target variable by learning simple decision rules inferred from prior data.
The primary challenge in the decision tree implementation is to identify which attributes to consider as the root node at each level. Handling this is known as attribute selection.
The ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking. It always makes the choice that seems to be the best at that moment.
Attribute selection in the ID3 algorithm involves various steps such as computing entropy, information gain and selecting the most appropriate attribute as the root node.

In this assignment you are given functions that calculate all these important parameters that help in construction of a **categorical variable decision tree.**

Your task is to complete the code for these functions.

You are provided with the following:

- A python file - *Assignment1.py*
    - The python file has the function definitions, the parameters the functions take and the expected output.

- Sample test cases to validate your code against.
    - *DT_SampleTestCase.py* has all the sample test cases.
      Simply run `python3 DT_SampleTestCase.py` to validate your functions.

- You are **not allowed** to import any other library other than the ones already imported for you.

Important Points:

- Please do not make changes to the function definitions that are provided to you. Use the skeleton as it has been given. Also do not make changes to the sample test file provided to you. Run it as it is.
- You are free to write any helper functions that can be called in any of these predefined functions given to you.
- The dataset we will be testing against will have N columns with N-1 attributes and Nth column being the target value.
- Your code will be auto evaluated by our testing script and our dataset and test cases will not be revealed. Please ensure you take care of all edge cases!
- Ensure you follow convention while returning from these functions.

DEADLINE : 15th October 2020

(Submission details and procedure will be communicated well before the deadline)