

Machine Learning: Programming Assignment 1: Decision Trees

Problem Statement:

Write a program to learn a decision tree. Decision tree learning should use *information gain* as the criterion for choosing the attribute for splitting. *Tree pruning should not be performed*. The tree should be tested on few test samples. The tree structure should be printed as output.

The program can be written c/C++/java/Python programming language from scratch. No machine learning/data science/statistics package/library should be used.

Data Set Description:

Training Data Filename: *data1_19.csv*

Training Data Description: The data set is about the passengers who survived the Titanic disaster. There are three input features: Passenger class (pclass), age, and gender. Each of the features are symbolic. Only the distinct values present in the training data comprise the domain of values of the feature. There is one output to be predicted: survived (yes/no).

Output Format:

The constructed tree structure should be printed as output. For output, you can choose how to draw the tree so long as it is clear what the tree is. Do not use any graphics package. You should use formatted text. You might find it easier to use indentation to show levels of the tree as it grows from the left. For example:

```
pclass = 1st
| gender = male: no
| gender = female : yes
pclass = 2nd: yes
pclass = 3rd
| age = adult: no
| age = child: yes
```

Submission Guidelines:

You may use one of the following languages: c/C++/Java/Python. You should name your file as `<rollnumber_a1.extension>` (e.g., `16CS10001_a1.c`). Your program should be standalone and should not use any *special purpose* library. `numpy` may be used. You should submit the program file only and not the output/input file. The submitted single program file *should* have the following header comments:

```
# Roll          # Name          # Assignment number          # Specific compilation/execution flags (if required)
```

Please submit the program in moodle by **August 21, 2018 midnight** (hard deadline). Copying from friends/web will lead to strict penalties.