

BIL712 Machine Learning

Fall 2016

Programming Assignment 1 – Decision Tree Learning

Due Date: November 16, 2016

In this programming assignment, you are going to implement decision tree algorithm ID3. Your program is going to use three input files:

functionAttributes.txt –

This file should contain the information about the attributes of the target function (including target attribute). This file contains exactly N lines for a function with N attributes (including target attribute). The first line should contain information about target attribute, the second line contains the information about the first attribute, the third line contains the information about the second attribute, and so on.

The first line contains the possible number of values for the target attribute (a value > 1), and possible values for the target attribute. For example, if the target function is a *boolean* function, the first line will be:

```
2 yes no
```

If the target attribute can take four possible values (a, b, c, d), the first line will be:

```
4 a b c d
```

The line for an attribute will be in the same format of the line for the target attribute.

trainingSet.txt –

This file contains the training examples. Each line contains a training example which consists of N attribute values. The first one is the value of the target attribute, the second value is the value of the first attribute, and so on. Each value is separated with one or more space character. The following is a sample training example file for a function with three attributes. The target attribute may take values “yes” and “no”; the first attribute may take values “a” and “b”; the second attribute may take values “c”, “d” and “e”.

```
yes a c
yes b d
no a e
no b c
```

evaluationSet.txt –

This file contains the evaluation set and its format is same as the file `trainingSet.txt`.

Your program should do the followings:

- It should read `functionAttributes.txt` and `trainingSet.txt` files in order to learn the decision tree using ID3 learning algorithm. It should convert the learned decision tree into rules and print them into `rules.txt` file.
- After the decision tree is learned, it should read `evaluationSet.txt` file in order to find the accuracy result for the instances in this file. It should print the accuracy result on the screen.
- Test your program with at least two example sets.
- Use Java, C or C++ to implement your homework.

You should hand-in:

- Put the following files into a `.rar` file (named as `yourname.rar`) and you will send it use an online submission system.
 - Your source program files
 - Sample input files for your program
 - Executable of your program and a message indicating how to run your executable
- You will use the online submission system (<https://submit.cs.hacettepe.edu.tr/>) to submit your projects. You should use your cs email accounts (with domain `cs.hacettepe.edu.tr`). You should write your student id with lowercase (i.e. `n12000000` or `o120000000`) to login into the system. If you have any problem with the online submission system, you may contact with the assistant Ali Osman Serhatoglu (`aoserhatoglu@cs.hacettepe.edu.tr`).