Introduction to Intelligent System

# Assignment 3

Instructor: Le Nguyen

**Date**: Monday, March 18, 2019

**Due Date**: April 8, 2019

**Total marks: 35.**

**This is an individual assignment.**

1. (3 marks) You are to explain three types of machine learning. Give an example for each type of learning.
    a. Supervised learning
    b. Unsupervised learning
    c. Reinforcement learning.
2. (12 marks) Decision Tree
   **A,B,C,D are attributes.**

| Example | A | B | C | D | Y |
|---------|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 0 |
| 4 | 1 | 0 | 0 | 1 | 1 |
| 5 | 1 | 1 | 0 | 0 | 1 |
| 6 | 0 | 1 | 1 | 0 | 1 |
| 7 | 0 | 0 | 1 | 1 | 1 |
| 8 | 0 | 0 | 0 | 1 | 1 |
| Test | | | | | |
| 9 | 1 | 1 | 1 | 1 | |
| 10 | 1 | 1 | 0 | 0 | |
| 11 | 0 | 1 | 0 | 1 | |

   1. Calculate information gain and draw the decision tree.

a. Calculate information gain of attributes for S (Original data set).
b. Choose the root of the decision tree.
c. Calculate the information gain of attributes for subsets of S.
d. Use the test case 9 to 11 to test your decision tree and write out the result of the test.

3. **(20 marks)** Coding: Decision Tree Learning.
There are **2 input files**: Scheme file and Data file. Both provides the training examples for the decision tree. A typical Scheme file and data file have the following structures.

| Scheme File | Data File | | | | |
|---|---|---|---|---|---|
| 5 | Outlook | Temperature | Humidity | Wind | Play_Golf |
| | S | H | H | W | N |
| Outlook | S | H | H | S | N |
| 3 | O | H | H | W | Y |
| S O R | R | M | H | W | Y |
| | R | C | N | W | Y |
| Temperature | R | C | N | S | N |
| 3 | O | C | N | S | Y |
| H M C | S | M | H | W | N |
| | S | C | N | W | Y |
| Humidity | R | M | N | W | Y |
| 2 | S | M | N | S | Y |
| H N | O | M | H | S | Y |
| | O | H | N | W | Y |
| Wind | R | M | H | S | N |
| 2 | | | | | |
| W S | | | | | |
| | | | | | |
| Play_Golf | | | | | |
| 2 | | | | | |
| N Y | | | | | |

The last column "Play_Golf" is the function output.

There are number of java classes

- **Attribute.java**: It provides a definition of an attribute. An attribute has its name and several possible values that stores in an array list.
- **Example.java**: It stores the attribute values of **a data example**. It uses **integer** to store attribute values.
- **Scheme.java**: It stores the attributes information into a list of attributes, and the function, and number of attributes. It should read the scheme file and store the scheme information.
- **DataSet.java**: It has training data set (a list of examples) and a scheme related to a data set. It has a method that read the data file, store data in the list and **getEntropy, getRemainder, getBestAttribute methods**.
- **Node.java**: It is a node in a decision tree. It has a node label, a reference to parent node, and a label link incoming from a parent node.
- **DTLearner.java**: It has a main method. It implements the decision tree learning algorithm, sameClassfication, majorityValue.
- **Util.java**: You might put some static methods here.

Running Decision Tree Learner:

Java DTLearner SchemeFile DataFile

Decision tree learning begins:

Outlook: Information gain = 0.246

Temperature: Information gain = 0.029

Humidity: Information gain = 0.151

Wind: Information gain = 0.048

Choose attribute: **Outlook**

Temperature: Information gain = 0.57

Wind: Information gain = 0.01

Humidity: Information gain: 0.97

Choose attribute: **Humidity**

Wind: Information gain = 0.97

Temperature: Information gain = 0.01

Choose attribute: **Wind**

The decision tree should be drawn:

Outlook = S

| humidity = H: **N**

| humidity = N: **Y**

Outlook = O: **Y**

Outlook = R

| wind = S: **N**

| wind = W: **Y**

**<u>Test cases:</u>**

There are several test cases provided for you. You should try to test with simple cases.

If scheme file and data file are different or not match, you should terminate and print out error.

Note: you should pay attention on getting sub data sets and sub attribute list.

I provided a detailed calculation of the above example in an excel sheet.

**Submission:**

Codes that **cannot compile** will receive a zero mark. Please make sure it can compile and it should work on Windows 10 and Linux. Oracle **Java 8 SDK** is used to mark the assignment.

- Please **comment** your code properly.
- Please don't have any **package statement in the Java classes**.

Please submit the following files:

Put all your files in **yourName_A3.zip** and submit**.**
1. A3.pdf.
2. README.txt File
3. The all Java files above