CS421: Advanced Artificial Intelligence
Assignment 4: Applying Ensemble Learning Methods and Others to Classification Problems

1 Overview

The main task of this assignment is to use existing software tools to compare ensemble learning methods or others on a selected data set. This assignment takes 10% in your final score, thus 10 marks.

You will work **individually** on the training set of the same Kaggle in-class competition of in Assignment 3, then write a report **individually**. The score you will get in this assignment depends on the ranking of your submission and your report.

2 Main tasks

2.1 Program (4 marks)

We will use the training set of the same Kaggle in-class competition of Assignment 3. You will work on **one ensemble learning method or some other methods** for classification. Thus, the limitation on method that you are using has been relaxed. You only need to make your classifier as accurate as possible and do not plagiarise.

2.1.1 Naming policy and submission

Each student will prepare one program using a classifier method, and name the file as **classifier-{studentNumber}.py** . Example: **classifier-12345678.py**. Each student will submit one classifier named as "classifier{studentnumber}" to the competition, e.g., **classifier12345678**.

Competition link: `https://www.kaggle.com/t/7e34beba9c6d46df81abd152cfa01a0d`

2.1.2 Ranking and scoring

The ranking is made on the performance of your classifier on the unknown test set.

Program (3 points) Your program should be clear, well documented. One of your colleague will read your program and try to repeat the experiments. If it is not possible to repeat or the program can not be executed properly, then you will receive 0.

Ranking (1 point)   The ranking of your classifier on the test set (ranked by test error) will be used as part of evaluation of this assignment. The point that you will receive is

$$\frac{accuracy - worstAccuracy}{bestAccuracy - worstAccuracy},$$

where $accuracy$ refers to the accuracy of your classifier, $bestAccuracy$ and $worstAccuracy$ refers to the best and worst accuracy of the submissions to Assignment 4 only.

The above evaluation policy is based on the fact that your classifiers will be executed without any problem.

2.2 Report (6 marks)

A report (in pdf format) must be submitted, named as "report{studentnumber}.pdf". Example: report12345678.pdf.

- (2 points) Detailed description of the method and the library/package that you are using, if any.

- (3 points) Detail how the model was trained (parameters, tuning, time, computational resource, etc.). All parameters and corresponding values used in the experiments should be reported.

- (1 points) Illustration of the results and conclusion.

2.3 What to submit

You should submit your program (including trained model) and report via *Sakai* **no later than 11pm, 13th Dec 2018**. Each student is asked to submit one single zip file respecting the following requirements.

- Create a folder named as **assignment4-{studentnumber}**.

- The zip file should contain

  - A **readme.txt** which explains how to launch your code.
  - Your program, which should create and save a file named as **{studentnumber}-submission.csv**, given an input of same format as in Assignment 3, in the current folder and respect the format of the competition submission. Thus the your .csv file should be a valid solution file to the competition.
  - A report named as **report{studentnumber}.pdf**.

- Zip the folder and name your zip: The zip should be named as **assignment4-{studentnumber}.zip**. Example: assignment4-12345678.zip

3 Remarks

Prohibition You should not copy other's program, except using existing public libraries or packages. Otherwise, you will receive 0 as mark in this assignment.

You will get 0 as mark in any of the following cases, but not limited to:

- You don't respect the naming policy of classifier.

- You don't respect the formats of the submission of Assignment 4.

- You have used (part of) the original dataset for training.

- You have delayed the submission for more than 3 days.

Contact For any question regarding this assignment, please email to *liujl@sustc.edu.cn*. The subject of the email should respect the format: **[AAI] Assignment 4 (LastName/FirstName-StudentNumber)**
Example: [AAI] Assignment 4 (Liu/Jialin-12345678)