CS421: Advanced Artificial Intelligence

Assignment 3: Comparing Neural Networks, Decision Trees and Support Vector Machine

1 Overview

The main task of this assignment is to use existing software tools to compare Neural Networks (NNs), Decision Trees (DTs) and Support Vector Machine (SVM) on a selected data set. This assignment takes 10% in your final score, thus 10 marks.

You will form a group of 3 students (except one group of 4 students) to participant a given Kaggle in-class competition, then give a presentation by group in a lab. The score you will get in this assignment depends on your group ranking and your group presentation including the individual contribution due to which the members in a team will possibly receive different marks. No report is required in this assignment.

2 Main tasks

2.1 Competition (5 marks)

A Kaggle in-class competition is set up by the TA. You will work on a data set for classification. A training set is given, a test set will be used for evaluating the trained classifiers.

Competition link: `https://www.kaggle.com/t/3dcad4be79d4400f8f0a1355eb7f00fc`

2.1.1 Data set

The given (public) data set has 6 classes. Each sample has 6 input features, indexed by 0 to 5.

#Public samples	$10,000$
#Input features	6
#Classes	6

2.1.2 Naming policy and submission

Each group should have a name. From now on, the group name "demogroup" will be used as an example. Each group will prepare 3 kernels the competition using a NN, DT and SVM, and name the corresponding classifiers using the group name as prefix followed by "_nn", "_dt" and "_svm", respectively. Example: "demogroup_nn", "demogroup_dt" and "demogroup_svm". Thus, each group will submit 3 classifiers to the competition.

Each classifier can be updated up to 5 times per day. After the competition submission deadline, the submitted classifiers will be tested on a private test set. The ranking is made on the performance of your classifiers on the test set.

2.1.3 Ranking and scoring

The ranking of your classifiers on the test set (ranked by test error) will be used as part of evaluation of this assignment. This ranking will be restructured to 3 distinct rankings by the category of algorithms, thus NN,DT, and SVM. Thus, each group will have 3 rankings of their 3 submitted classifiers. For any classifier $j \in \{nn, dt, svm\}$, let $c_{i,j}$ denote the percentage of correct prediction, and ordered as $c_{1,j} \geq c_{2,j} \geq \ldots, c_{7,j}$, the mark of the group ranked at i^{th} position for classifier j is $mark_{i,j} = 1 + \frac{2}{3} \times \frac{c_{i,j} - c_{7,j}}{c_{1,j} - c_{7,j}}$ if classifier j is submitted, otherwise $mark_{i,j}$. The sum of marks for the 3 classifier is the part of mark based on ranking of this group.

The above evaluation policy is based on the fact that your classifiers will be ran without problem.
You should submit your trained model and code via *Sakai* **no later than 11pm,** 22^{nd} **Nov 2018**.
Each student is asked to submit one single zip file respecting the following requirements.

- Create a folder named as **assignment3-source-{studentnumber}**.
- The zip file should contain
  - A **readme.txt** which explains how to launch your code.
  - Your code, which should create and save a file named as **{studentnumber}-submission.csv** 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 file named as **sample-submission.csv** which is a prediction of your submitted model (a valid solution file).
- Zip the folder and name your zip: The zip should be named as **assignment3-source-{studentnumber}.zip**.
  Example: assignment3-source-12345678.zip

2.2 Group presentation (5 marks)

You will give a group presentation (**no more than** 10 minutes' talk + questions) in the lab on 30^{nd} **Nov 2018**.
You are expected to present your classifiers, results and credits of work, thus individual contribution. Please be serious with the presentation. **The time limit (10 minutes) is strict.**

You should send your presentation slides via *Sakai* **no later than 11pm,** 29^{nd} **Nov 2018**. No update of the slides will be possible after the 11pm, 29^{nd} Nov 2018, no matter how minor the edit will be.

The order of the presentations is random and will be decided on 30^{nd} Nov 2018.

Each student is asked to submit one single zip file respecting the following requirements.

- Create a folder named as **assignment3-presentation-{studentnumber}**.
- The zip file should contain a **.pdf** file of your presentation slides.
- Zip the folder and name your zip: The zip should be named as **assignment3-presentation-{studentnumber}.zip**.
  Example: assignment3-presentation-12345678.zip

2.3 Important dates

- Trained model and relevant code submission deadline: 10pm59, 22^{nd} Nov 2018
- Presentation slides submission deadline: 10pm59, 29^{th} Nov 2018

No delay is allowed.

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.
- You have delayed the submission (strict deadline).

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 3 (LastName/FirstName-StudentNumber)**
Example: [AAI] Assignment 3 (Liu/Jialin-12345678)