CSE472 (Machine Learning Sessional)
L-4, T-2, January 2017 Term

Assignment 2: Ensemble Learning

In ensemble learning, we combine decisions from multiple weak learners to solve a classification problem. It is expected that the combined decision will perform better than the individual models in the process of one model correcting the errors of the other.

There are many ensemble methods such as stacking, bagging and boosting. In this assignment you will be implementing the **discrete AdaBoosting** algorithm. The necessary details are as follows.

1. As the weak/base learner use ID3 decision stump. A decision stump is a decision tree of depth one (i.e., it branches on only one attribute and then makes decision).
2. There are several implementation of AdaBoosting algorithm. Follow the pseudocode given in the class.
3. You should make your code as modular as possible. Namely, your main module of Ada-Boosting should treat the base learner as a blackbox and communicate with it via a generic interface that inputs weighted examples and outputs a classifier, which then can classify any instances. For the decision stump, you can modify your ID3 implementation in assignment 1 or implement it from scratch.
4. To incorporate effect of weights in ID3, use sampling with replacement strategy. Use information-gain as the evaluation criterion. Do not use pruning.
5. To train and test your model, use the breast cancer dataset given for assignment 1.
6. Analyze the expected performance of your model using k-fold cross validation for k=5, 10 and 20, leave-one-out cross validation.

Instructions for Report Writing

1. Your final report will contain the following points:
   - Tabulate the expected accuracies obtained using k-fold cross validation for k=5, 10 and 20, leave-one-out cross validation.
   - Compare and analyze the accuracies obtained by different learners: decision stump alone,boosting with 30 rounds, your ID3 implementation.

- Compare and analyze the accuracies obtained by boosting with different numbers of rounds:5, 10, 20, 30.

2. Never copy the report. Just answer the questions precisely. Make it as simple as possible. Too much description is not needed!

Special Instructions

- Don't Copy anything! If you do copy from internet or from any other person or from any other source, you will be severely punished and it is obvious. More than that, we expect Fairness and honesty from you. Don't disappoint us!
- The report should be in .docx/.pdf (No hardcopy is required). Write precisely in your own language and keep it as simple as possible.
- For python and matlab, you may not get supporting softwares in the lab. If you do program in these languages, bring your computer in the sessional.
- You are allowed to show the assignment in your own laptop during the final submission. But in that case, ensure an internet connection as you have to instantly download your code from the moodle and show it.

Submission Deadline

- Upload the codes in moodle within 11 A.M. of 27th March, 2017 (Monday).

Instructions for moodle upload

Upload the assignment within the specified time. Otherwise, we can't accept it. For submission, use the following rules:

- If you write code in a single file, then rename it as<Studentid><code>.<extension>. For example, if your student id is 1205123 and you have done in java, then your file name should be "1205123code.java".
- If you write code in multiple files, then put all the necessary files in a folder and rename it as<Studentid><code>. For example, if your student id is 1205123 and you have done in java, then your folder name should be "1105123code".
- The report name should be <Studentid><report>.<extension>. For example, if your student id is 1205123 and it is in pdf format, then the report name should be "1205123report.pdf".

- Finally make a main folder, put the code (whether file or folder) and report in it, and rename the main folder as your<Student id><Programming language>. For example,”1205123Java”.Then zip it and upload it.