

CS534 — Implementation Assignment 3 — Due May 30th, 2012

General instruction.

1. You are strongly recommended to use Matlab for your implementation. However, Java, C/C++, R are also accepted.
2. You can work solo or in team of 2 people. Each team will only need to submit one copy of their solution.
3. Your source code will be submitted through the TEACH site (Due at noon on the due date).

`https://secure.engr.oregonstate.edu:8000/teach.php?type=want_auth`

Please include a readme file, which should contain the team member information.

4. You will also need to bring a hard copy of your report to the class on the due date. Similarly, clearly indicate the team members on your report.
5. Your report should be clearly written, presented in an organized manner. Figures should be correctly labeled, with axes and legends. Be careful when printing in black and white, that your figure should still be readable. Please, provide appropriate comments and notes about your work.

Ensemble of Decision Stumps

In this assignment you will implement Bagging and AdaBoost using the decision stump algorithm as the base learner. You will test your implementation on the SPECT data sets, which is a two-class classification problem, with 22 binary features. You will train your classifiers using the SPECT-train.csv file, and test on the SPECT-test.csv file.

You need to submit:

- a. Source code of your implementations (Decision stump, bagging, Adaboost).
- b. A report that contains the following components.
 - 1). For each ensemble method, please plot the training and testing errors as the size of the ensemble varies (please consider sizes 5, 10, 15, 20, 25 and 30). Note that for bagging, due to its stochastic nature, different random runs may lead to different results. To increase the robustness of the results, please show the average error rates over five random runs.
 - 2) A brief discussion of the results. What trend do you observe from the results? Can you provide any explanation for the observed results?