COSC470 Assignment 1

Due: August 10, 2018
Worth: 25% of the mark for COSC470

July 16, 2018

Tasks:

1. Implement:
   (a) Decision trees with a simple stopping criterion (no need to do pruning)
   (b) the Random Forests algorithm with your implementation of decision trees
   (c) the AdaBoost algorithm with the same decision trees
2. Evaluate both algorithms with the 4 classification datasets in sklearn. You should use cross-validation to get a measure of the accuracy of the algorithm on the data set.
3. Write up your experiments as a scientific report. If you are unsure how to structure such a report, a good model is to split the report into the sections: Introduction; Methods (describe AdaBoost, Random Forests, Decision trees); Evaluation (how you did the evaluation - e.g. cross-validation); Results (report the results); Discussion/conclusion (is there a clear winner?).
4. In addition, you are to choose one of the papers from ICML 2018 and do a brief review (one page is enough). The list of orals can be obtained here: `https://icml.cc/Conferences/2018/Schedule?type=Oral`. There are only a few papers on boosting or random forests, so they might be good ones to choose. There could very well be maths in these papers you don't understand. That's OK. In the review focus on:
   (a) The problem they are trying to solve (what limitation of previous work has been identified)?
   (b) How they solved the problem.
   (c) Have they presented convincing evidence? Are there limitations to their experimental method (if they had one).

All the code and the writing should be your own work.

To hand in:

1. Your code. I should be able to run the code with a minimum of fuss.
2. Your report as a pdf.

Your code should follow the standard sklearn interface for classifiers. That is, you should create a class with at least two methods: fit and predict. fit is the method used to learn the classifier, and predict is the method used to make a prediction for a new data point. If you use the standard interface, then you can use the standard cross-validation functionality described here: `http://scikit-learn.org/stable/modules/cross_validation.html#cross-validation`