

# Homework 2, due September 12th, 11:59pm

September 6, 2017

This project should be submitted both to Blackboard and to the peer evaluation website <https://pevals.com/>. Each team will need to create an account in Pevals to submit the homework there. For anonymity reasons, only write your FSU username on this project instead of your full name. Each team will have to grade 5 homeworks of their peers and enter their grades in Pevals. The homework will be graded by our grader Lizhe only for the teams that have completed their grading of the five projects in Pevals.

1. Use a programming language or package where random forests can be trained and applied. Examples include Python (scikit-learn package), R and Matlab. Using the training and test sets specified in the syllabus, perform the following tasks:

- a) On the `madelon` dataset, for each of  $k \in \{3, 10, 30, 100, 300\}$  train a random forest with  $k$  trees and report the training and test misclassification errors in a table.. Use the trained trees to predict the class labels on the training and test sets, and obtain the training and test misclassification errors. Plot on the same graph the training and test errors vs tree depth as two separate curves. Report the training and test misclassification errors in a table. (6 points)
- b) Repeat point a) on any of the other binary or multi-class datasets, with maximum tree depths  $d$  from 1 to 6. (1-3 points each as specified in the syllabus).

The datasets with 4-fold cross-val mean that 4-fold cross-validation should be performed on that dataset.