|  |
| --- |
| **Homework 3****COSC 6342: Machine Learning****University of Houston****Department of Computer Science****Sent on: October 21, 2019****Due: November 6, 2019 (midnight)** |

|  |
| --- |
| Name(s): |

Ensemble Learning

**Submission: Pdf with answers, description of the solution, and a zipped file with code.**

a) Get ionosphere: <https://archive.ics.uci.edu/ml/datasets/Ionosphere>

c) Using scikit-learn, run classification on Ionosphere first with Bagging and then with AdaBoost. Select 2 algorithms you are going to compare, ideally, you should pick weak learners like Decision Tree. (So, you will be using 2 boosted classifiers and then the same 2 but bagged). Perform your classification using K-fold cross-validation, with K being incremented from 2 to 4 to 6 to 8 to 10. Plot the test and train error rates as K changes. Make sure you plot them. (50 points)

d) Is anything interesting happening as K changes? If so, explain in terms of bias/variance, overfitting/underfitting. (10 points).

e) How do you think Bagging and Boosting compare in terms of overfitting/underfitting? Which component of the error (bias/variance) does each affect and how? (10 points)

f) When would you choose Boosting over Bagging? What about the other way around? (10 points)

g) Read about the Random Forest ensemble learning algorithm and, in a few brief sentences, explain your understanding of it. (10 points)