

CSE 6363 - *Machine Learning*

Project 2- Summer 2020

Due Date: Aug 2, 2020

Ensemble Classifiers

1. Consider the problem where we want to predict whether a banknote is authentic or not from a set of real numbered attributes characterizing properties of small image patches taken of the banknote, namely: i) the variance of the Wavelet transformed image, ii) the skewness of the Wavelet transformed image, iii) the kurtosis of the Wavelet transformed image, and iv) the entropy of image.

There is a training and a test data set for this problem on Blackboard (datasets are derived from the more expansive UCI machine learning banknote data set and are structured as comma separated lists with one data point per line where the first 4 numbers are the features and the final entry - 0 or 1 - indicates the class the data point belongs to).

- a) Implement a bagging routine for a logistic regression classifier. You have to implement bagging yourself but can use an existing implementation of logistic regression.
 - b) Apply bagging 10, 50, and 100 times to the training data. For each of the three cases, evaluate the resulting ensemble classifier on the test data set and compare the error rates for a single classifier and the three ensemble classifiers. Briefly discuss the results you obtained.
2. Using the data and logistic regression classifier, apply boosting to the problem.
 - a) Implement AdaBoost on top of your logistic regression classifier.
 - b) Apply boosting 10, 25, and 50 times to the training data. For each of the three cases, evaluate the resulting ensemble classifier on the test data set and compare the error rates for a single classifier and the three ensemble classifiers. Briefly discuss the results you obtained.