

CS771: Machine learning: tools, techniques, applications
Assignment #3: Decision trees and forests, ensembles

Due on: 03-04-2016, 23.59

16-3-2016

MM: 150

1. Use the MNIST data set for this problem. It contains 60000 images for training and 10000 images for testing. You can use 10000 images for training and 1000 images for testing to shorten run times. Ensure that your training and test data sets are balanced with respect to class labels.

You should use the histogram of gradients (HoG) as your features.

- (a) Build a single decision tree classifier.
- (b) Build a random forest classifier and show variation in performance with the number of trees in the forest.
- (c) Create a set of weak decision tree classifiers using different cell sizes in the HoG feature generator and/or by restricting tree depth. Now use the Adaboost algorithm to generate an ensemble classifier. Compare the 'best' boosted ensemble classifier with the corresponding 'best' random forest classifier. Best performance for random forest is the once error rate does not change much with the number of trees in the forest. Similarly, to get the best Adaboost classifier find the right number of trees in the Adaboost ensemble by using a validation set chosen from the MNIST training data set that is distinct from the training set used in your Adaboost algorithm.

[30,50,70]