

Assignment on Ensemble Learning

A decision stump is a decision tree of depth *one* (i.e., it branches on only *one* attribute and then makes decision). Implement the discrete AdaBoosting algorithm using ID3 decision stump as the *base learner*. You should make your code as modular as possible. Namely, your main module of AdaBoosting should treat the base learner as a blackbox and communicate with it via a generic interface that inputs weighted examples and outputs a classifier, which then can classify any instances. For the decision stump, you can modify your ID3 implementation in Homework 1 or implement it from scratch. To accommodate weights, create a wrapper which takes a training set of weighted examples and use *sampling with replacement* to generate a *new* training set of the *same* size, according to the weight of each example; the wrapper then passes on this set to the decision stump and obtains the classifier. Plug this wrapper and your decision stump implementation into the AdaBoost algorithm. Run the experiments with the dataset from Homework 1 and answer the questions. Use *information-gain* as the evaluation criterion. Do not use pruning.

- Compare and analyze the accuracies obtained by different learners: decision stump alone, boosting with 30 rounds, your ID3 implementation.
- Compare and analyze the accuracies obtained by boosting with different numbers of *rounds*: 5, 10, 20, 30.