

ComS 573 Machine Learning

Lab 4 (100 points)

Ensemble Learning

Note:

- For help with laboratory assignments, please contact TA.
- You may perform the experiments using a machine learning package. (For example the package "caret" in R (<https://topepo.github.io/caret/available-models.html>) has all the algorithms used in this lab.)

In this lab assignment, you will experiment with ensemble classifiers.

1 Dataset

- We will use the Blood Transfusion Service Center Data Set from the UC Irvine Machine Learning Repository:
<https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>

We have split the above data into training (train.csv) and testing (test.csv) sets. In the experiments, please train your models with the training data, and report the performances on the test data.

2 Tasks

2.1 Task 1

- Train ensemble models Random Forest (RF) and AdaBoost.M1 with decision stumps as base classifiers (Note you may use other versions of AdaBoost if the machine learning package you choose to use does not have this particular version.). Experiment with different values of hyper-parameters such as number of base classifiers etc.
- Report your experimental results. Report, for the best models you have learned, the corresponding hyper-parameters and the performance including overall classification accuracy and confusion matrix over the test data. Discuss the results.

2.2 Task 2

- Train 4 individual models: Neural Network (NN), Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT). Report the confusion matrix and classification accuracy on the test data for each of them. When training the 4 models, slightly tune the hyper-parameters (extensive grid search is not required). Report the experiments you have done.
- Construct an ensemble classifier using unweighted majority vote over the 4 models you have trained. Report the performance on the test data.

- Construct an ensemble classifier using weighted majority vote over the 4 models you have trained. Report the performance on the test data. You might use one of the following strategies to decide weights: make weights proportional to the classification accuracy, tune weights as hyperparameters, use stacking, or some other strategies. Report the experiments you have done.
- Discuss the results.

3 What to turn in

Turn in via Canvas (a compressed .zip file) the following:

- A lab report (in pdf file) with your experimental results and discussions of these results.
- All of your commented source code that you may have written.
- Readme file with instructions on how to reproduce your experiments.