

Coding project: Multi-class boosting

Kai Wu

December 17, 2014

1 VALUE-ADDED

- Implement different split criteria for the decision tree, including information gain for ID3 and information gain ratio for C4.5, can switch between these two using *options.dt_type = 'ID3'* and *options.dt_type = 'C4.5'*.
- Implement post-prune for the tree construction algorithm using reduced error pruning. (sub-routine $t = \text{prune}(t, Xvld, yvld, y)$ in *decision_tree.mat*).
- Implement both one-vs-rest multi-class classification using decision stumps as base learners and SAMME [1] using decision trees as base learners, can switch between these two using *options.classifier == 'decision_stump'* and *options.classifier = 'decision_tree'*.

2 COLLABORATION

I collaborated with Shashin Sharan who implemented multi-class bagging classification.

Each algorithm consisted of 50 trials. At each trial, 0.3 of the data examples were selected at random and set as the test set. The remaining 0.7 of examples were used to train the algorithms. The number of base learners for boosting is 30, and the number of bootstrap sample data is 200. I record the test errors and train-test time per iteration to compare the performance of these two algorithms. I tested using the data set *iris* from *UCI Machine Learning Repository* [2], results in Figure 2.1, source code see *model_compare_demo.mat*.

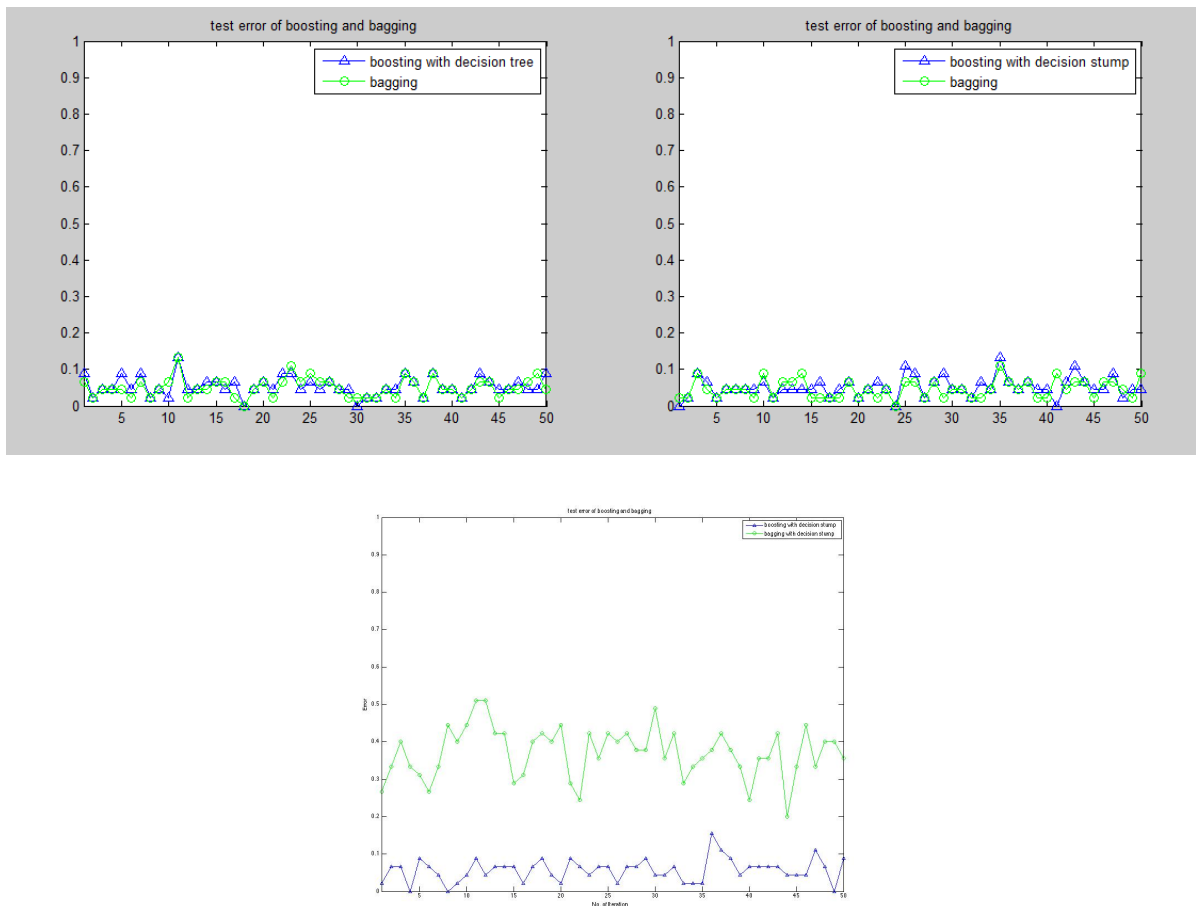


Figure 2.1: Test error rate of boosting using decision trees (SAMME) vs. bagging with decision trees, boosting using decision stump (one-vs-rest) vs. bagging with decision trees, boosting using decision stump (one-vs-rest) vs. bagging with decision stumps (one-vs-rest).

Table 2.1: Training and test time for each method

Method	Time(sec)
Adaboost with decision trees	10.212
Adaboost with decision stumps	0.854
Bagging with decision trees	73.741

We can conclude from the comparison that Adaboost with decision trees (SAMME) and Adaboost with decision stumps (one-vs-rest) can achieve similar results as bagging using

decision trees, but need significantly less time to train the model. And Adaboost with decision trees or decision stumps can achieve much better results than bagging using decision stump

REFERENCE

- [1] Zhu J, Zou H, Rosset S, et al. Multi-class adaboost[J]. Statistics and Its, 2009.
- [2] UCI Machine Learning Repository. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.