CS 189: HW5 Report

Ben Augarten, Richard Hwang, Max Johnson

April 29, 2013

1 Features Implemented

- Decision Trees
  - Impurity Function: Entropy
  - Impurity Function: Gina
  - Impurity Function: Misclassification
  - Stopping Criteria: Number of Points
  - Stopping Criteria: Impurity Reduction
- Random Forest
  - Random selection of features
  - Random selection from K best split points
- Boosted Trees (AdaBoost)

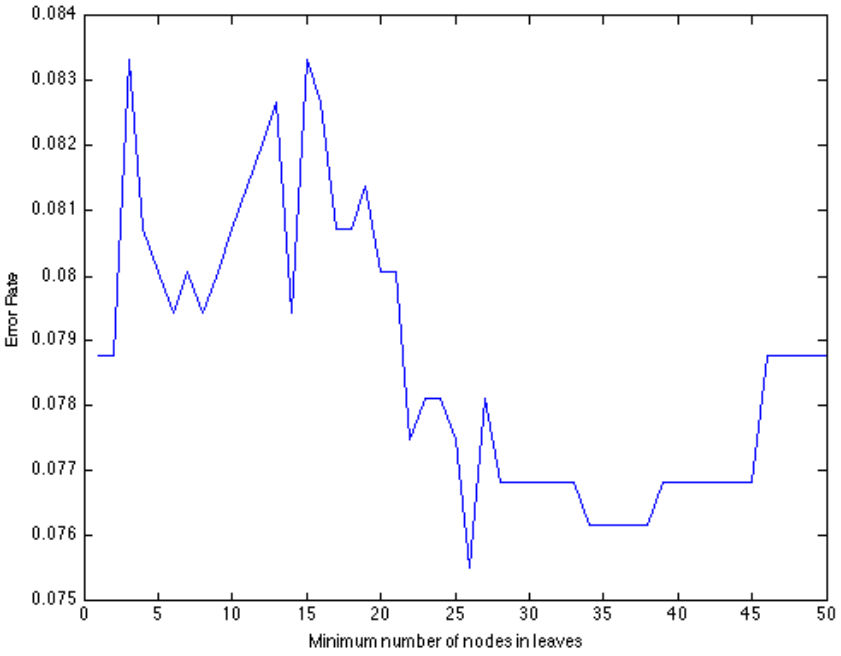
2 Results

2.1 Decision Trees

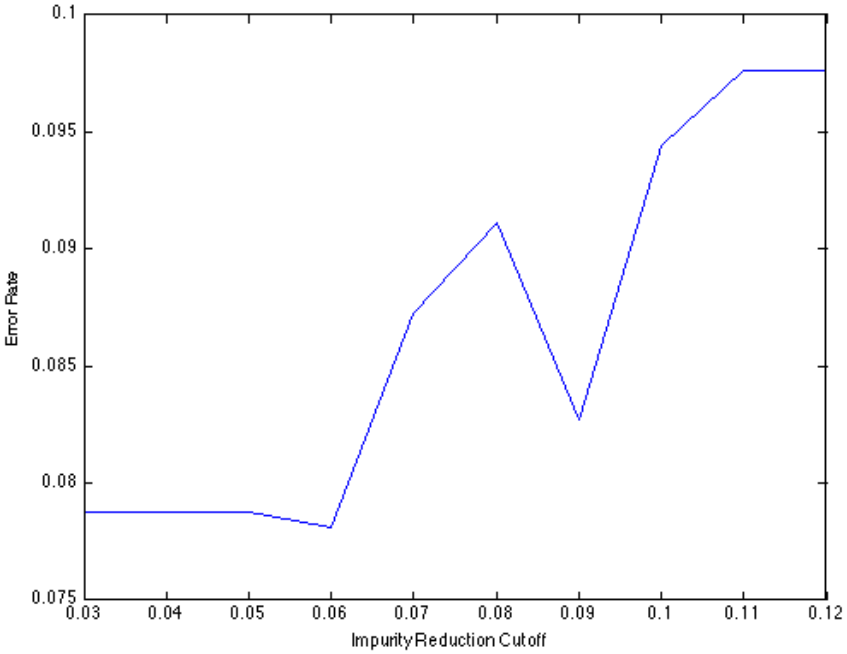
We implemented Decision Trees with 2 different stopping criteria: Stopping based on the number of points, and stopping based on the reduction of impurity. We also implemented three different impurity functions; Entropy, Gina Impurity, and Misclassification.

2.1.1 Stopping Criteria

Entropy, fully grown: 0.0788
Entropy, with leaves of at least 26 nodes: 0.0755
Entropy, error vs n nodes:

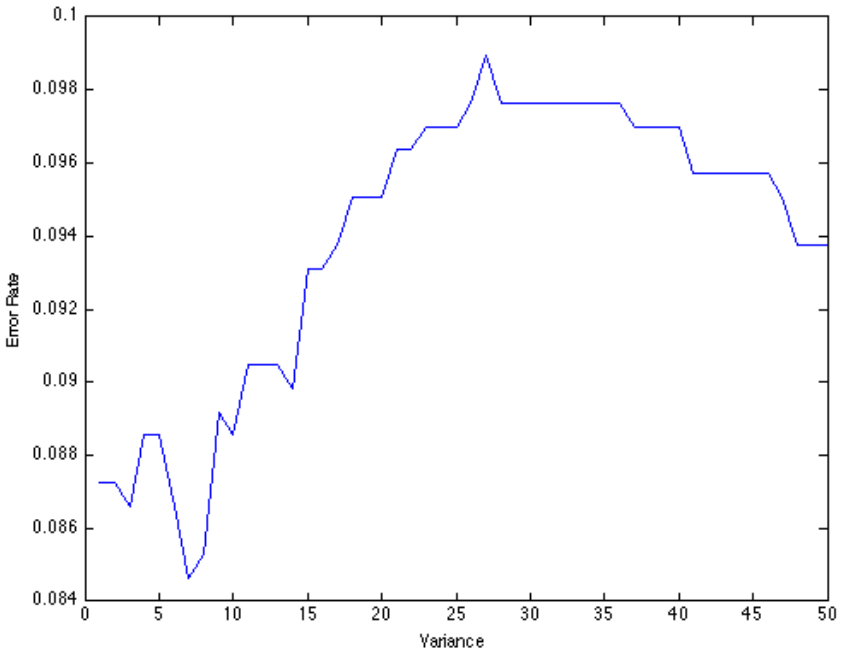


Entropy, fully grown: 0.0788
Entropy, with an impurity reduction cutoff of 0.6: 0.0781 Entropy, error vs n nodes:

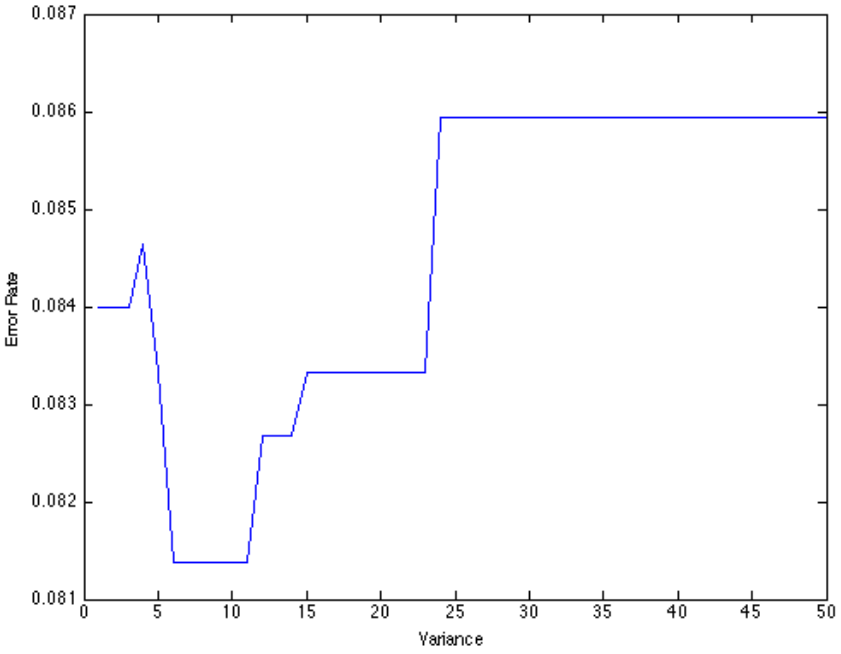


2.1.2 Impurity Functions

Variance, fully grown: 0.0872
Variance, with leaves of at least 7 nodes: 0.0846
Variance, error vs n nodes:



Misclassification, fully grown: 0.0840
Misclassification, with leaves of at least 6-11 nodes: 0.0814
Misclassification, error vs n nodes:



2.2 Random Forests

2.2.1 Random feature selection

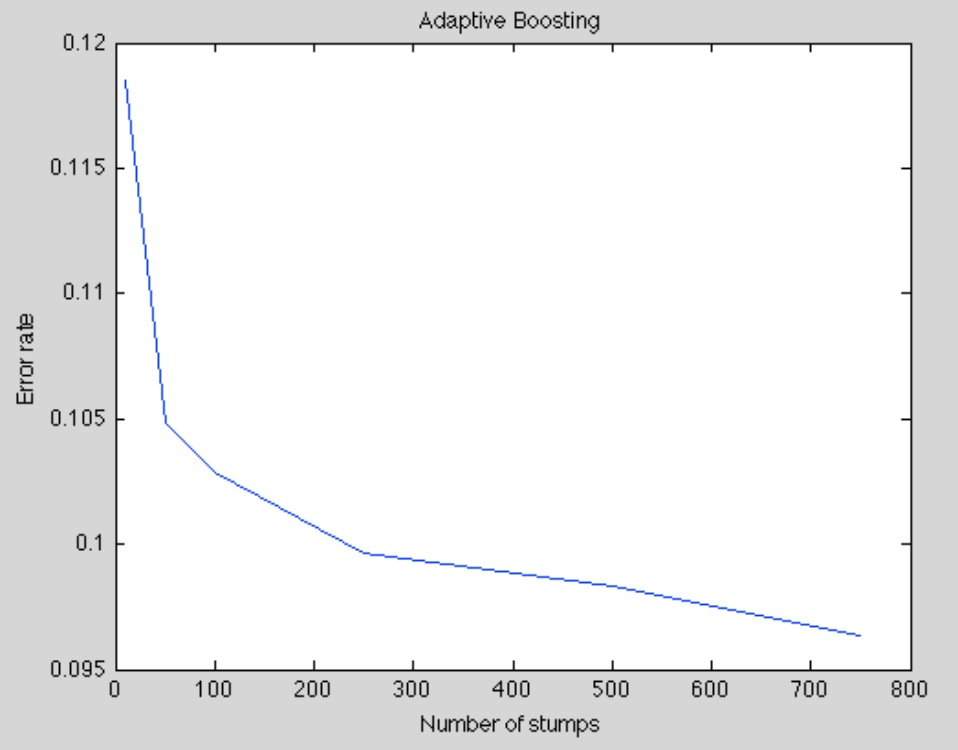
Achieved an error rate of 0.0456 with a forest of 150 nodes selecting splits from 5 random features.

2.2.2 Random selection from top-k split points

Achieved an error rate of 0.0612 with a forest of 200 trees, selecting from the top 200 nodes.

2.3 Boosted Trees

We implemented the AdaBoost algorithm by boosting weak decsion trees of depth 2. We found this worked better than stumps, as too many of the stumps were identical.
With 750 trees, we achieved an error rate of 0.0964.



3 Summary

We achieved the best result using Random Forests with random feature selection. The best error rate we achieved was 4.6%.

4 References

- Included Materials
- `http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm`
- `http://en.wikipedia.org/wiki/Random_forest`