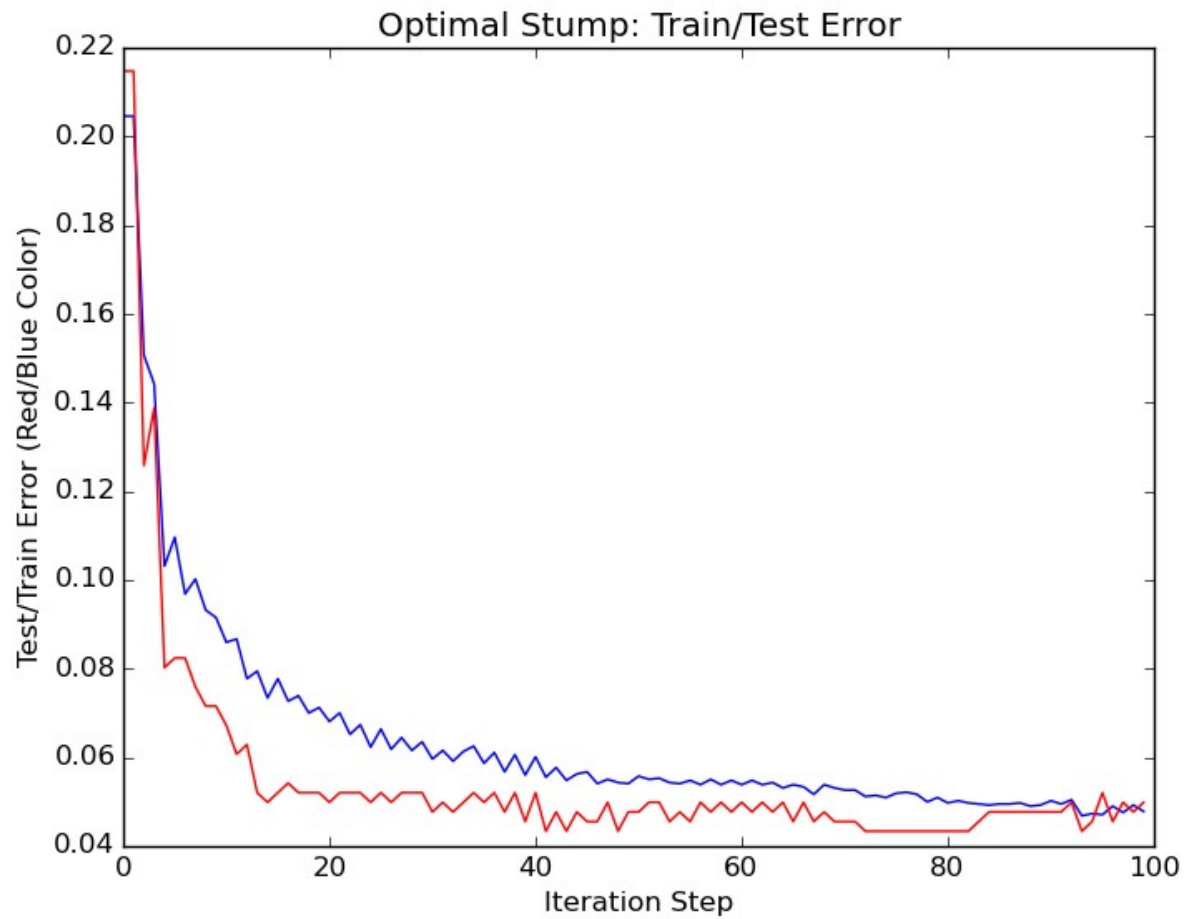


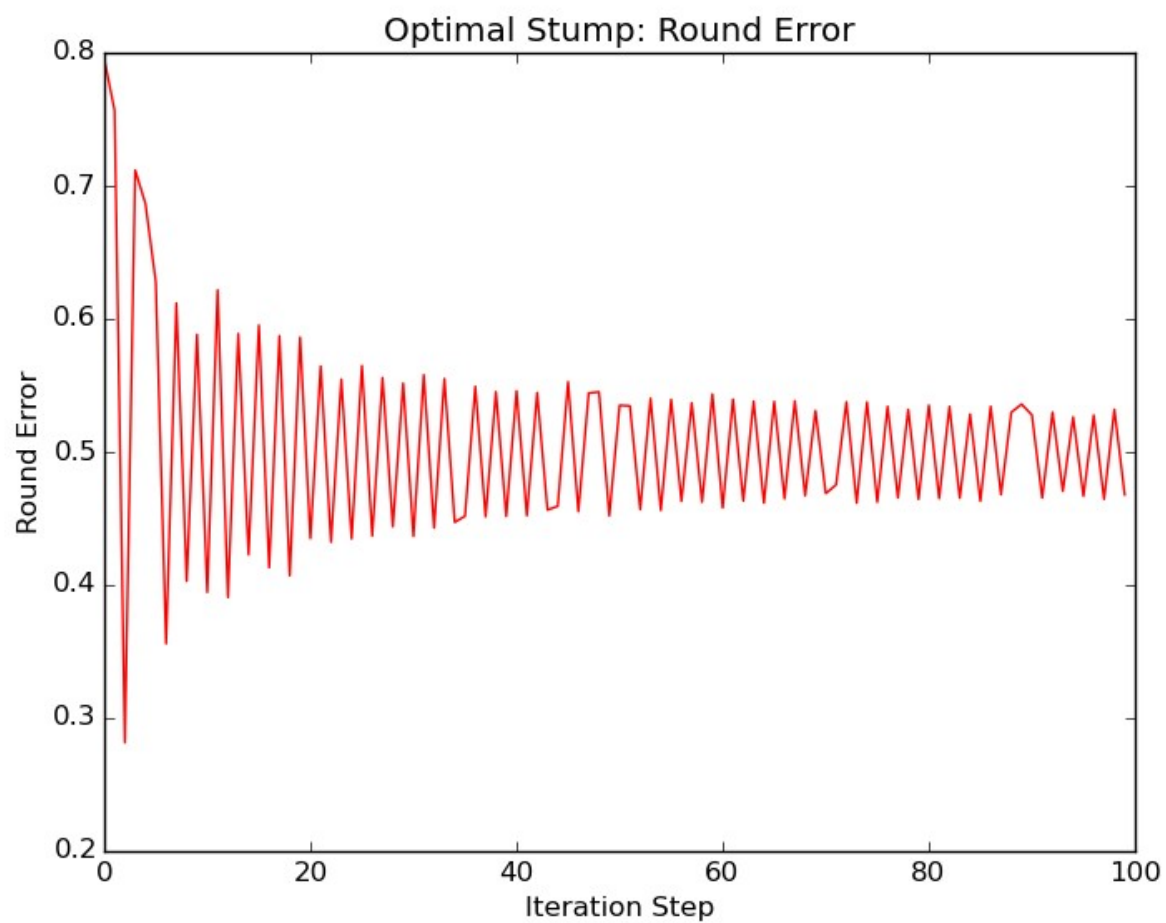
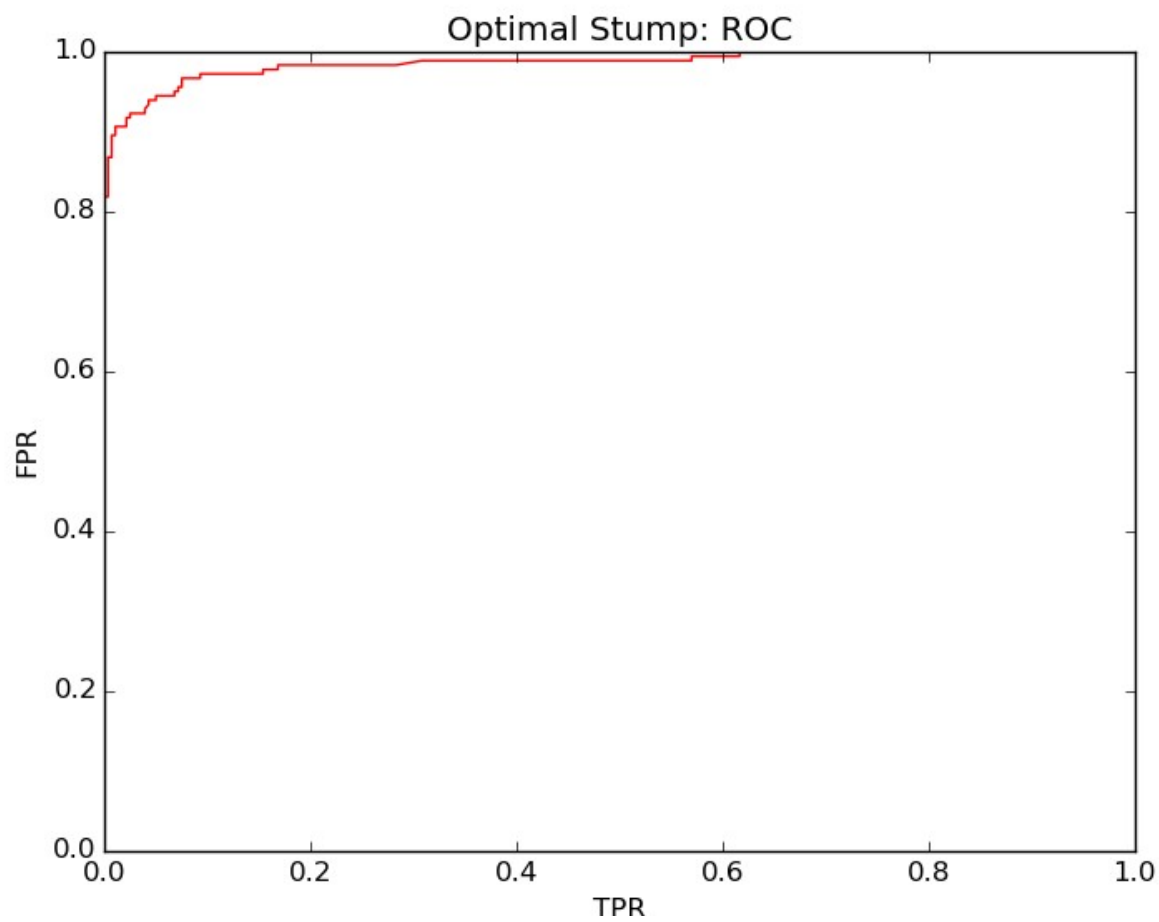
Luke Jaffe
CS6140 Machine Learning
11/07/15

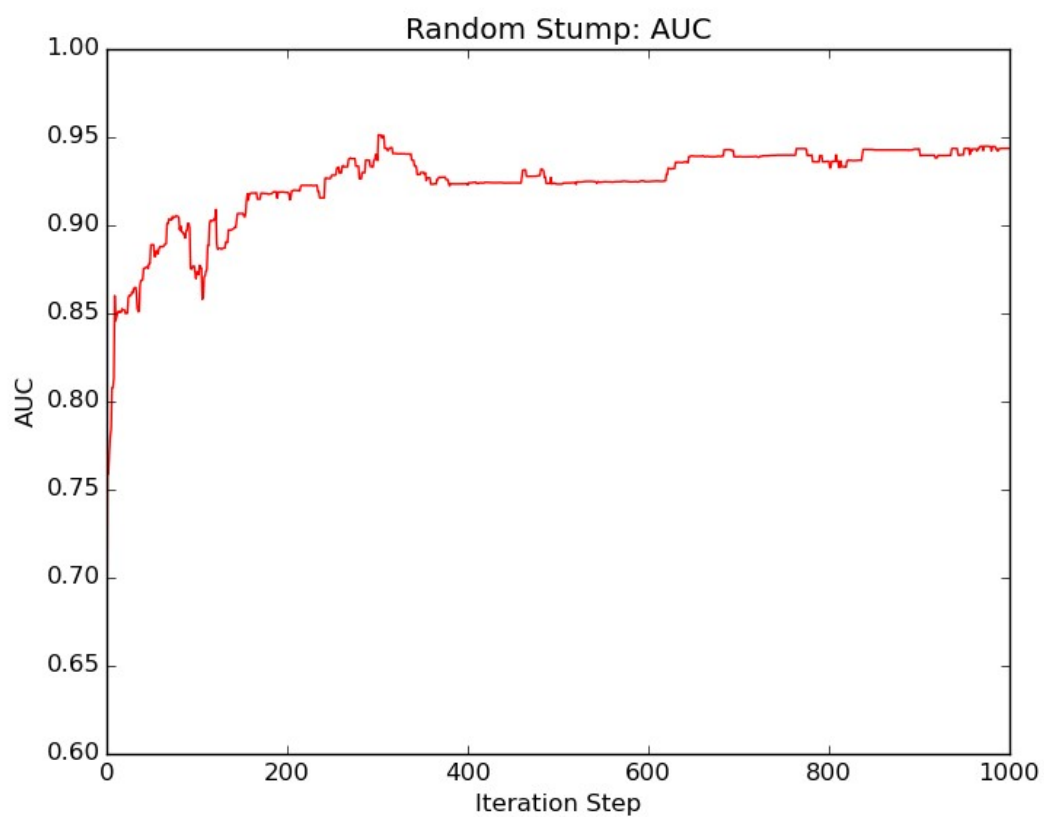
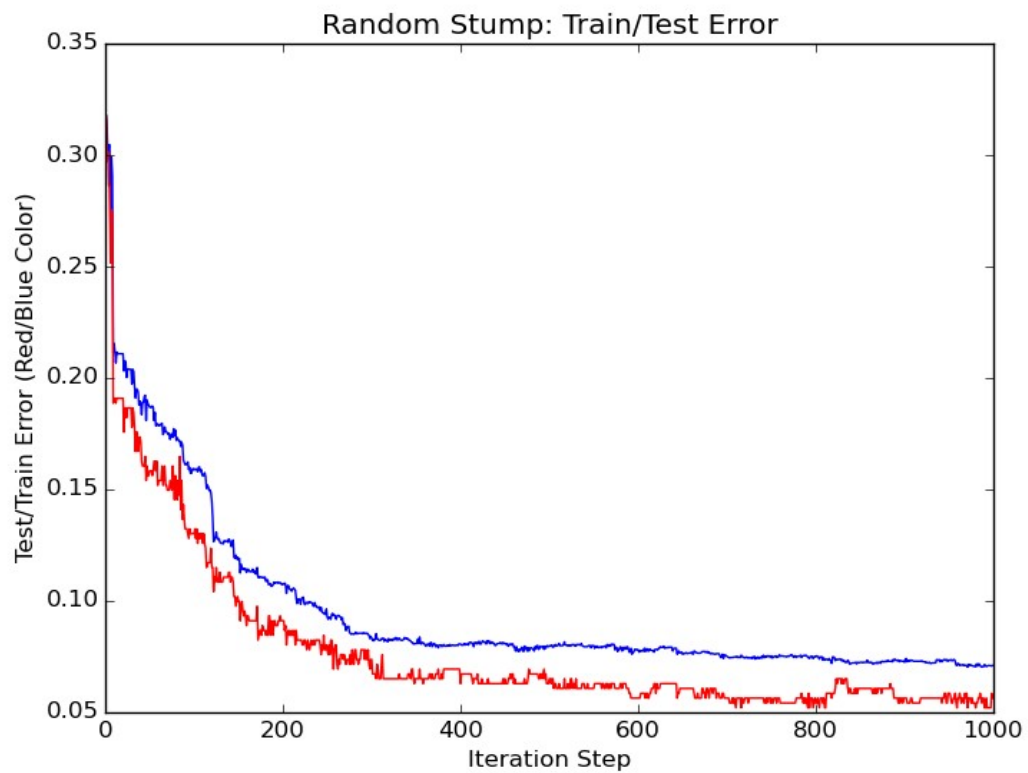
HW4 – Boosting and Bagging

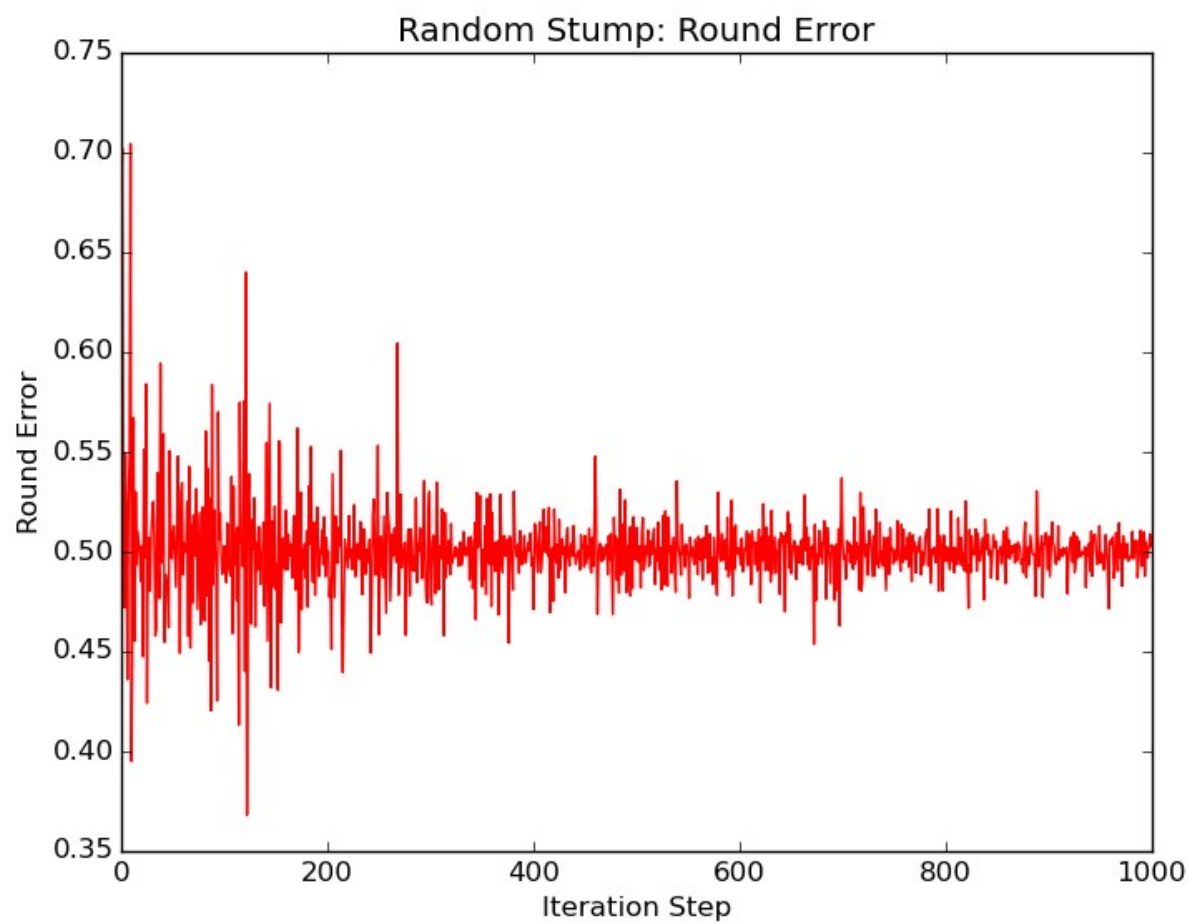
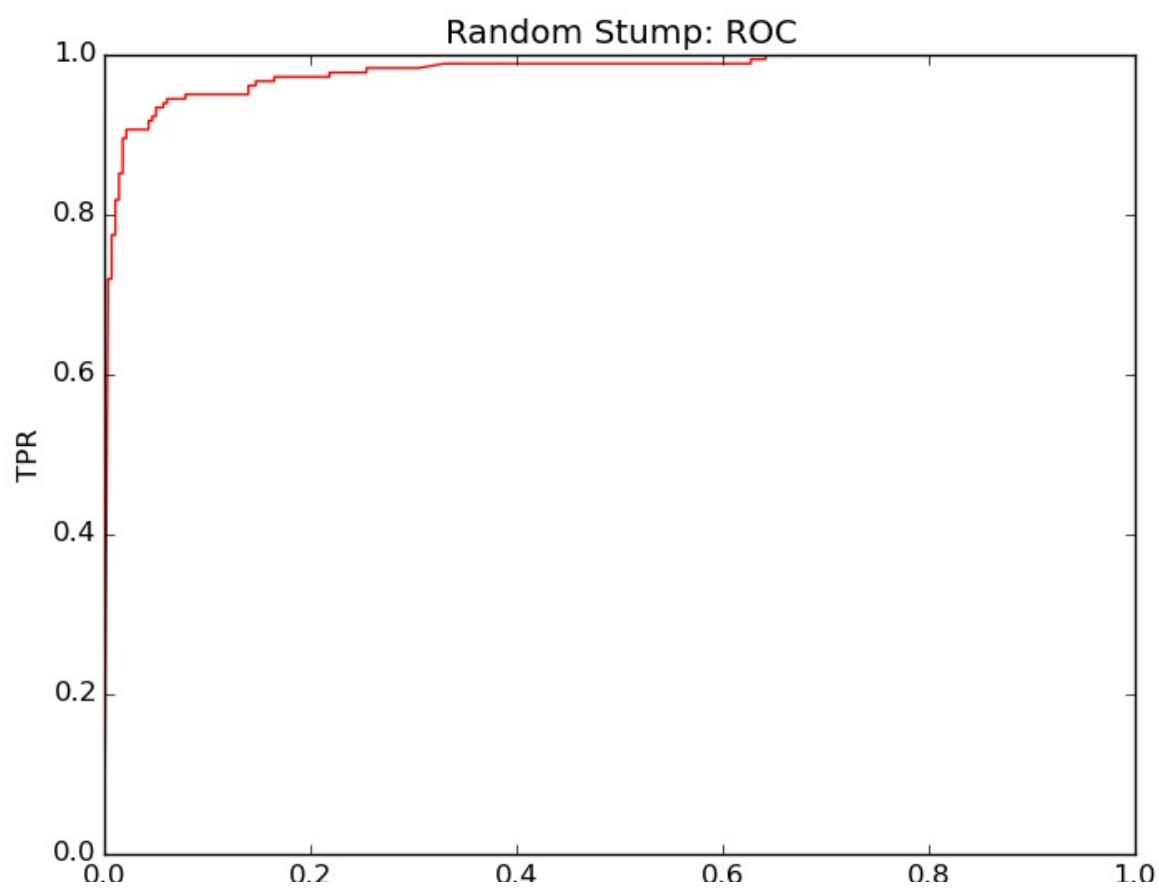
Problem 1. Adaboost Code

Usage: `python adaboost.py`









Problem 2. Adaboost on UCI Datasets

a)

Usage: *python kboost.py -d {'crx', 'vote'}*

Vote:

Average test_acc: 0.951638477801

CRX:

Average test_acc: 0.859420289855

b)

Usage: *python cboost.py -d {'crx', 'vote'}*

Vote:

c%=5%, train error: 0.00 test error: 0.14 AUC: 0.87

c%=10%, train error: 0.00 test error: 0.14 AUC: 0.86

c%=15%, train error: 0.00 test error: 0.11 AUC: 0.88

c%=20%, train error: 0.00 test error: 0.07 AUC: 0.93

c%=30%, train error: 0.01 test error: 0.06 AUC: 0.96

c%=50%, train error: 0.04 test error: 0.04 AUC: 0.97

c%=80%, train error: 0.04 test error: 0.04 AUC: 0.98

CRX:

c%=5%, train error: 0.00 test error: 0.41 AUC: 0.61

c%=10%, train error: 0.00 test error: 0.26 AUC: 0.82

c%=15%, train error: 0.00 test error: 0.19 AUC: 0.87

c%=20%, train error: 0.00 test error: 0.19 AUC: 0.86

c%=30%, train error: 0.02 test error: 0.20 AUC: 0.86

c%=50%, train error: 0.06 test error: 0.14 AUC: 0.90

c%=80%, train error: 0.09 test error: 0.11 AUC: 0.95

Problem 3. Active Learning

Usage: *python active.py*

Problem 4. Error Correcting Output Codes

Usage: *python ecoc.py*

Problem 5. VC Dimension

a)

b)

c)

d)

Problem 6. Bagging

Usage: python tree.py

Problem 7. Gradient Boosted Trees for Regression

Usage: python rtree.py

With 10 rounds of boosting:

Training MSE (boosting): [16.83690229]

Testing MSE (no boosting): [496.32841925]

Testing MSE (boosting): [34.53110545]