James Abundis and Thomas Fahrner

jlabundi@calpoly.edu and tfahrner@calpoly.edu

Decision Tree Project

In this report we examine the results of running decision tree prediction algorithms against ELECTION data and IRIS data. Specifically, we compare the accuracy of C4.5 with 10-fold cross validation to Random Forests (5 trees).

**Tree 01 100:**

* 10-fold: 0.91
* Random Forest: 0.955

**Tree 01 1000:**

* 10-fold: 0.969
* Random Forest: 0.987

**Tree 02 20:**

* 10-fold: 0.9
* Random Forest: 0.95

**Tree 02 100:**

* 10-fold: 1.0
* Random Forest: 1.0

**Tree 03 20:**

* 10-fold: 0.8
* Random Forest: 0.9

**Tree 03 100:**

* 10-fold: 0.97
* Random Forest: 0.96

**Iris:**

* 10-fold: 0.78
* Random Forest: 0.836

We see that by and large Random Forests perform much better than just C4.5 even with a forest size of 5. Despite the k-fold validation efforts applied to the C4.5 algorithm, the predictive power that bagging provides outweighs the subsampling benefits of validation techniques.