

# CS57300: Assignment 4

Pavani Guttula  
pguttula@purdue.edu  
late days used -2

April 2, 2019

## 1 Implement Decision Trees, Bagging and Random Forests

My accuracies for Decision Trees with a tree depth = 8:

Training Accuracy DT 0.78

Test Accuracy DT: 0.71

My accuracies for Bagging with depth = 8 and number of trees =30:

Training Accuracy BT: 0.79

Test Accuracy BT: 0.75

My accuracies for Random Forests with depth = 8 and number of trees =30:

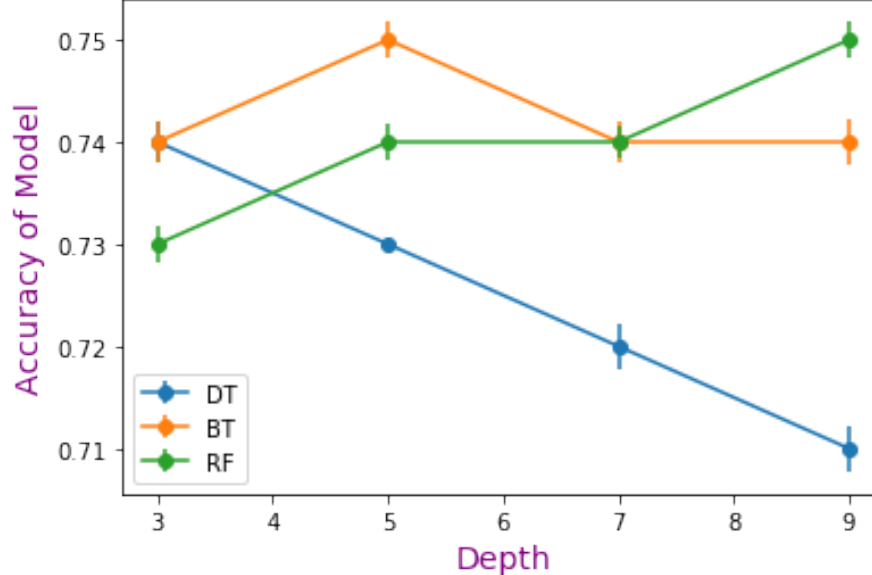
Training Accuracy RF: 0.78

Test Accuracy RF: 0.73

## 2 Influence of Tree Depth on Classifiers

Below is my graph for all accuracies of all the 3 models over a range of tree depths:

## Depth vs Accuracies for DT,BT& RF



I am comparing the accuracies of my Decision Tree and Random Forest models for different depths of tree. My Null Hypothesis is that performance is same for both these models.

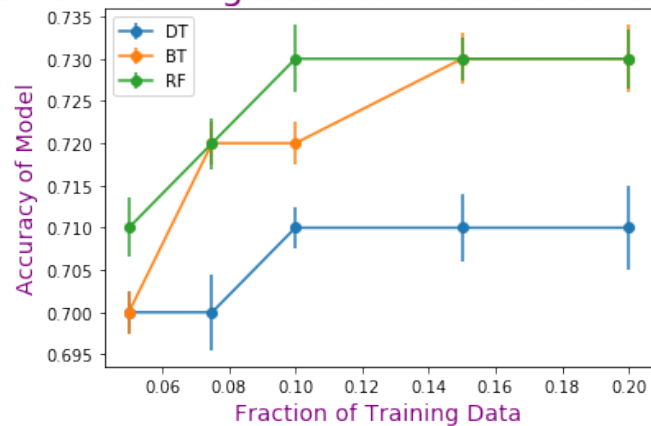
I tested my hypothesis again using two tailed paired t test. I compared the distribution of accuracies for different fractions of Training data for both the models. My t test resulted in a probability of 0.24 which is large enough(alpha value = 0.05) such that we cannot reject our Null hypothesis. So we have enough evidence to believe that the probability of our null hypothesis being true is not really low.

So I accept my null hypothesis and reject my non directional alternative Hypothesis that there is a difference in the distribution of accuracies of Decision Tree and Random Forests.

### 3 Performance of Different Models for various fractions of training data

Below is my graph for all accuracies of all the 3 models trained over different fractions of data.

## Fraction of Training Data vs Accuracies for DT,BT& RF



I am comparing the accuracies of my Decision Tree and Bagging models for different fractions of training data. My Null Hypothesis is that performance is same for both these models.

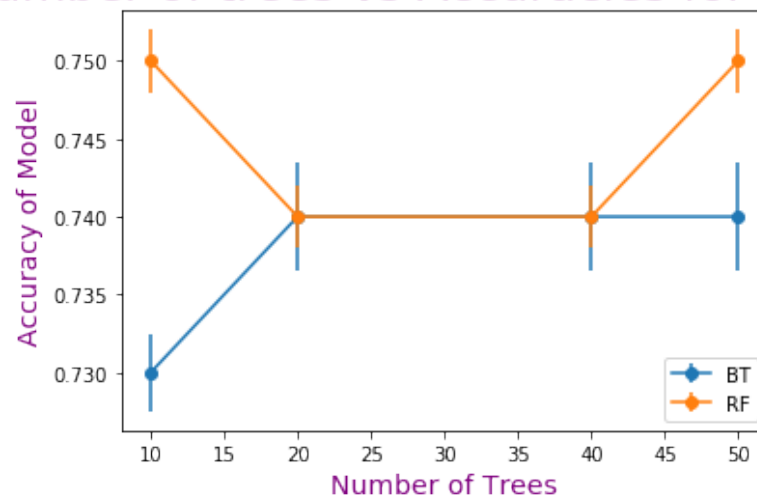
I tested my hypothesis using two tailed paired t test. I compared the distribution of accuracies for different fractions of Training data for both the models. My t test resulted in a probability of 0.024 which is less than the alpha value 0.05. So we have enough evidence to believe that the probability of our null hypothesis being true is very low.

So I reject my null hypothesis and accept my non directional alternative Hypothesis that there is a difference in the distribution of accuracies of Decision Tree and Bagging.

## 4 The Influence of Number of Trees on Classifier Performance

Below is my graph for the accuracies of all the ensemble models trained over different fractions of data.

### Number of trees vs Accuracies for BT& RF



I am comparing the accuracies of my ensemble models Bagging and Random Forest for by training them on different number of trees. My Null Hypothesis is that performance is same for both these models.

I tested my hypothesis again using two tailed paired t test. I compared the distribution of accuracies for different fractions of Training data for both the models. My t test resulted in a probability of 0.215 which is large enough(alpha value  $> 0.05$ ) such that we cannot reject our Null hypothesis. So we have enough evidence to believe that the probability of our null hypothesis being true is high.

So I accept my null hypothesis and reject my non directional alternative Hypothesis that there is a difference in the distribution of accuracies of Decision Tree and Random Forests. I conclude that there is no difference between the performances of both the models.