ENGR421 – HW5 REPORT

BERK BAHADIR BAHCETEPE

0071533

For this homework, I started with splitting the test and training data as required in the homework 5 guideline. I assigned the first 150 data points to training set and the remaining 122 data points to test set. Later, with the help of Lab7 and things Mehmet Hoca said at the end of the class, I created 3 functions such as one for rmse calculation, one for decision tree regression calculation and one for predicted value calculation. In decision tree calculation I used the material that Mehmet Hoca provided us in Lab7. I defined the stopping condition as such: if total number of data points in a node becomes less than or equal to P, which is the pre-pruning parameter, that node becomes a terminal node. After that I defined the error function as Mehmet Hoca told as in the end of the Lab7. I coded this function for both univariate and multivariate cases. Even though I knew the data set given to me was an univariate data set and I did not have to thus iterate through the features, in order to it to work for all types of data set, I iterated through the features just in case (for this case I iterated only once since it was an univariate data set). I returned node\_splits and is\_terminal from this function since they would be beneficial in calculating the predicted values. Later, I defined a prediction function and a rmse function which is similar to the one in the previous homework. After all, I calculated the rmse values for both training and test sets and for different values of P and obtained accurate graphs compared to the ones in the homework 5 guidelines. Here are the graphs that I obtained:

