İrem Şahin

ENGR421

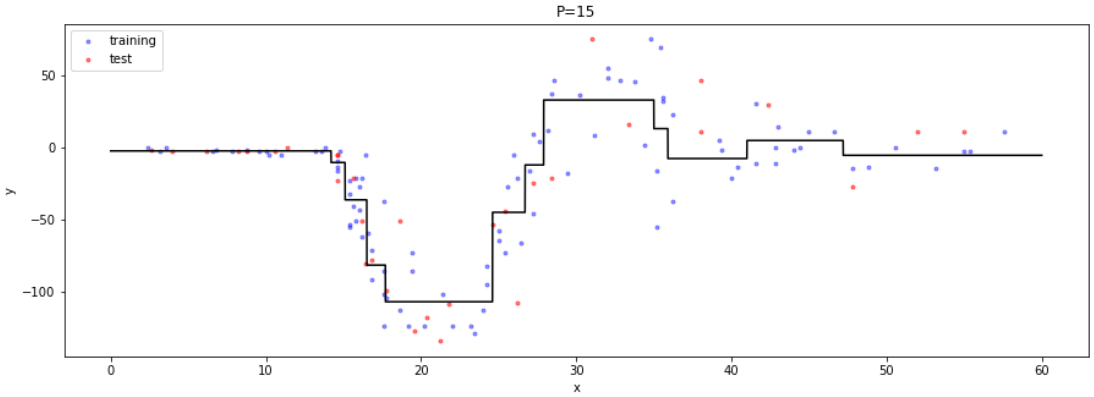
Mehmet Gönen

HW5-Report

For this project, I have written my code in a Jupyter Notebook .ipynb file. I have started my code by importing necessary libraries and writing our parameters down which were given in the pdf file. After this I have imported our data set from the given .csv files. Then, as requested in the pdf file, I have divided the first 100 samples of each class for the training set, and the remaining 33 samples for the test set. Then, I have also divided them as x and y for future use. After using normal list appending, I just turned them into numpy arrays to use their transpose,shape etc. built-ins.

In the next cell, I have defined the decision tree learning algorithm. In this algorithm, if a node has less points than our P value(prune), then we have determined that node as a terminal node(leaf in a tree). Else, by trying every split candidate and calculating their errors, we have determined the best split position, which is the one with the minimum total error/length of right+left indices. We also make a node terminal after this step if all data are the same, i.e. len of unique(x_train) is 1. After selecting the best split, we have determined the new right and left indices and created them for the next iteration. After this I have also defined the prediction function which predicts the nodes place according to the tree we have created.

Next, I have plotted the data according to our Lab6, which looked like below:



Lastly, I have done the same for other P datas given in the pdf and calculated their RMSE values in a list. Then, I have plotted this like below to see the performance: