Mertcan Aşgün

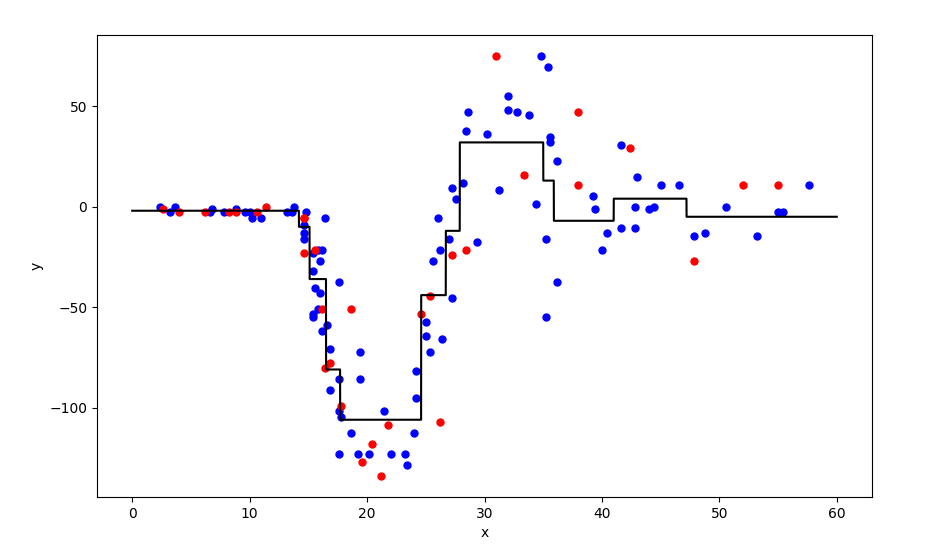
63948

Homework 05: Decision Tree Regression

Our aim was to classify the given dataset by using decision tree regression. For this whole assignment, I mainly used the code from the Decision Tree Lab. I started by reading the data from the given .csv file. As a remark, the name of the .csv file was labeled as “hw4”, that is why I used those name in my code. After reading the data, I split it into training and test sets.

Then comes the main tree inference. I used the same structure from the lab. Just changed some parts to make it regression, instead of classification. For this purpose, for example I removed node\_frequencies since we did not need it anymore, changed the split\_scores etc.

After defining tree inference as a method, I was able to call it to make predictions on any given data set. For this purpose, I defined a new method to make this prediction. And directly applied it on our train set to get predicted y values with P value set as 15. After I have both data points and predicted y values, I plotted the graph:



Then I needed to calculate the RMSE value. Since in the next parts we will need to calculate other RMSE values for different P values, I defined a method to get RMSE values. The main idea is same as the previous assignment’s RMSE calculations:



I don’t know the reason why the RMSE value is not exactly the same with the RMSE value in the handout, but I suppose it won’t be much problem.

After defining a method for getting RMSE values, I was ready to get different RMSE values for different P values. For this purpose, in a for loop, I calculated different RMSE values. The resulted graph is:

