Tutku Kılıçaslan

54201

ENGR 421 HW5

Regression Tree

1. First and foremost, I used the code structure of our instructor that he had used in the lab session. Since the classification and regression trees are very similar I only made few changes.
2. I deleted for loop (for (d in 1:D)) since we do not have classes in our case.
3. Since we do not have classes, I erased the dimension of all values. x\_train and y\_train have only one feature.
4. I changed our constraint as being smaller than 25 data points in each node.
5. I sorted unique x values.
6. I try to find all possible split values between each unique x values.
7. I found left and right indicies for each split.
8. I found the mean value of x values going to left node and right node separately.
9. I calculated errors for each split.
10. I chose the split giving minimum error for each node.
11. I created left and right node according to the split I have chosen.
12. I applied these until each node has more than 25 data points.
13. After creating all nodes, I extract the rules of each terminal nodes.
14. I ploted the graph of splits.
15. I calculated the RMSE value as given in the homework description.
16. I put the my code into for loop for each P values given in the homework.
17. I caculated all the RMSE values for each P values and put them into an array RMSE.
18. I ploted the RMSE values versus P values and highlight the RMSE values for each P by putting dots as in the given plot.

Note: I encounter with lots of dimensional problems with my data points. Therefore, I used x\_train and y\_train as if they were matrices. This may confuse the grader but my code gives the same values as the instructor’s.