

DASC521 HW#5 REPORT

In this homework, we have implemented regression trees with the training data and evaluated the models with the test data. Data is divided into two from the best split location, recursively, until the number of data points in the node is less than or equal to the pre-pruning parameter. The error of each possible split is calculated, split with the minimum error value is selected as the split location. The error of a split is calculated using the following formula:

$$\frac{N_{left\ node} * RMSE_{left\ node} + N_{right\ node} * RMSE_{right\ node}}{N_{parent}}$$

However, my implementation did not exactly result as the one in the description.

Results are shown in the following figures:

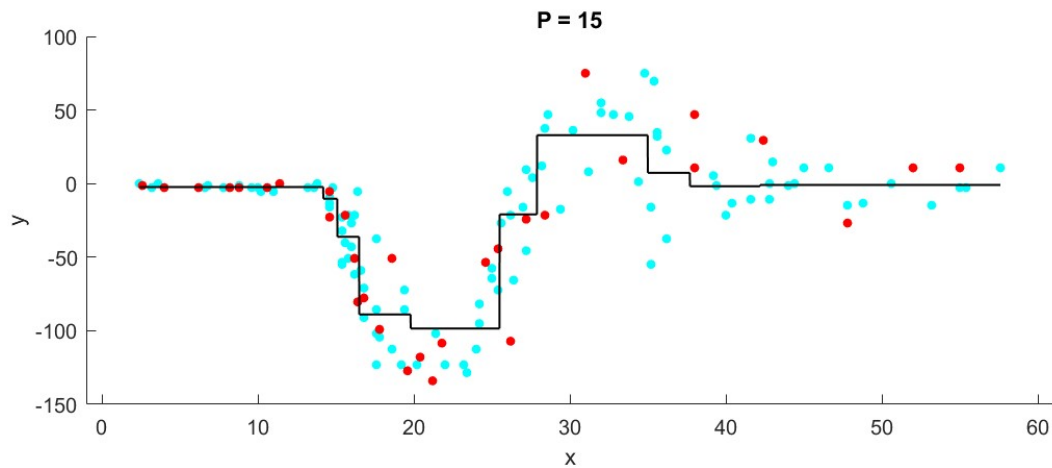


Figure 1 Decision tree when pre-pruning parameter is set to 15

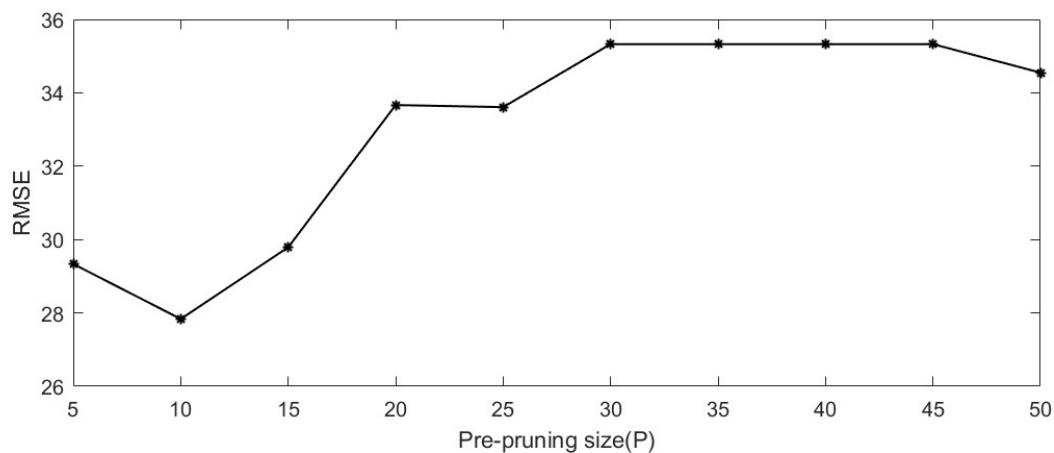


Figure 2 RMSE for different P values on a plot