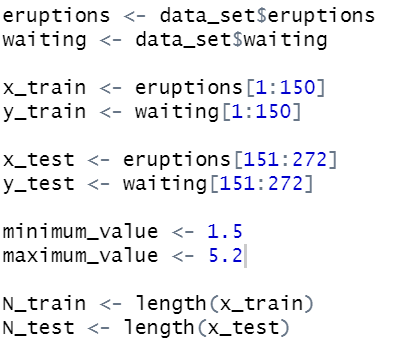
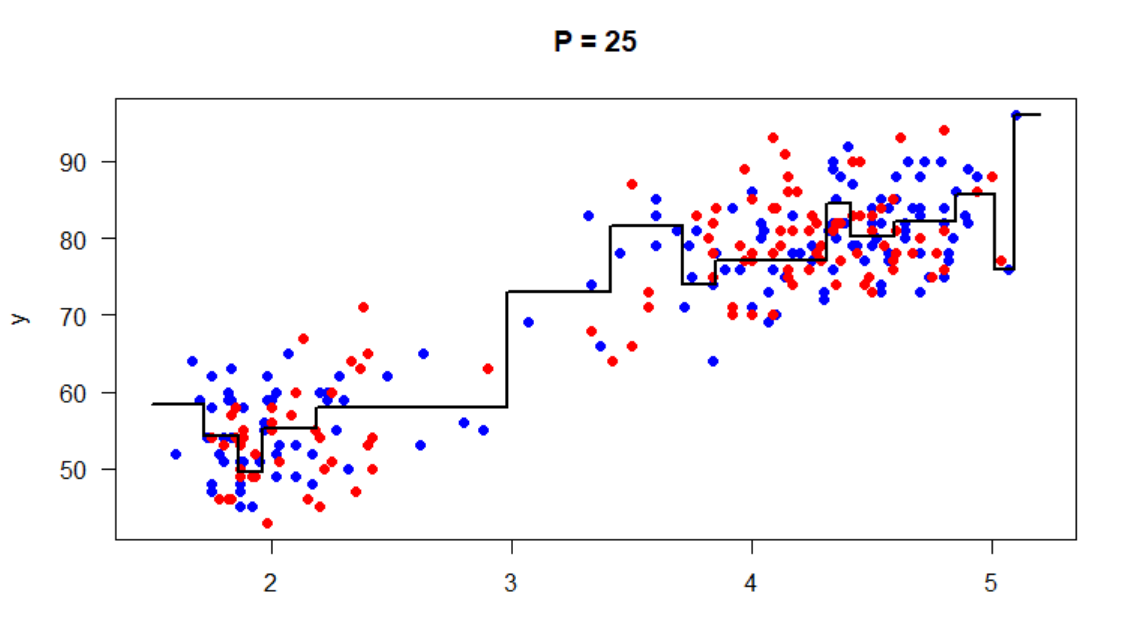
HW05 REPORT

In this homework, I implemented a decision tree regression algorithm in R. We are given a univariate regression data set, which contains 272 data points about the duration of the eruption and waiting time between eruptions for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

As a first step I divided data set into two parts by assigning the first 150 data points to the training set and the remaining 122 data points to the test set after dividing data set into eruptions and waitings vectors as shown below.

I implemented a decision tree regression algorithm using the following pre-pruning rule: If a node has P or fever data points, convert this node into a terminal node and do not split further, where P is a user-defined parameter which is DecisionTreeRegression function in my implementation.

After learning a decision tree by setting the pre-pruning parameter P to 25, I plot the training data points, test data points, and fit in the same figure that is shown below.



After that, I calculated the root mean squared error for test data points using get\_prediction function in my implementation. I found the following result.

C:\Users\26045323\AppData\Local\Microsoft\Windows\INetCache\Content.Word\3.png

As a final step, I learned decision trees by setting the pre-pruning parameter P to 5, 10, 15, …, 50. Then, I plot the graph of RMSE (root mean squared error) for test data points as a function of P. Graph can be shown below.

