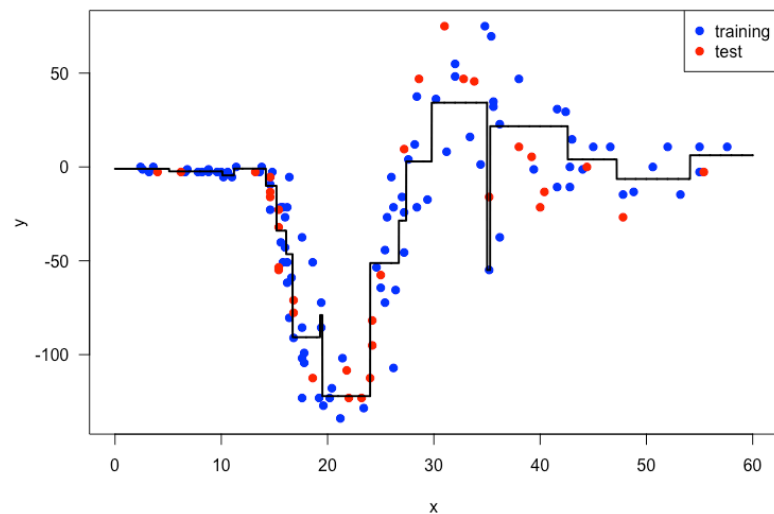


## COMP/INDR 421/521 HW05: Decision Tree Regression

Deadline: November 24, 2017, 11:59 PM

In this homework, you will implement a decision tree regression algorithm in R, Matlab, or Python. Here are the steps you need to follow:

1. You are given a univariate data set, which contains 133 data points, in the file named hw05\_data\_set.csv. Randomly pick 100 data points as the training set and the remaining 33 data points will be used as the test set.
2. Implement a decision tree regression algorithm using the following pre-pruning rule: If a node has  $P$  or fewer data points, convert this node into a terminal node and do not split further, where  $P$  is a user-defined parameter.
3. Learn a decision tree by setting the pre-pruning parameter  $P$  to 10. Draw training data points, test data points, and your fit in the same figure. Your figure should be similar to the following figure.



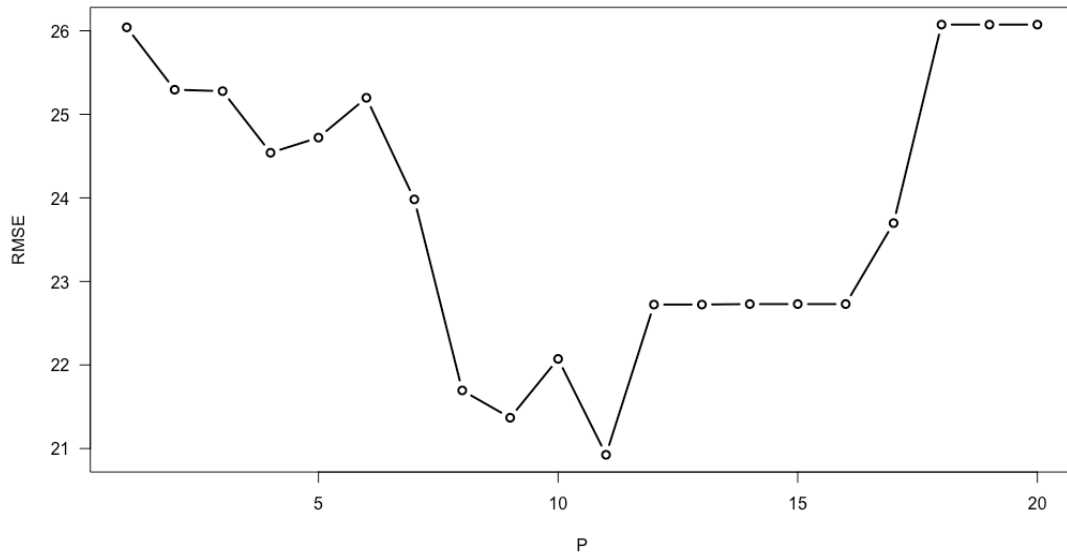
4. Calculate the root mean squared error for test data points. The formula for RMSE can be written as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{\text{test}}} (y_i - \hat{y}_i)^2}{N_{\text{test}}}}$$

Your output should be similar to the following sentence.

```
## [1] "RMSE is 22.0726 when P is 10"
```

5. Learn decision trees by setting the pre-pruning parameter  $P$  to 1, 2, 3, ..., 20. Draw RMSE for test data points as a function of  $P$ . Your figure should be similar to the following figure.



**What to submit:** You need to submit your source code in a single file (.R file if you are using R, .m file if you are using Matlab, or .py file if you are using Python) and a short report explaining your approach (.doc, .docx, or .pdf file). You will put these two files in a single zip file named as ***STUDENTID.zip***, where ***STUDENTID*** should be replaced with your 7-digit student number.

**How to submit:** E-mail the zip file you created to [mehmetgonen@ku.edu.tr](mailto:mehmetgonen@ku.edu.tr) with the subject line ***Intro2MachineLearningHW05***. Please follow the exact style mentioned for the subject line and do not send a zip file named as ***STUDENTID.zip***. Submissions that do not follow these guidelines will not be graded.

**Late submission policy:** Late submissions will not be graded.

**Cheating policy:** Very similar submissions will not be graded.