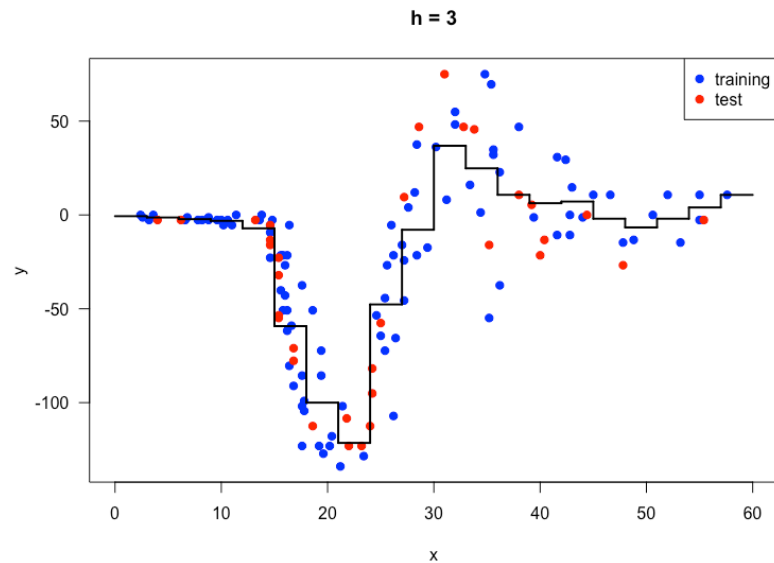


## COMP/INDR 421/521 HW04: Nonparametric Regression

Deadline: November 17, 2017, 11:59 PM

In this homework, you will implement three nonparametric regression algorithms 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 `hw04_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. Learn a regressogram by setting the bin width parameter to 3. Draw training data points, test data points, and your regressogram in the same figure. Your figure should be similar to the following figure.



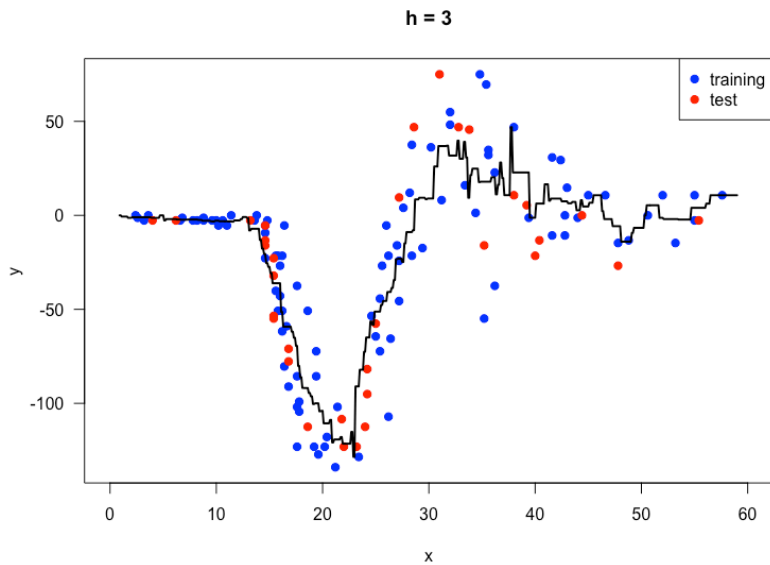
3. Calculate the root mean squared error for test data points. The formula for RMSE can be written as:

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

Your output should be similar to the following sentence.

```
## [1] "Regressogram => RMSE is 24.3717 when h is 3"
```

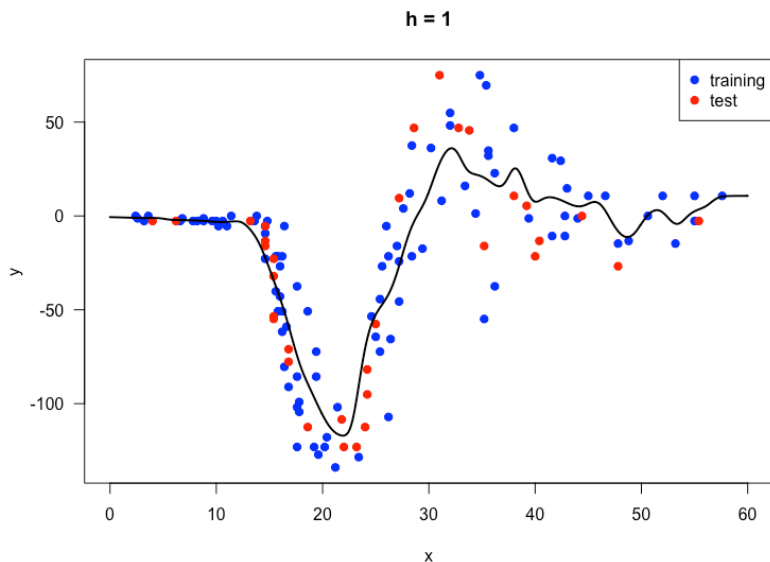
4. Learn a running mean smoother by setting the bin width parameter to 3. Draw training data points, test data points, and your running mean smoother in the same figure. Your figure should be similar to the following figure.



5. Calculate the root mean squared error for test data points. Your output should be similar to the following sentence.

```
## [1] "Running Mean Smoother => RMSE is 21.9220 when h is 3"
```

6. Learn a kernel smoother by setting the bin width parameter to 1. Draw training data points, test data points, and your kernel smoother in the same figure. Your figure should be similar to the following figure.



7. Calculate the root mean squared error for test data points. Your output should be similar to the following sentence.

```
## [1] "Kernel Smoother => RMSE is 22.1088 when h is 1"
```

**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 ***Intro2MachineLearningHW04***. 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.