İrem Şahin

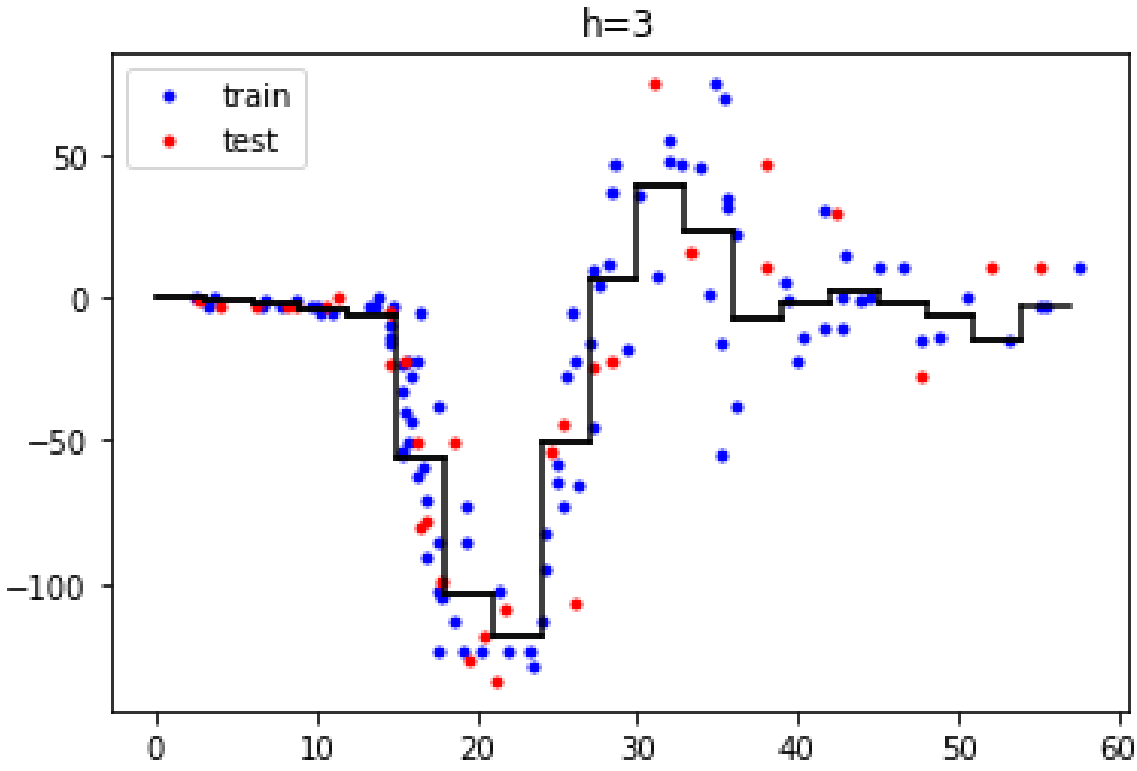
ENGR421

Mehmet Gönen

HW4-Report

For this project, I have written my code in a Jupyter Notebook .ipynb file. I have started my code by importing necessary libraries and writing our parameters down which were given in the pdf file. After this I have imported our data set from the given .csv files. I have continued by taking N(# of samples) values for future use. Then, as requested in the pdf file, I have divided the first 100 samples of each class for the training set, and the remaining 33 samples for the test set. Then, I have also divided them as x and y for future use. After using normal list appending, I just turned them into numpy arrays to use their transpose,shape etc. built-ins.

In the next cell, I have calculated the data intervals and left-right borders for regression. Then, I have calculated the p_hat. After implementing the regressogram with the given bin width, and left& right borders, my regressogram was like this:



I have implemented this function for the purposes above.

$$\hat{g}(x) = \frac{\sum_{t=1}^{N} b(x, x^t) r^t}{\sum_{t=1}^{N} b(x, x^t)} \quad (8.24)$$

where

$$b(x, x^t) = \begin{cases} 1 & \text{if } x^t \text{ is the same bin with } x \\ 0 & \text{otherwise} \end{cases}$$

Then, I have found root mean squared error in my rmse function. The error was

Regressogram RMSE is 28.96065633310612 when h is 3

... which is similar to the expected value in the pdf with some error.

Unfortunately, I got sick after writing this far and couldn't do the other parts. This is my first homework with is partially missing and I am sorry for that.