ENGR421 – HW4 REPORT

BERK BAHADIR BAHCETEPE

0071533

For this homework, I started with reading the data and setting the training and test data points. Later I defined the maximum value and the minimum value which are 5.2 and 1.5 respectively. The main factor in deciding the values were, the latter is the origin point indicated in the instructions and the former was deciding using the following logic: I realized that in regressogram there were 10 bins with width of 0.37. When we increase the starting point (1.5) with the bin width 10 time, we obtain 5.2. Later, I wrote functions for regressogram, running mean smoother and kernel smoother respectively with the help of Lab06 of which I modified the algorithms of histogram, naïve and kernel estimator respectively and the cheat sheet. I plotted the graphs using these functions. After, I coded a function for the root mean squared error in order to compare y\_predicted and y\_test. In order to determine y\_predicted I used x\_test and x\_train data points. The calculations of y\_predicted with the running mean and kernel smoother were easy. The only difference with the functions was instead of iterating through data\_interval, I iterated through x\_test and obtain the accurate results. However, for the regressogram I had to do it from scratch. I calculated checked whether x\_train and x\_test are in the same bin or not one-by-one and if yes multiplied it with y\_train and divided it with the number of x\_test and x\_train that are in the same bin and stored in the array called y\_pred. I did this operation for every x\_test data point and obtained an array with 122 elements. Then, compared this array with y\_test and obtained a rmse value for regressogram which was accurate.