ENGR 421
Homework04: Nonparametric Regression
İrem Demir
69563

1. I read data and stored it. After, discriminating X and Y values I specified train and test indices as given in the HW Description. I split data into train and test sets.
2. I specified some common parameters for the tasks.
3. For the regressogram, I defined several functions.
   1. In order to having bins that we will work on I created a function called bin_calculate. It basically generates two arrays one with starting point of each bin other with end point for corresponding bin at the same position.
   2. Having the bins, I can now calculate regressogram values. I summed corresponding y value of data points within each boundary and took average. Now, we had the regressogram for train set.
   3. For predicting I wrote a different function called regressogram_predict that takes a test set regressogram values based on train set and boundaries. For the points within bins I predicted corresponding trained regressogram value and others are zeros.
   4. I calculated RMSE using RMSE function which I also used for other parts. It was a straightforward calculation
   5. For plotting the regressogram, I defined plot_regressogram function. It first plots train and test data sets puts needed labels, legend etc.
   For regressogram plot within each bin I plotted the value and afterward I added the horizontal lines for better graph.
4. For the running mean smoother I defined several functions
   1. This time I defined points on an interval in our range. For each point in there I took the average of X_train points on the "neighborhood" (within bin_width). I derived the formula using 8.25 on chapter 8.8 of our course book and cheat sheet given. In order to avoid a zero devision I checked them after getting a warning.
   2. For the predict I used X_test values instead of my defined points. For each test point I took the average around a binwidth neighborhood.
   3. The rest of the parts plotting and calculating RMSE were similar with regressogram.
5. For the kernel smoother I defined several functions
   1. As in the running mean smoother I defined points along our axis. I used this to calculate value wtihin a neighborhood however this time the further you get our initial point lesser effect you have with a Gaussian distribution of effects. I did implement the formula using 8.26 on chapter 8.8 and cheat sheet given and Gaussian distribution.
   2. For the predict, I followed a similar logic with running mean smoother
   3. The rest of plotting and calculation RMSE were similar.