**ENGR 421 Introduction to Machine Learning Assignment #4 Report**

Orçun Özdemir

54020

A screenshot of a cell phone

Description automatically generatedA picture containing object

Description automatically generated I have coded my assignment with Python, therefore I have imported pandas,numpy and matplotlib to handle data import and visualization. I have retrieved X and y data by taking the first and second columns of the given data respectively. Then, I have retrieved train data by taking the first 150 rows from both X and y and subsequently, I have retrieved test data by taking the other 120 rows from them as well. Then, I get the minimum and maximum of the X\_train to create left and right borders of my regressogram. Left borders are in between origin point (which is given as 1.5) and maximum value of X\_train(max\_val). Subsequently, right borders are in between origin + bin\_width (which is given as 0.37 too) and max\_val + bin\_width. Then, I created a data\_interval for future calculations of running mean smoother and kernel smoother. After reading book’s section 8.8, I have started to implement the algorithms. Then, I implemented regressogram with a helper function which iterates through X\_train and checks whether the element of X\_train is smaller or equal to right\_border’s related index and left\_border’s related index is smaller than the element of X\_train. If conditions are good to go, then add the y\_train value with the same index to a sum and after the iteration average them. Then, I implemented running mean smoother with a helper function which iterates through X\_train to check conditions are hold in figure 1. Then, I have implemented kernel smoother with a halper function which iterates through X\_train to achieve Gaussian Kernel calculations in figure 2. In figure 2, K stands for Gaussian Kernel which can be implemented in python in figure 3. Then, I have implemented RMSE functions of each three algorithms according to given RMSE function in the pdf. Subsequently, I have created three different p\_head values and filled them with the helper functions’ outputs. Then, I have plotted scatter plots of the train and test data and plotted the algorithms’ lines according their conditions, which basically controls the related point in the scatter plot is in the specific bins of the algorithms p\_heads. After each plot, I have called RMSE function to observe the related algorithm’s RMSE value. At the very end of my assignment I have compared my results with the expected results in the pdf and concluded that my algorithms are in good shape.

Figure 1 Running Mean Smoother Formula

Figure Gaussian Kernel Implementation

Figure 2 Kernel Smoother with Gaussian Kernel