2)

First I read univariate regression data set. Since dataset has header, we will skip it.



3)

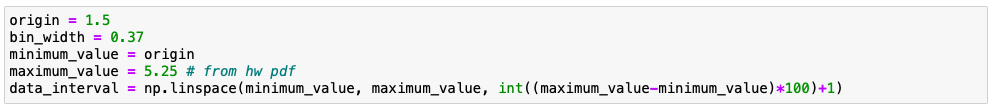
I Divided the data set into two parts by assigning the first 150 data points to the training set and the remaining 122 data points to the test set.

Text

Description automatically generated

4)

I set the bin width parameter to 0.37 and the origin parameter to 1.5. Since maximum value is not given. I put 5.25 by trying values which fits best for graphs.



If we define an origin and a bin width and average the r values in the bin as in the histogram, we get a regressogram. So, to learn regressogram, I used following formulas which is given in the Section 8.8 of the textbook.

Text, letter

Description automatically generated

Graphical user interface, text, application

Description automatically generated

the output is:

Chart, scatter chart

Description automatically generated

5)

Calculate the root mean squared error (RMSE) of your regressogram for test data points. The formula for RMSE is:

For regressogram we need to check borders to check whether test data is within bin\_width. Otherwise numerator of condition is 0. Then we calculate rmse with given formula.

Graphical user interface, text, application

Description automatically generated

Output is:

Text

Description automatically generated

6)

Having discontinuities at bin boundaries is disturbing as is the need to fix an origin. As in the naive estimator, in the running mean smoother, we define a bin symmetric around x and average in there (from book).

Using formula from the book is shows us how to Learn a running mean smoother.

Diagram

Description automatically generated with low confidence

implementation of given formula is follows. As in the formula we should check |u|<1/2.

Text, application

Description automatically generated

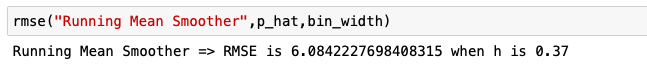
the output is:

Chart, scatter chart

Description automatically generated

7)

rmse is calculated with given formula and I specified implementation in part 5 of this report.



8)

As In the kernel estimator, we can use a kernel giving less weight to further points, and we get the kernel smoother. I implemented given formula from the book.

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

the output is:

Chart, scatter chart

Description automatically generated

9)

RMSE value of Kernel Smoother, is similar to the Running Mean Smoother. And implemented with given formula from the book.

Text

Description automatically generated with medium confidence