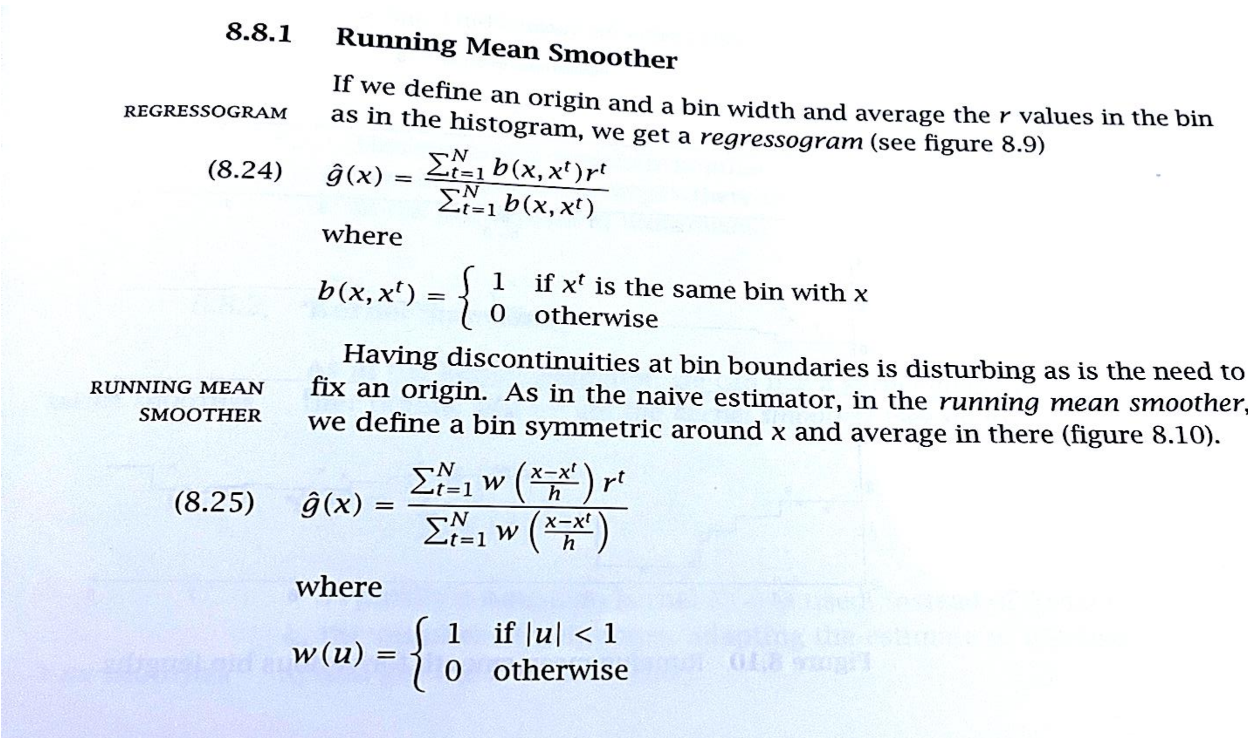
<u>ENGR-421 HW-4 REPORT</u>   (Name-Surname: Barış KAPLAN)

Firstly, I have imported the necessary libraries. The necessary libraries are as follows:

```python
import numpy as np
import math
import matplotlib.pyplot as plt
import pandas as pd
```

By using the genfromtxt function of numpy library, I have read the csv file. After that, I have learned the regressogram by making the value of the bin width parameter (h) 0.37 and by making the value of the origin parameter 1.5. In the Figure 3, you can find the generated regressogram plot with the training & test data points. While learning the regressogram, I have utilized the formula in the part (8.24) of the Figure 1.



8.8.1 Running Mean Smoother

REGRESSOGRAM

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* (see figure 8.9)

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

where

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

RUNNING MEAN SMOOTHER

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 (figure 8.10).

(8.25) $$\hat{g}(x) = \frac{\sum_{t=1}^{N} w\left(\frac{x-x^t}{h}\right) r^t}{\sum_{t=1}^{N} w\left(\frac{x-x^t}{h}\right)}$$

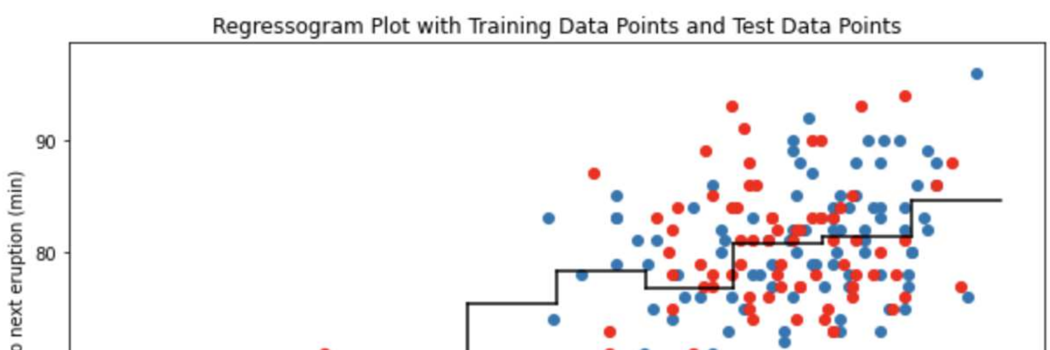
where

$$w(u) = \begin{cases} 1 & \text{if } |u| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Figure 1: The formula used for learning regressogram (8.24) and the formula used for learning the running mean smoother (8.25)

Subsequently, I have calculated the value of RMSE (the root mean squared error) for the regressogram that I have learned. While calculating the value of RMSE for the "Regressogram", the RMSE for the "Running Mean Smoother", and the RMSE for the "Kernel Smoother" ; I have utilized the formula in Figure 2. In Figure 4, you can find the output value of the RMSE for the "Regressogram". In Figure 6, you can find the output value of the RMSE for the "Running Mean Smoother". (See Figure 9 for the output value of RMSE for the "Kernel Smoother")

REGRESSOGRAM:



ating the
es for the
her, and

Figure 3: The regressogram , test data points , and the training data points

```
Regressogram => RMSE is 5.962617204275407 when h is 0.37
```

Figure 4: The output value of RMSE (Root Mean Squared Error) for "Regressogram"

Subsequently, I have learned a "Running Mean Smoother" by making the value of the bin width parameter (h) 0.37. In the Figure 5, you can find the "Running Mean Smoother".

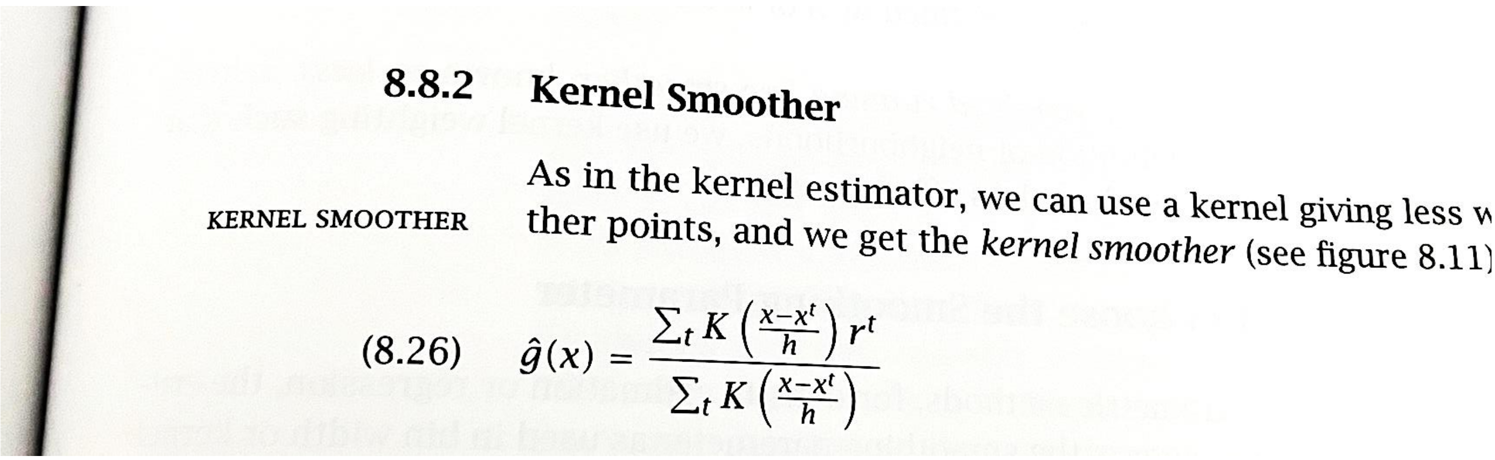
```
RUNNING MEAN SMOOTHER:
```

Figure 5: The Running Mean Smoother, training data points and test data points

```
Mean Smoother => RMSE is 6.08545774692462 when h is 0.37
```

Figure 6: The output value of RMSE (Root Mean Squared Error) for the "Mean Smoother"

Finally, I have learned the "Kernel Smoother" by making the value of the bin width parameter (h) 0.37. You can see the generated "Kernel Smoother" with the traning data points and test data points in Figure 8. While learning the "Kernel Smoother", I have utilized the formula in the part (8.26) of the Figure 7. While learning the "Kernel Smoother", I have also used the formula in Figure 10.



8.8.2 **Kernel Smoother**

KERNEL SMOOTHER

As in the kernel estimator, we can use a kernel giving less w ther points, and we get the *kernel smoother* (see figure 8.11)

$$\hat{g}(x) = \frac{\sum_t K\left(\frac{x-x^t}{h}\right) r^t}{\sum_t K\left(\frac{x-x^t}{h}\right)} \qquad (8.26)$$

Figure 7: The formula used for learning the "Kernel Smoother" (The formula in the part (8.26))

```
KERNEL SMOOTHER:
```

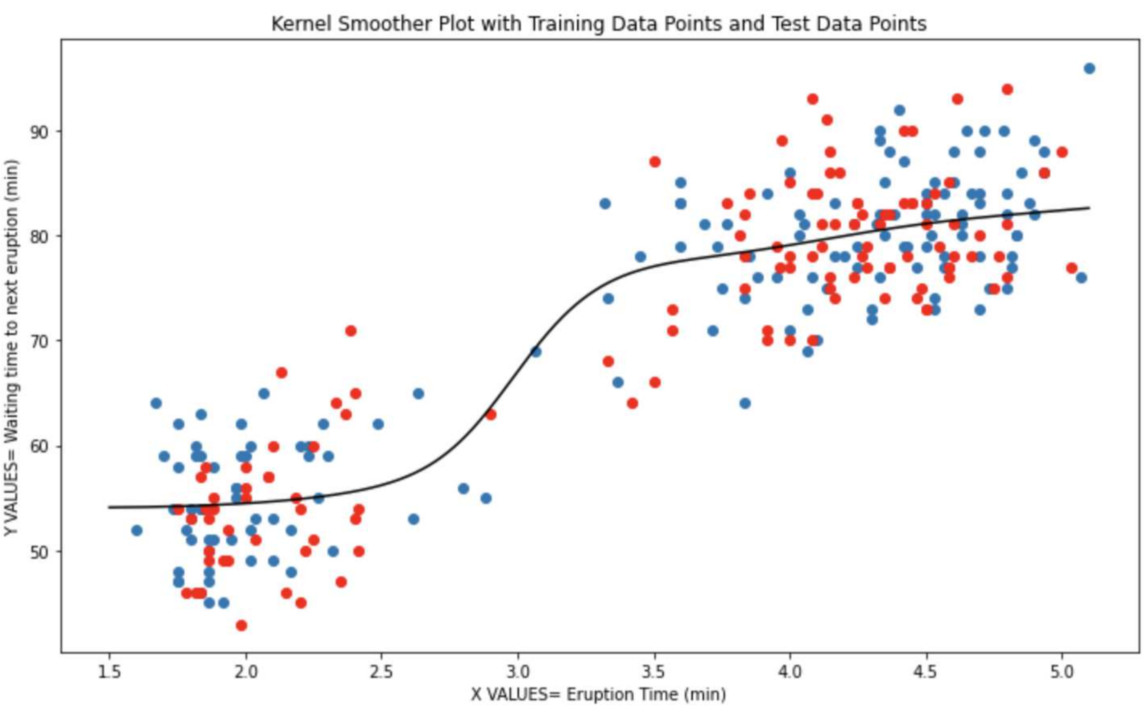


Figure 8: The Kernel Smoother, training data points, and test data points

```
Kernel Smoother => RMSE is 5.874019808179156 when h is 0.37
```

Figure 9: The output value of the RMSE (Root Mean Squared Error) for the "Kernel Smoother"

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

Figure 10: The formula of the Kernel Function

When we compare the algorithms, we can say that the Kernel Smoother is better than the other smoothing models. Because the RMSE value for the Kernel Smoother is less than the other smoothing models (Running Mean Smoother and Regressogram). So, we can say that “Kernel Smoother” fits the data better than the other given smoothing models(Running Mean Smoother and Regressogram) in this homework.