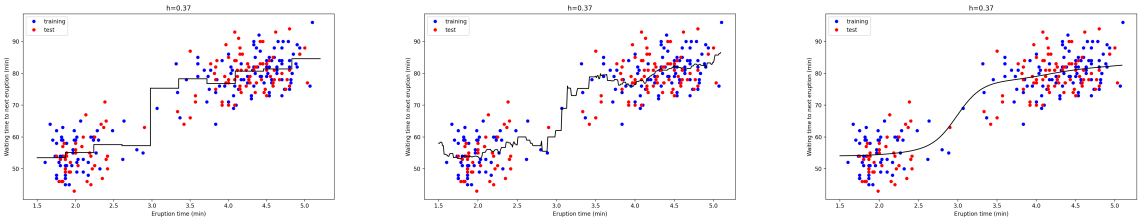
ENGR 421 - HOMEWORK 3
Zeynep Öner
ID: 64912

Nonparametric Regression

In this homework, we implemented a nonparametric regression algorithm using the regressogram, the running mean smoother and the kernel smoother. In the regressogram, we count the data points in the same bin as x and conduct our operation. In the running mean smoother, we use the weight function we discussed in class while covering the naive estimator. In the kernel smoother, we use the kernel function; which is a standard normal distribution.

$$\hat{p}_{reg} = \frac{\sum_{t=1}^{N} b(x,x^t)r^t}{\sum_{t=1}^{N} b(x,x^t)}, \hat{p}_{rms} = \frac{\sum_{t=1}^{N} w(x,x^t)r^t}{\sum_{t=1}^{N} w(x,x^t)}, \hat{p}_{k} = \frac{\sum_{t=1}^{N} K(x,x^t)r^t}{\sum_{t=1}^{N} K(x,x^t)} \tag{0.1}$$

$b(x, x^t)$ is 1 when x^t is in the same bin as x; $w(u)$ is 1 when $|u| < 1/2$, and $K(u) = \frac{1}{\sqrt{2\pi}} exp(\frac{-u^2}{2})$. The textbook says that the weight function returns one when $|u| < 1$, but the naive estimator weight function we covered in class is the one that makes the plot look exactly like the one in the homework description. Therefore, I stuck with our weight function even though the textbook gives a different weight function. The figures obtained from the $\hat{p}$ calculations for regressogram, running mean smoother and kernel smoother respectively are below:



We see that the regressogram has fixed and non-overlapping bins. In the running mean smoother, we have a moving window. In the kernel estimator, we have a smooth moving window.

The RMSE calculations are:

```
/usr/local/bin/python3.9 "/Users/Zeynep/Desktop/spring_2021/ENGR421/Homework Assignments/hw3/engr421_hw3.py"
Regressogram => RMSE is 5.9626 when h is 0.37
Running Mean Smoother => RMSE is 6.0443 when h is 0.37
Kernel Smoother => RMSE is 5.8723 when h is 0.37

Process finished with exit code 0
```

We see that our estimates for the labels and our smoothers yield good results and we have low error.