Berke Ataseven
0054326

# DASC521 HW#4 Report

In this homework, we were asked to implement 3 different non-parametric regression algorithms. These three algorithms are: regressogram, running mean smoother and kernel smoother.

Data points required to draw the relevant curves are generated by the functions "*calculate_regressogram ()*", "*calculate_mean_smoother ()*", "*calculate_mean_smoother ()*". In these functions, I have discretized the x-axis, then calculated the $\hat{g}(x)$ values for every discretized x value. Following equations are used in the implementation of the functions:

**REGRESSOGRAM**

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

where,

$$b(x,\, x^t) = \begin{cases} 1 \; if \;\; x^t \; and \; x \; in \; the \\ \quad same \; bin \\ 0 \; otherwise \end{cases}$$

**RUNNING MEAN SMOOTHER**

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

where,

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

**KERNEL SMOOTHER**

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

where,

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

Notice that in the equation for kernel smoother, $K(u)$ function is the formula for the normal distribution whose mean ($\mu$) is 0 and covariance parameter ($\sigma$) is 1. After the curve functions are generated, I have calculated the root mean squared error (RMSE) for the test data. Calculated RMSE values turned out to be slightly different from what is given in the homework description. I reckon the reason for this is the size of the increment in the discretization of the x-axis. Resulting RMSE values are given below:

```
Regressogram => RMSE is 24.7260 when h is 3.00

Running Mean Smoother => RMSE is 24.2600 when h is 3.00

Kernel Smoother => RMSE is 24.1688 when h is 1.00
```

I have commented my code for the sake of intelligibility.