Data Science Assignment 3

Nike Marie Pulow – Henri Paul Heyden
stu239549 – stu240825

1 Linear Regression

1.1 Regression parameter calculation

We get:

$$\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 8 & 6 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 8 \\ 1 & 6 \\ 1 & 4 \end{bmatrix} \right)^{-1} \cdot \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 8 & 6 & 4 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 2 \\ 2 \\ 4 \end{bmatrix} \right) \\
&= \begin{bmatrix} 4 & 20 \\ 20 & 120 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 16 \\ 60 \end{bmatrix} \\
&= \begin{bmatrix} 1.5 & -0.25 \\ -0.25 & 0.05 \end{bmatrix} \cdot \begin{bmatrix} 16 \\ 60 \end{bmatrix} \\
&= \begin{bmatrix} 9 \\ -1 \end{bmatrix}
\end{aligned}$$

1.2 Error evaluation

For the individual errors ε, we know $y = X\hat{\beta} + \varepsilon$, we conclude $\varepsilon = y - X\hat{\beta}$ and get values:

$$\varepsilon = \begin{bmatrix} 8-9+2 \\ 2-9+8 \\ 2-9+6 \\ 4-9+4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

As we know that the SSE is $\sum_{i=1}^{n} (\beta \cdot x_i - y_i)^2$, we can also write $\sum_{i=1}^{n} {\varepsilon_i}^2$ which is 4.

1.3 Plot