COEN 240 Machine Learning

Assignment # 1

The data points are

$$(2,2),(4,6),(5,4),(7,8),(8,10), \text{ and } (10,12)$$

Find the equation of the regression line (linear regression). Use a high-level language like C, Java, or Python. You should obtain the following quantities. The relevant quantities (answers) are:

$$\begin{aligned}
\overline{x} &= 6, \\
\overline{y} &= 7, \\
S_x^2 &= 7, \\
S_y^2 &= 11.6667, \\
S_{xy} &= 8.6667, \\
r^2 &= 0.9197, \\
r &= 0.9590, \\
E &= 5.6190, \\
a &= -0.42857, \\
b &= 1.2381.
\end{aligned}$$

The equation of the regression line is

$$y = -0.42857 + 1.2381x$$

We have

$$\begin{aligned}
\widehat{y}_1 &= 2.047619, \\
\widehat{y}_2 &= 4.5238095, \\
\widehat{y}_3 &= 5.7619048, \\
\widehat{y}_4 &= 8.2380952, \\
\widehat{y}_5 &= 9.4761905, \\
\widehat{y}_6 &= 11.952381.
\end{aligned}$$

Also since $e_i = -\left(\widehat{y}_i - y_i\right)$, for $1 \leq i \leq 6$, we have

$$\begin{aligned}
e_1 &= -0.047619, \\
e_2 &= 1.476190, \\
e_3 &= -1.761905, \\
e_4 &= -0.238095, \\
e_5 &= 0.523810, \\
e_6 &= 0.047619
\end{aligned}$$

It can be checked that $\sum_{i=1}^{6} e_i = 0$.The solution via EXCEL spread-sheet is also provided.

You should submit the high-level code and the output in a pdf file. Please label the file with your name on it.