Linear Regressions

Justin Hong

November 8, 2020

Problem 28-1

Suppose you are given the following dataset:

$data = [(1, 0.2), (2, 0.25), (3, 0.5)]$

Fit a linear regression model $y = a + bx$ *by hand by*

1. setting up a system of equations,

2. turning the system into a matrix equation,

3. finding the best approximation to the solution of that matrix equation by using the pseudo-inverse,

4. and substituting your solution for the coefficients of the model. Show all of your steps.

No code allowed!

Solution

Step 1.

If we plug in all of the numbers, we will have

$0.2 = a + (1)b$

$0.25 = a + (2)b$

$0.5 = a + (3)b$

Step 2.

We then turn the system of equations into a matrix equation:

$$\begin{bmatrix} 0.2 \\ 0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix}$$

Step 3.

We then use use the transposition of the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ and then the inverse of that (the pseudo-inverse) to solve for $\begin{bmatrix} a \\ b \end{bmatrix}$:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

Now we can use the inverse of $\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$ to find $\begin{bmatrix} a \\ b \end{bmatrix}$:

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 0.95 \\ 2.2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} 0.016667 \\ 0.15 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

Step 4.

$a = 0.016667$ and $b = 0.15$, so the line of best fit is $y = 0.15x + 0.016667$