Exercise 1

Given the data points $(1,2),(2,3),(3,5)$, find the least squares line $y=a+bx$ that best fits the data.

Solution

We aim to find the values of a and b that minimize the sum of squared residuals:

$$S(a,b)=\sum_{i=1}^{n}\left(y_i-(a+bx_i)\right)^2$$

To minimize this function, we take the partial derivatives of $S(a,b)$ with respect to a and b, and set them equal to zero.

Step 1: Compute the partial derivatives of $S(a,b)$

The sum of squared residuals is:

$$S(a,b)=\sum_{i=1}^{n}\left(y_i-(a+bx_i)\right)^2$$

Now, we compute the partial derivatives of $S(a,b)$ with respect to a and b.

$$\frac{\partial S}{\partial a}=-2\sum_{i=1}^{n}\left(y_i-(a+bx_i)\right)$$

$$\frac{\partial S}{\partial b}=-2\sum_{i=1}^{n}x_i\left(y_i-(a+bx_i)\right)$$

To minimize the function, we set both partial derivatives equal to zero:

$$\frac{\partial S}{\partial a}=0 \quad \Rightarrow \quad \sum_{i=1}^{n}\left(y_i-(a+bx_i)\right)=0$$

$$\frac{\partial S}{\partial b}=0 \quad \Rightarrow \quad \sum_{i=1}^{n}x_i\left(y_i-(a+bx_i)\right)=0$$

These equations are the normal equations, which we solve to find a and b.

Step 2: Solve the system of equations

After expanding and simplifying, we obtain:

$$\sum y_i=na+b\sum x_i$$

$$\sum x_iy_i=a\sum x_i+b\sum x_i^2$$

We solve this system of equations to find a and b. This is exactly what we did in the original solution by calculating the necessary sums.

Let us define the data:

x_i	y_i	x_i^2	x_iy_i
1	2	1	2
2	3	4	6
3	5	9	15

Now compute the necessary sums:

$$\sum x_i = 1 + 2 + 3 = 6, \quad \sum y_i = 2 + 3 + 5 = 10$$
$$\sum x_i^2 = 1 + 4 + 9 = 14, \quad \sum x_i y_i = 2 + 6 + 15 = 23$$
$$n = 3$$

Step 3: Compute the slope b

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{3 \cdot 23 - 6 \cdot 10}{3 \cdot 14 - 6^2} = \frac{69 - 60}{42 - 36} = \frac{9}{6} = \frac{3}{2}$$

Step 4: Compute the intercept a

We compute the sample means:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6}{3} = 2, \quad \bar{y} = \frac{\sum y_i}{n} = \frac{10}{3}$$

Then:

$$a = \bar{y} - b\bar{x} = \frac{10}{3} - \frac{3}{2} \cdot 2 = \frac{10}{3} - 3 = \frac{1}{3}$$

Final Result

The least squares regression line is:

$$y = \frac{1}{3} + \frac{3}{2}x$$