Exercise:

The given linear equation is:

$$y = mx + b$$

Where:
- initial $m = -1$
- initial $b = 1$
- Learning rate $\alpha = 0.1$
- Given points:

i	x	y
1	1	3
2	3	6

Loss function: MSE

$$J(m,b) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2 ,$$

n here is equal 2 : $n = 2$

1-) Derive the gradient:

$$J(m,b) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2$$

* Let's calculate $\frac{dJ}{dm}$

$$\frac{dJ}{dm} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2\right]'$$

We know that for f a function:

$$(f^n)' = n f' f^{n-1}$$

So:

$$\frac{dJ}{dm} = \frac{1}{n}\sum_{i=1}^{n} 2\left[\left(y_i - (mx_i + b)\right)'\left(y_i - (mx_i + b)\right)\right]$$

$$\frac{dJ}{dm} = \frac{2}{n} \sum_{i=1}^{n} \left[\left(-x_i \left(y_i - (mx_i + b) \right) \right) \right]$$

cause for $n \in \mathbb{R}$ and x :

$$\frac{d}{dx}(nx) = n \quad \text{and} \quad \frac{dn}{dx} = 0$$

so :

$$\underline{\underline{\frac{dJ}{dm} = -\frac{2}{n} \sum x_i \left(y_i - (mx_i + b) \right)}}$$

* Let's calculate $\frac{dJ}{db}$

$$\frac{dJ}{db} = \left[\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (mx_i + b) \right)^2 \right]'$$

$$\frac{dJ}{db} = \frac{1}{n} \sum 2 \left[(y_i - (mx_i + b))' (y_i - (mx_i + b)) \right]$$

$$\frac{dJ}{db} = \frac{2}{n} \sum -1 \times (y_i - (mx_i + b))$$

Cause:

$$\frac{db}{db} = 1 \quad \text{and} \quad \frac{dm}{db} = 0 \quad \text{with } m \neq b.$$

$$\frac{dJ}{db} = -\frac{2}{n} \sum (y_i - (mx_i + b))$$

* Gradient descent update rules:

$$m_{new} = m_{old} - \alpha \frac{dJ}{dm} \; , \; b_{new} = b_{old} - \alpha \frac{dJ}{db}$$

because $n=2$, the multipliers simplify: $-\frac{2}{n} = -\frac{2}{2} = -1$. So in our numeric steps we can use:

$$\frac{d\mathcal{J}}{dm} = -1 \times \sum x_i \left(y_i - (mx_i + b)\right)$$

$$\frac{d\mathcal{J}}{dm} = -1 \times \sum x_i (y_i - \hat{y}_i)$$

as $\hat{y}_i = mx_i + b$ from the given linear equation.

$$\frac{d\mathcal{J}}{dm} = -1 \times \sum x_i e_i$$

as: $e_i = y_i - \hat{y}_i$

$$\frac{d\mathcal{J}}{dm} = -1 \times \sum x_i e_i$$

Let's proceed with: $\frac{d\mathcal{J}}{db}$

$$\frac{d\mathcal{J}}{db} = -1 \times \sum \left(y_i - (mx_i + b)\right)$$

$$= -1 \times \sum (y_i - \hat{y}_i)$$

$$\frac{d\mathcal{J}}{db} = -1 \sum e_i$$