PART3: Gradient Descent Manual Calculation.

Linear Equation ⇒ $\hat{y} = mx + b$

points ⇒ (1, 3) and (3, 6)

initial values ⇒ ~~$m_0 =$~~ , $b_0 = 1$

$m_0 = -1$

learning rate ⇒ $\alpha = 0.1$

Number of data points ⇒ $n = 2$

Cost function (MSE) using the common convention with factor $1/2n$: $J(m,b) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - (mx + b))^2$

for $n = 2$, the prefactor is $\frac{1}{2n}$

$$\Rightarrow \frac{1}{2(2)} = \frac{1}{4} = 0.25$$

Derive gradients

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

let $e_i = y_i - (mx+b)$

$$J(m,b) = \frac{1}{2n} \sum_{i=1}^{n} e_i^2$$

partial derivative of m:

$$\frac{\partial J}{\partial m} =$$

$$\Rightarrow \frac{1}{2n} \sum_{i=1}^{n} 2e_i \cdot \frac{\partial e_i}{\partial m} \text{ (chain rule)}$$

where $e_i = y_i - (mx+b)$

$$\frac{\partial e_i}{\partial m} = -x$$

so $\frac{\partial J}{\partial m} = \frac{1}{2n} \sum_{i=1}^{n} 2e_i(-x)$

$$= \frac{-1}{n} \sum_{i=1}^{n} e_i x$$

partial derivative of b:

$$\frac{\partial J}{\partial b} = \frac{1}{2n} \sum_{i=1}^{n} 2e_i \frac{\partial e_i}{\partial b}$$

where $\frac{\partial e_i}{\partial b} = -1$

therefore $\frac{\partial J}{\partial b} = \frac{-1}{n} \sum_{i=1}^{n} e_i$

Gradient formulas:

$$\frac{\partial J}{\partial m} = \frac{-1}{n} \sum_{i=1}^{n} x(y_i - (mx+b))$$

$$\frac{\partial J}{\partial b} = \frac{-1}{n} \sum_{i=1}^{n} (y_i - (mx+b))$$

update rules (gradient descent):

$$m_{new} = m_{old} - \alpha \frac{\partial J}{\partial m}$$

$$b_{new} = b_{old} - \alpha \frac{\partial J}{\partial b}$$

Iteration

Given information:

$m_0 = -1$, $b_0 = 1$, Points (1, 3), (3, 6)

let's compute initial predictions:

for $x_1 = 1$, $\hat{y}_1 = m_0 x_1 + b_0$

$$= (-1)(1) + 1$$

$$= 0$$

For $x_2 = 3$, $\hat{y}_2 = m_0 x_2 + b_0$

$$= (-1)(3) + 1$$

$$= -2$$

Error$_1 = y_1 - \hat{y}_1 = 3 - 0 = 3$

Error $_2 = y_2 - \hat{y}_2 = 6 - (-2) = 8$

Compute cost:

$$J_0 = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y})^2$$

$$J_0 = \frac{1}{2(2)} ((\text{loss}_1 + \text{loss}_2)^2)$$

$$= \frac{1}{4} ((3^2) + (8^2))$$

$$= \frac{73}{4} = 18.25$$

Lets' compute gradient:

$$\frac{\partial J}{\partial m} = \frac{-1}{n} \sum_{i=1}^{n} e_i x \quad , n = 2$$

and

$$\frac{\partial J}{\partial b} = \frac{-1}{n} \sum_{i=1}^{n} e_i$$

For $\frac{\partial J}{\partial m} = \frac{-1}{n} \sum_{i=1}^{n} e_i x$

$$= \frac{-1}{n} (\text{error}_1 x_1 + \text{error}_2 x_2)$$

$$= \frac{-1}{2} (3(1) + 8(3))$$

$$= \frac{-1}{2} (3 + 24)$$

$$= -\frac{27}{2} = -13.5$$

For $\frac{\partial J}{\partial b} = \frac{-1}{n} \sum_{i=1}^{n} e_i$

$$= \frac{-1}{n} (\text{error}_1 + \text{error}_2)$$

$$= \frac{-1}{2} (3 + 8)$$

$$= \frac{-11}{2} = -5.5$$

Next, updating the parameters:

$$m_1 = m_0 - \alpha . \frac{\partial J}{\partial m}$$

$$= -1 - (0.1)(-13.5)$$

$$= 0.35$$

$$b_1 = b_0 - \alpha . \frac{\partial J}{\partial b}$$

$$= 1 - (0.1)(-5.5)$$

$$= 1.55$$

After the iteration, this is our new values:

$m_{new} = 0.35$

$b_{new} = 1.55$

Given $m_0 = -1$ and $b_0 = 1$ and also $m_{new} = 0.35$ and $b_{new} = 1.55$ by calculation our new values of m and b are increasingly where by m we are increasingly by the change of $+1.35$ and b is increasingly by the change of $+0.55$. so,

Yes, the values of m and b are moving towards reducing the error because of the negative gradient $(-13.5, -5.5)$ combined with the gradient descent are the one pushing or moves the parameters in the correct direction to fit the line of the points $(1,3)$ and $(3,6)$.