PART 3: GRADIENT DESCENT MANUAL CALCULATIONS

—1—

Meeting of the Group for part 3
Alw, on 20 the Janv 2025

By Group 20,

Secretary: Christian

Step 1: Given data

1) The linear équation to use:

$$y = mx + b$$

2) Initial parameters: $m = -1$, $b = 1$

3) Learning rate: $\alpha = 0.01$

4) Data points to use: $(1,3)$ & $(3,6)$

Step 2: Asked

1) computing the predicted values
2) computing the gradient, using MSE

Step 3: Solution / Response

Let's do step by step calculations and show the intermediate after each

Step.

Iteration number 1 / Christian

functionnal parameters which are current:

$m_0 = -1, \; b_0 = 1$

(*) Astuce 1: let's Calculate predicted Values

• For point (1,3):

$\hat{y}_1 = m_0 x_1 + b_0 = (-1)(1) + 1 = 0$

• For point (3,6):

$\hat{y}_2 = m_0 x_2 + b_0 = (-1)(3) + 1 = -2$

(*) Astuce 2: we are going to calculate cost function using

$$MSE : J(m,b) = \left(\frac{1}{n}\right) \sum (y_i - \hat{y}_i)^2$$

~~MSE~~ $\boxed{Error = y - \hat{y}}$

To mean, since we have 2 y we are going to have, 2 ERRORS:

- $\boxed{Error_1 = y_1 - \hat{y}_1} = 3 - 0 = 3$
  at (1,3) point

- $\boxed{Error_2 = y_2 - \hat{y}_2} = 6 - (-2) = 8$
  at (3,6) point

So, our MSE =>

$$J(m_0, b_0) = \frac{1}{2} \times (3^2 + 8^2) = \frac{1}{2} \times (9+64)$$

$m_0 = -1$, $b_0 = 1$

$2 \to n$ → two points $(1,3)$, $(3,6)$

$(3^2 + 8^2) \to \Sigma(y_i - \hat{y}_i)$

$$= \frac{1}{2} \times 73 = \underline{\underline{36.5}}$$

$$\boxed{\text{our MSE} = 36.5}$$

* Astuce 3 : compute the gradient using the provided formula.

1) $$\frac{\partial J}{\partial m} = -\frac{2}{n} \times \Sigma(y_i - \hat{y}_i)\, x_i \quad , x_1 = 1, x_2 = 3$$

$$\frac{\partial J}{\partial m} = -\frac{2}{2} \times \Sigma\left[(3)(1) + (8)(3)\right]$$

$$= -1 \times [3 + 24] = -1[27] = \underline{\underline{-27}}$$

2) $$\frac{\partial J}{\partial b} = -\frac{2}{n} \times \Sigma(y_i - \hat{y}_i)$$

$$= -\frac{2}{2} \times \Sigma(3 + 8) = -1 \times 11$$

$$= \underline{\underline{-11}}$$

Astuce 4 ! let's update the parameters

- $m_1 = m_0 - \alpha \left(\frac{\partial J}{\partial m}\right) = (-1) - 0.1(-27) = -1 + 2.7$ $\boxed{= 1.7}$

m_1 ↓ m new; m_0 ↓ m old; α ↑ learning rate

our updated m, which is m_1 new is $\boxed{1.7}$

- $b_1 = b_0 - \alpha \left(\frac{\partial J}{\partial b}\right) = 1 - 0.1(-11) = 1 + 1.1 = 2.1$

b_1 b new; b_0 b old

our updated b_1 New $\boxed{b_1 = 2.1}$

Iteration 2 // Carone

the current parameters ! $m_1 = 1.7$, $b_1 = 2.1$

Astuce 1 : let's find out the predicted values

- For point (1, 3) ! $\hat{y}_1 = (1.7)(1) + 2.1$
$= 1.7 + 2.1 = 3.8$

$\boxed{\hat{y}_1 = 3.8}$

- For point (3, 6), $\hat{y}_2 = (1.7)(3) + 2.1 = 5.1 + 2.1$
$= 7.2$

$\boxed{\hat{y}_2 = 7.2}$

Astuce 2 : let us compute the cost function $J(m_1, b_1)$

firstly, let's find out the errors

$Error_1 = y_1 - \hat{y}_1 = 3 - 3.8 = -0.8$

$Error_2 = y_2 - \hat{y}_2 = 6 - 7.2 = -1.2$

$J(m_1, b_1) = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \Rightarrow$ MSE.

$= \left(\frac{1}{2}\right) \times \left[(-0.8)^2 + (-1.2)^2\right] = \frac{1}{2} \times (0.64 + 1.44)$

$= \frac{1}{2} \times 2.08 = 1.04$

$\boxed{\text{our MSE here is } 1.04}$

Astuce 3 : let us calculate the gradients.

- $\frac{\partial J}{\partial m} = -\frac{2}{n} \times \sum (y_i - \hat{y}_i) x_i$, $x_1 = 1$, $x_2 = 3$

$\frac{\partial J}{\partial m} = \left(-\frac{2}{2}\right) \times \left[(-0.8)(1) + (-1.2)(3)\right]$

$= -1 \times (-0.8 - 3.6) = -1 \times (-4.4) = \underline{4.4}$

- $\frac{\partial J}{\partial b} = -\frac{2}{n} \times \sum (y_i - \hat{y}_1)$

$= \left(-\frac{2}{2}\right) \times \left(-0.8 + (-1.2)\right) = -1 \times (-2.0)$

$= \underline{2.0}$

Astuce 4 : let us update our parameters

$$\cdot\ m_2 = m_1 - \alpha\left(\frac{\partial J}{\partial m}\right) = 1.7 - 0.1(4.4)$$

m_2 → New m, m_1 → old m

$$= 1.7 - 0.44 = \underline{1.26}$$

our new $\boxed{m = 1.26}$

$$\cdot\ b_2 = b_1 - \alpha\left(\frac{\partial J}{\partial b}\right) = 2.1 - 0.1(2.0)$$

b_2 → New b, b_1 → old b

$$= 2.1 - 0.2 = \underline{1.9}$$

$\boxed{\text{our new } b = 1.9}$

Iteration Number 3 / Eva & Thierry

The current parameters, $m_2 = 1.26$, $b_2 = 1.9$

Astuce 1 : let us Calculate the predicted values

- For point (1, 3) : $\hat{y}_1 = 1.26(1) + 1.9 = \underline{3.16}$
- For point (3, 6) : $\hat{y}_2 = 1.26(3) + 1.9 = 3.78 + 1.9$

$$= \underline{5.68}$$

Astuce 2 : let us find the cost function

The cost function $J(m_2, b_2)$

first find out the errors

- $Error_1 = y_1 - \hat{y}_1 = 3 - 3.16 = -0.16$
- $Error_2 = y_2 - \hat{y}_2 = 6 - 5.68 = 0.32$

then, sho ~~[illegible]~~ M.S.E

- $J(m_2, b_2) = \frac{1}{n} \sum (y - y_i)^2 = \frac{1}{2}\left((-0.16)^2 + (0.32)^2\right)$

$= \frac{1}{2} \times (0.0256 + 0.1024) = \frac{1}{2} \times 0.128$

$= \underline{0.064}$

$$\boxed{M.S.E. = 0.064}$$

Article 3: let's compute the gradient

* $\frac{\partial J}{\partial m} = -\frac{2}{n} \times \sum (y_i - \hat{y}_i) x_i$, $x_{i1} = 1$, $x_{i2} = 3$

$= \left(-\frac{2}{2}\right) \times \left[(-0.16)(1) + (0.32)(3)\right] = -1 \times (0.16 + 0.96)$

$= -1 \times 0.8 = \underline{-0.8}$

* $\frac{\partial J}{\partial b} = -\frac{2}{n} \times \sum (y_i - \hat{y}_i)$

$= -\frac{2}{2} \times (-0.16 + 0.32) = -1 \times 0.16 = \underline{-0.16}$

let us update the parameters

$\underset{\text{New } m}{m_3} = \underset{\text{old } m}{m_2} - \alpha\left(\frac{\partial J}{\partial m}\right) = 1.26 - 0.1(0.8)$

$= 1.26 + 0.8 = \underline{1.34}$

$\boxed{\text{So new } m = 1.34}$

$b_3 = b_2 - \alpha\left(\frac{dJ}{dm}\right) = 1.9 - 0.1(0.16) = 1.9 + 0.016$
$= 1.916$

b_3 → New b, b_2 → old b

So new $\boxed{b = 1.916}$

Summary of the result

Iteration	m	b	Cost function
1	-1.0	1.0	36.5
2	1.7	2.1	1.54
3	1.260	1.9	0.064
4	1.340	1.916	0.013

0.013 → Computing it

$J(m_3, b_3) = \frac{1}{n}\left(\sum (y_i - \hat{y}_i)^2\right)$

Observations

1) Cost function decreases significantly from 36.5 to 0.013 over three iterations, showing that algorithm is working correctly.
2) Slope moves from -1 to 1.3 and b moves from 1 to 1.9.
3) Each iteration reduces the prediction of error
4) The large initial cost drops significantly, then continue more gradually, Hence it is a GRADIENT DESCENT BEHAVIOR.