PART B: GRADIENT DESCENT MANUAL CALCULATION

Given information

model: $y = mx + b$

initial $m = -1$

initial $b = 1$

Learning rate $\alpha = 0.1$

Data points: $(1,3)$ and $(3,6)$

COST FUNCTION (mean square error)

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

when $n = 2$

Step 1: computing prediction with m and b

$$\hat{y}_1 = mx_1 + b = (-1)(1) + 1 = 0$$

$$\hat{y}_2 = mx_2 + b = (-1)(3) + 1 = -2$$

Step 2: computing error

$$e_1 = y_1 - \hat{y}_1$$

$$e_2 = y_2 - \hat{y}_2$$

$$e_1 = 3 - 0 = 3$$

$$e_2 = 6 - (-2) = 8$$

Step 3: computing partial derivative of J with respect to m

$$J = \frac{1}{2}\left((3 - m - b)^2 + (6 - 3m - b)^2\right)$$

Differentiating the first term

$$\frac{\partial}{\partial m}(3 - m - b)^2$$

Let $u = 3 - m - b$, then $\frac{du}{dm} = -1$

chain rule:

$$\frac{d}{dm}(u^2) = 2u \cdot \frac{du}{dm} =$$

$$2(3 - m - b)(-1)$$

$$\therefore \frac{\partial}{\partial m}(3 - m - b)^2$$

$$= -2(3 - m - b)$$

Differentiating 2nd term

$$\frac{\partial}{\partial m}(6 - 3m - b)^2$$

Let $v = 6 - 3m - b$, then $\frac{dv}{dm} = -3$

chain rule:

$$\frac{d}{dm}(v^2) = 2v \cdot \frac{dv}{dm} = 2(6 - 3m - b)(-3)$$

$$\frac{\partial}{\partial m}(6 - 3m - b)^2 = -6(6 - 3m - b)$$

Sum it up and multiply by $\frac{1}{2}$

$$\frac{\partial J}{\partial m} = \frac{1}{2}\left[-2(3 - m - b) + (-6)(6 - 3m - 3)\right)$$

$$\frac{1}{2}\left(-2(3 - m - b) - 6(6 - 3m - b)\right)$$

Step 4: Substitute

First term

$$3 - m - b = 3 - (-1) - 1 =$$

$$3 + 1 - 1 = 3$$

2nd term

$6-3m-b = 6-3(-1)-1 = 6+3-1$
$= 8$

$\therefore \frac{\partial J}{\partial m} = \frac{1}{2}\left[-2(3)-6(8)\right]$

$= \frac{1}{2}\left[-6-48\right] = \frac{1}{2}(-54) = -27$

Step 5: computing partial derivative of J with respect to b

let $u = 3-m-b$, $\frac{dy}{db} = -1$

chain rule:

$2u \cdot \frac{dy}{db} = 2(3-m-b)(-1) =$
$-2(3-m-b)$.

let $v = 6-3m-b$, $\frac{dy}{db} = -1$

chain rule:

$2v \cdot \frac{dv}{db} = 2(6-3m-b)(-1) =$
$-2(6-3m-b)$

Sum and multiply by ½

$\frac{\partial J}{\partial b} = \frac{1}{2}\left[-2(3-m-b)-2(6-3m-b)\right]$

$= \frac{1}{2}\left[-2(3-m-b+6-3m-b)\right]$

$= -1\left[9-4m-2b\right]$

Substituting m and b

where $m=-1$ and $b=1$

$9-4(-1)-2(1) = 9+4-2 = 11$

$\therefore \frac{\partial J}{\partial b} = -11$

Step 6 updating m and b

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

$-1 - 0.1 \cdot (-27)$

$= -1 + 2.7 = 1.7$

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

$1 - 0.1 \cdot (-11)$

$= 1 + 1.1 = 2.1$

Therefore the values of m and b are:

$m = 1.7$

$b = 2.1$

Second Iteration:

$m = 1.7$

$b = 2.1$

Computation of Prediction:

Points - Provided: $(1,3)$ and $(3,6)$

For $x_1 = 1$

$\hat{y}_1 = mx_1 + b = (1.7)1 + 2.1 = 3.8$

for $x_2 = 3$

$\hat{y}_2 = mx_2 + b = (1.7)3 + 2.1 = 7.2$

Computation of Errors

$e_1 = \hat{y}_1 - y_1 = 3.8 - 3 = 0.8$

$e_2 = \hat{y}_2 - y_2 = 7.2 - 6 = 1.2$

Computation of Gradients.

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

- $n = 2$ (as we have 2 points)
- Due to Constant Multiple rule - we will exempt ourselve from the $\frac{1}{n}$ part of the formula.

$$\frac{d}{dx}[c \cdot f(x)] = c \cdot f'(x)$$

Second term Iteration

$J(m,b) = \frac{1}{2}\left[(3-m-b)^2 + (6-3m-b)^2\right]$

$\frac{d}{dm}(3-m-b)^2$

let $t = 3-m-b$; $\frac{dt}{dm} = 0-1-0 = -1$

Chain rule :

$\frac{d}{dm}(t^2) = 2t \cdot \frac{dt}{dm} = 2(3-m-b)(-1) = -2(3-m-b)$

Differentiating Term2 :

$\frac{d}{dm}(6-3m-b)^2$; let $r = 6-3m-b$

$\frac{dr}{dm} = 0-3-0 = -3$.

$\frac{d(r^2)}{dm} = 2r \cdot \frac{dr}{dm} = 2(6-3m-b)(-3)$

$(=)\ -6(6-3m-b)$

$\frac{dJ}{dm} = \frac{1}{2}\left[-2(3-m-b) + (-6)(6-3m-b)\right]$

Substitution

$\frac{dJ}{dm} = \frac{1}{2}\left[-2(3-1.7-2.1) + (-6)[(6-3(1.7)-2.1)\right]$

$= 4.4$

Partial derivative with respect to b:

$$\frac{d}{db}(3-m-b)^2, \text{ let } s = 3-m-b$$

$$\frac{ds}{db} = 0-0-1 = -1$$

Chain rule

$$\frac{d(s)^2}{dm} = 2s \cdot \frac{ds}{dm} = 2(3-m-b)(-1) = -2(3-m-b)$$

Differentiating Term 2:

$$\frac{d}{db}(6-3m-b)^2, \text{ let } z = 6-3m-b$$

$$\frac{dz}{db} = 0-0-1 = -1$$

Chain rule:

$$\frac{d(z^2)}{db} = 2z \cdot \frac{dz}{db} = 2(6-3m-b)(-1)$$

$$\frac{dJ}{db} = \frac{1}{2}\left[-2(3-m-b) + (-2)(6-3m-b)\right]$$

$$= \frac{1}{2}\left[-2(3-1{\cdot}7-2{\cdot}1) + (-2)(6-3(1{\cdot}7)-2{\cdot}1)\right]$$

$$= 2.$$

Updating m and b

$$m_{new} = m_{old} - \alpha \frac{dJ}{dm}$$

$$b_{new} = b_{old} - \alpha \frac{dJ}{db}$$

$$m_{new} = 1.7 - (0.1)\, 4.4 = 1.26$$

$$b_{new} = 2.1 - (0.1)\, 2 = 1.9$$

$m = 1.26 \qquad b = 1.9$

Third Iteration:

Given:

Model $\Rightarrow y = mx + b$

Initial values $\begin{cases} m = 1.26 \\ b = 1.9 \end{cases}$

$\alpha = 0.1$

Provided Points: $(1,3)$ and $(3,6)$

For $x_1 = 1$:

$$\hat{y}_1 = mx_1 + b = (1.26)(1) + 1.9 = 3.16$$

For $x_2 = 3$:

$$\hat{y}_2 = mx_2 + b = (1.26)(3) + 1.9 = 5.68$$

Error computation:

$$e_1 = \hat{y}_1 - y_1 = 3.16 - 3 = 0.16$$

$$e_2 = \hat{y}_2 - y_2 = 5.68 - 6 = -0.32$$

Gradient computation:

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

where: $n = 2$ (no. of points)

- Now, let us compute the partial derivative of J with respect to m:

$$J = \frac{1}{2}\left[(3-m-b)^2 + (6-3m-b)^2\right]$$

From differentiating the 1^{st} term, we get:

$$\frac{d}{dm}(3-m-b)^2$$

- Let t be $(3-m-b)$

$$\therefore \frac{dt}{dm} = -1$$

- Using the chain rule:

$$\frac{d(t^2)}{dm} = 2t \cdot \frac{dt}{dm} = 2(3-m-b)(-1)$$

$$= -2(3-m-b)$$

From differentiating the 2^{nd} term, we get:

$$\frac{d(6-3m-b)^2}{dm}$$

- Let $u = 6-3m-b$; $\frac{du}{dm} = -3$

- Using the chain rule.

$$\frac{d(u^2)}{dm} = 2u \cdot \frac{du}{dm} = 2(6-3m-b)(-3)$$

$$= -6(6-3m-b)$$

Now;

$$\frac{dJ}{dm} = \frac{1}{2}\left[-2(3-m-b) + (-6)(6-3m-b)\right]$$

$$= \frac{1}{2}\left[-2(3-1.26-1.9) - 6(6-3(1.26)-1.9)\right]$$

$$= \frac{1}{2}\left[-2(-0.16) - 6(0.32)\right]$$

$$= \boxed{-0.8}$$

- Let us now compute the partial derivative of J with respect to b:

- Let $u = 3-m-b$; $\frac{du}{db} = -1$

Using the chain rule:

$$2u \cdot \frac{du}{db} = 2(3-m-b)(-1)$$

$$= -2(3-m-b)$$

- Let $t = 6-3m-b$; $\frac{dt}{db} = -1$

$$2t \cdot \frac{dt}{db} = 2(6-3m-b)(-1)$$

$$= -2(6-3m-b)$$

Now;

$$\frac{dJ}{db} = \frac{1}{2}\left[-2(3-m-b) + (-2)(6-3m-b)\right]$$

$$= \frac{1}{2}\left[-2(3-1.26-1.9) - 2(6-3(1.26)-1.9)\right]$$

$$= \frac{1}{2}\left[-2(-0.16) - 2(0.32)\right] = \boxed{-0.16}$$

Let us now update the values of m and b.

$$M_{new} = M_{old} - \alpha \frac{dJ}{dm}$$

$$b_{new} = b_{old} - \alpha \frac{dJ}{db}$$

$M_{new} = 1.26 - (0.1)(-0.8)$

$= \underline{\underline{1.34}}$

$b_{new} = 1.9 - (0.1)(-0.16)$

$= \underline{\underline{1.916}}$

$\therefore$ $\boxed{\begin{array}{l} m = 1.34 \\ b = 1.916 \end{array}}$

PAGE
DATE

Given values from 3rd iteration

$m = 1.34$ Learning rate $= 0.1$
$b = 1.916$ Data points $= (1,3)(3,6)$

Step 1: computing predictions

$\hat{y}_1 = mx_1 + b = (1.34)(1) + 1.916$
$= 3.256$

$\hat{y}_2 = mx_2 + b = (1.34)(3) + 1.916$
$= 5.936$

Step 2: computing errors

$e_1 = y_1 - \hat{y}_1 = 3 - 3.256 = -0.256$
$e_2 = y_1 - \hat{y}_2 = 6 - 5.936 = 0.064$

Step 3: partial derivative of J with respect to m

using Formula:

$$\frac{\partial J}{\partial m} = \frac{1}{2}\left[-2(3-m-b) - 6(6-3m-b)\right]$$

Substitute:
$m = 1.34$, $b = 1.916$

First term:
$3 - m - b = 3 - 1.34 - 1.916$
$= 0.256$

Second term

$6 - 3m - b = 6 - 3(1.34) - 1.916$
$= 6 - 4.02 - 1.916 =$
0.064

so: $\frac{\partial J}{\partial m} = \frac{1}{2}\left[-2(-0.256) - 6(0.064)\right]$

$= \frac{1}{2}(0.512 - 0.384]$.

$= \frac{1}{2}(0.128) = 0.064$

Step 4: partial derivative of J with respect to b

$$\frac{\partial J}{\partial b} = -1[9 - 4m - 2b]$$

Substituting m and b
$m = 1.34$ $b = 1.916$

$9 - 4(1.34) - 2(1.916) =$

$9 - 5.36 - 3.832. = -0.192$

so! $\frac{\partial J}{\partial b} = -1(-0.192) = 0.192$

HD

PAGE
DATE

Step 5: updating m and b

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

$= 1.34 - 0.1 \times (0.064)$

$= 1.34 - 0.0064$

$= 1.3336$

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

$= 1.916 - 0.1 \times (0.192)$

$= 1.916 - 0.0192 = 1.8968$

Result after 4th Iteration

$m = 1.3336$

$b = 1.8968$

SUMMARY OF ALL ITERATIONS

1st Iteration: $m = 1.7$ $b = 2.1$

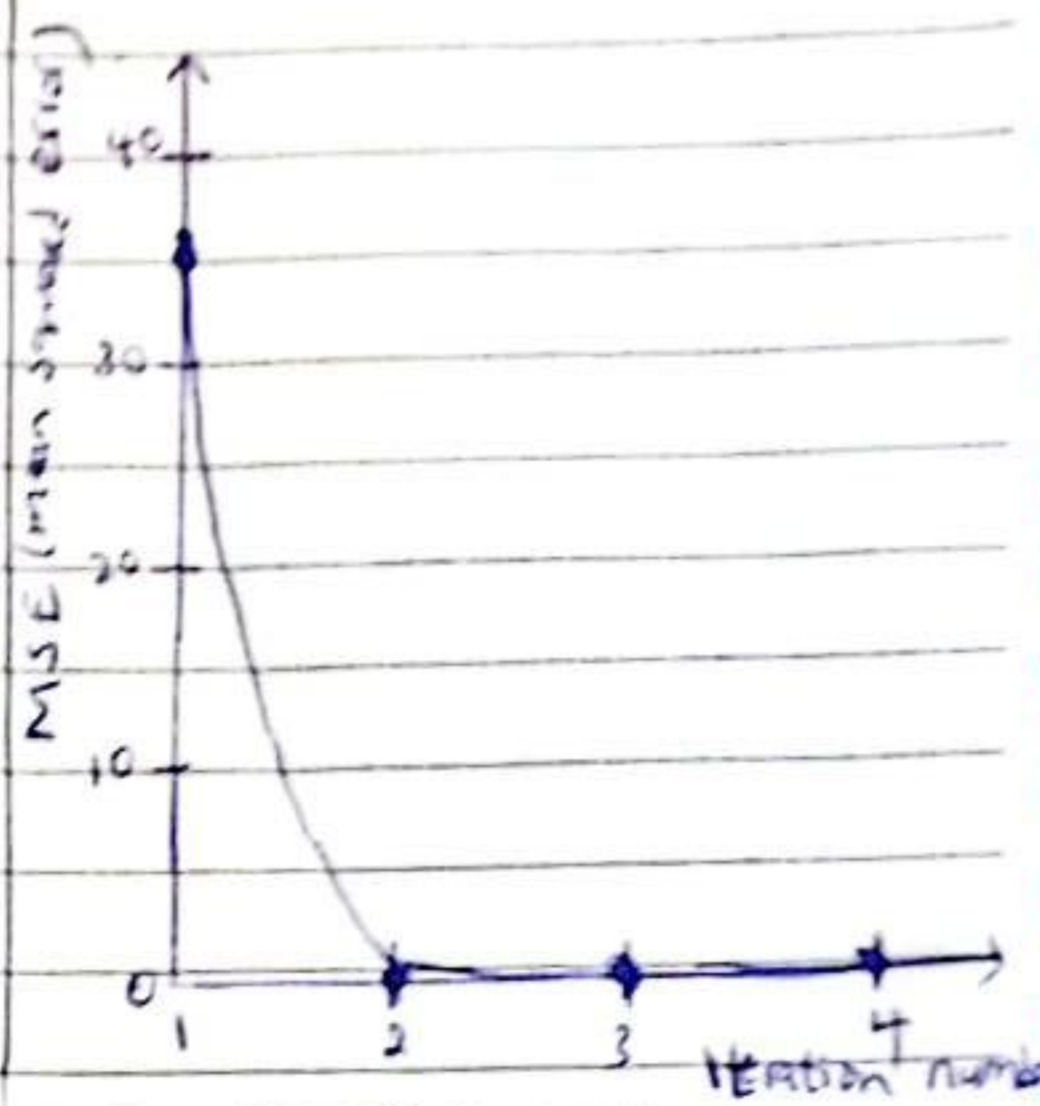
2nd Iteration: $m = 1.26$ $b = 1.9$

3rd Iteration: $m = 1.34$ $b = 1.916$

4th Iteration: $m = 1.3336$ $b = 1.8968$

The parameters are converging towards the true optimal values ($m = 1.5$, $b = 1.5$) and the error is decreasing with each Iteration.

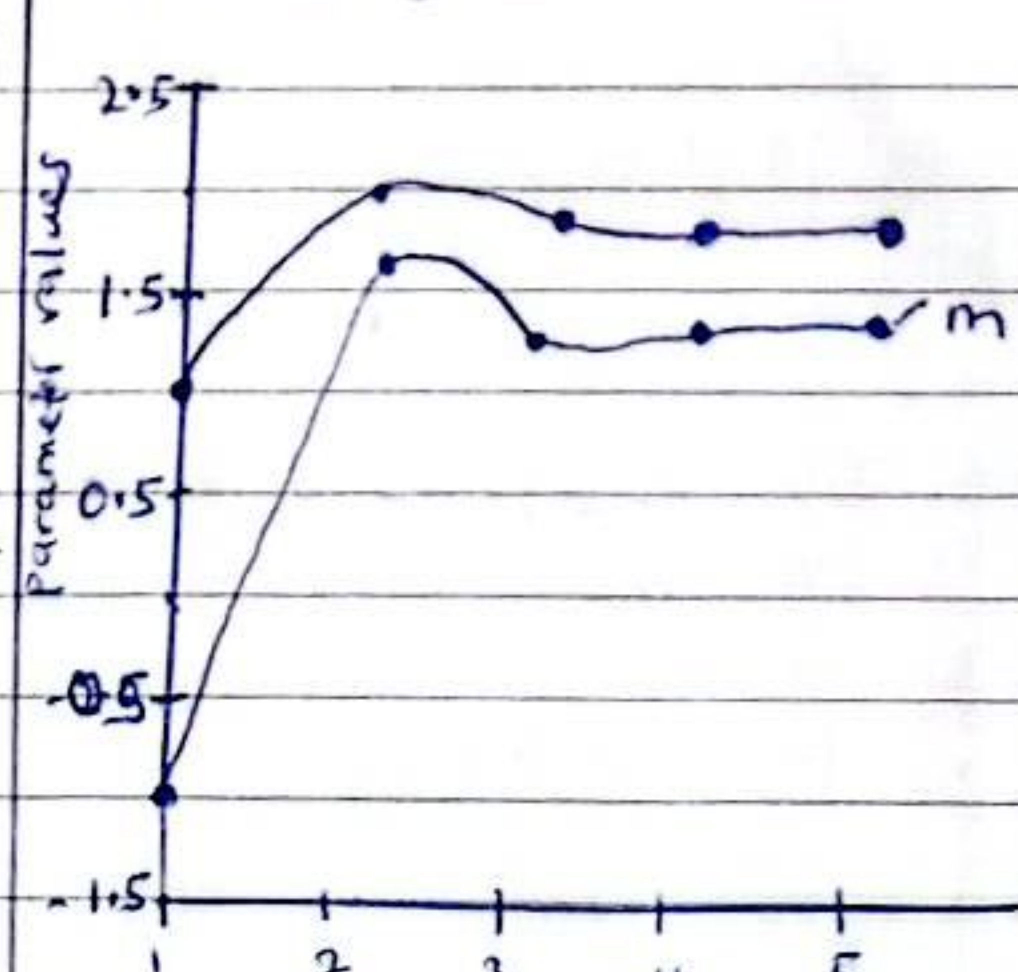
MSE VS ITERATION GRAPH



KEY OBSERVATION

MSE decreased from 36.5 to 0.034816 - a reduction of over 99%

Parameter evolution graph (m and b)



HD