PART 3: Gradient Descent Manual Calculations

$y = mx + b$

Given that

Initial $m = -1$

Initial $b = 1$

Learning rate $= 0.1$

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

Using Mean Squared Error (MSE)

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

to see how wrong the line is. We should reduce the value of J.

We will also use the following Gradient formulas to be able change m and b

$$\frac{\partial J}{\partial m} = -2/n \sum (y_i - \hat{y}_i)\, x_i$$

$$\frac{\partial J}{\partial b} = -2/n \sum (y_i - \hat{y})$$, n = no data points

Iteration 1

$y = mx + b$ to predict the values

For $(1, 3) \Rightarrow \hat{y}_1 = (-1)(1) + 1 = 0$

For $(3, 6) \Rightarrow \hat{y}_2 = (-1)(3) + 1 = -2$.

Computing errors :

Error $= y - \hat{y}$

for $y = 0$, Error $= 3 - 0 = 3$ for point $(1, 3)$

for $y = -2$, Error $= 6 - (-2) = 8$ for point $(3, 6)$

$$\frac{\partial J}{\partial m} = -2/n \sum (y_i - \hat{y}_i) x_i$$

$$\Rightarrow -2/2 \left((3 \times 1) + (8 \times 3) \right)$$

$$\Rightarrow -1 (3 + 24)$$

$$= -27.$$

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

$$\Rightarrow -2/n \; (3+8) = -11$$

Updated m and b using

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

$$\Rightarrow -1 - 0.1 (-27) = 1.7$$

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

$$\Rightarrow 1 - 0.1 (-11) = 2.1$$

Iteration 2

Predictions

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

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

Errors

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

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

Gradients

$$\frac{dJ}{dm} = -2/2\left((0.8)(1) + (-1.2 \times 3)\right) = 4.4$$

$$\frac{dJ}{db} = -2/2\,(-0.8 + 1.2) = 2$$

updated m and b

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

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

Date: / /

Iteration 3.

Predict values

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

$y_2 = 1.26(3) + 1.9 = 5.68$

Errors are

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

$y - y_2 = 6 - 5.68 = 0.32.$

Gradients Formula

$\frac{dJ}{dm} = -\frac{2}{n} \sum (y_i - \hat{y}_i)(x_i) = -\frac{2}{2} \sum (-0.16 \times 1 + 3 \times 0.32) = -0.8$

$\frac{dJ}{db} = -\frac{2}{n} \sum (y_i - \hat{y}_i) = -\frac{2}{2}(-0.16 + 0.32) = -0.16.$

Replace the m and b in the formula

$y = mx + b.$

$m_{new} = 1.26 - 0.1(0.8) = 1.34$

$b_{new} = 1.9 + 0.1(-0.16) = 1.96.$

V

Iteration 4

Predict values

$\hat{y}_1 = 1.34(1) + 1.91 = 3.25$

$\hat{y}_2 = 1.34(3) + 1.91 = 5.93$

Errors

$y - \hat{y}_1 = 3 - 3.25 = -0.25$

$y - \hat{y}_2 = 6 - 5.93 = 0.07$

Gradients

$$\frac{\Delta J}{\Delta m} = \frac{-2}{2}\left((-0.25 \times 1) - (0.07 \times 3)\right) = 0.46$$

$$\frac{\Delta J}{\Delta b} = -2/2\,(-0.25 - 0.07) = 0.32$$

Updated m and b

$m_{new} = 1.34 - 0.1(0.46) = \boxed{1.29}$

$b_{new} = 1.91 - 0.1(0.32) = \boxed{1.89}$

Over the four iterations, the values of m and b steadily move toward reducing the error. Initially, both parameters adjust significantly to correct large prediction errors, but with each iteration, the changes become smaller and more refined. This gradual adjustment indicates that the model is converging toward the optimal line of best fit, with the error decreasing at every step. The shrinking gradients and stabilizing values confirm that m and b are effectively minimizing the mean squared error.