Part 3: Gradient Descent Manual Calculation

Initial values:

$m = -1, b = 1, \alpha = 0.1$

x	1	3
y	3	6

Iteration 1:

Find predicted values $\hat{y}$ for each data point.

$\hat{y}_1 = mx + b$

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

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

Expected y	Predicted y
3	0
6	-2

Find the new m and b

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

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

$$\frac{\Delta J}{\Delta m} = -\cancel{\frac{2}{2}}\left[(3-0)(1) + (6-(-2))(3)\right]$$

$$= -1\,(3+24)$$

$$= -1(27) = -27$$

$$m_{new} = -1 - 0.1(-27)$$

$$m_{new} = 1.7$$

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

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

$$= -\cancel{\frac{2}{2}}\left[(3-0) + (6-(-2))\right]$$

$$= -1(11) = -11$$

$$b_{new} = 1 - 0.1(-11) = 2.1$$

2nd Iteration

From the first Iteration:

$$m = 1.7$$

$$b = 2.1$$

Values of $\hat{y}$ using values of m and b from 1st Iteration.

For (1,3); $\hat{y}_1 = (1.7) \times 1 + 2.1 \Rightarrow 1.7 + 2.1 = 3.8$

For (3,6); $\hat{y}_2 = (1.7) \times 3 + 2.1 \Rightarrow 5.1 + 2.1 = 7.2$

Expected values of y	Predicted values of y
3	3.8
6	7.2

New values of m and b

$$m_{new} = m_{1st_i} - \alpha \frac{\Delta J}{\Delta m}$$

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

$$\frac{\Delta J}{\Delta m} = \frac{-\cancel{2}}{\cancel{2}} \left[(3-3.8)(1) + (6-7.2)(3)\right]$$

$$\frac{\Delta J}{\Delta m} = -\left[(-0.8) + (-1.2)(3)\right]$$

$$\frac{\Delta J}{\Delta m} = -(-0.8 - 3.6)$$

$$= -(-4.4)$$

$$\frac{\Delta J}{\Delta m} = 4.4$$

$$m_{new} = 1.7 - (0.1 \times 4.4) \Rightarrow 1.7 - 0.44 = \underline{\underline{1.26}}$$

$$m_{new} = 1.26.$$

$$b_{new} = b_{1st_i} - \alpha \frac{\Delta J}{\Delta b}$$

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

$$\frac{\Delta J}{\Delta b} = \frac{-2}{2} \left[(3-3.8) + (6-7.2)\right]$$

$$-\left[(-0.8) + (-1.2)\right]$$

$$= \underline{2}$$

$$b_{new} = 2.1 - (0.1 \times 2) \Rightarrow 2.1 - 0.2$$

$$b_{new} = \underline{1.9}$$

2nd Iteration

$$b_{new} = 1.9$$

$$m_{new} = 1.26$$

3^{rd} iteration

Using m & b from iteration 2

$m = 1.26$

$b = 1.9$

y - Predictions:

for (1,3): $\hat{y}_1 = (1.26 \times 1) + 1.9 = \underline{\underline{3.16}}$

for (3,6): $\hat{y}_2 = (1.26 \times 3) + 1.9 = \underline{\underline{5.68}}$

expected y	Predicted y
3	3.16
6	5.68

Finding m_{new} & b_{new}

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

$$m_{new} = 1.26 - \left[0.1 * \left[\left(-\frac{2}{n}\right)\left[(3-3.16)(1) + (6-5.68)(3)\right]\right]\right]$$

$$m_{new} = 1.26 - \left[0.1 \left(-\frac{2}{2}\right)\left[(-0.16) + (0.96)\right]\right]$$

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

$$m_{new} = 1.34$$

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

$$b_{new} = 1.9 - \left[0.1 * \left(-\frac{2}{n}\right)\left[(3-3.16) + (6-5.68)\right]\right]$$

$$b_{new} = 1.9 - \left[0.1 \times \left(-\frac{2}{2}\right)\left[(-0.16) + (0.32)\right]\right]$$

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

$$b_{new} = 1.9 + 0.016$$

$$b_{new} = 1.916$$

Iteration 4

1. Predicted values

point (1,3) $\hat{y} = 1.34 \times 1 + 1.916 = 3.256$

(3,6) $\hat{y} = 1.34 \times 3 + 1.916 = 5.936$

2. Gradients

$$\frac{\partial J}{\partial m} = -\frac{2}{2}\left[(3-3.256)(1) + (6-5.936)(3)\right]$$

$$= -[-0.256 + 0.192] = 0.064$$

$$\frac{\partial J}{\partial b} = -\frac{2}{2}\left[(3-3.256) + (6-5.936)\right]$$

$$= -[-0.256 + 0.064] = 0.192$$

To update parameters

$m = 1.34 - 0.1 \times 0.064$

$= 1.34 - 0.0064 = 1.3336$

$b = 1.916 - 0.1 \times 0.192$

$= 1.916 - 0.0192 = 1.8968$

Summary

Iteration Results	m	b
Initially	-1	1
1	1.7	2.1
2	1.26	1.9
3	1.34	1.916
4	1.3336	1.8968

CamScanner

Our observation

The parameters show an oscillating but converging pattern

1. The first update made large changes

   m from -1 to 1.7

   b " " 1 to 2.1

2. Subsequent updates made smaller adjustments (3rd and)

3. By the 4th iteration changes became very smaller

   $\Delta m \simeq 0.0064$, $\Delta b \simeq 0.0192$

4. The values are stabilizing around $m \simeq 1.33$, $b \simeq 1.9$ suggesting convergence

Each iteration successfully reduced the error with the magnitude of updates decreasing as we approach optimal values.