Given data

Linear equation $y = mx + b$

Initial $m = -1$, $b = 1$

learning rate $\alpha = 0.1$

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

As m = slope (updated)
b = Intercept (updated)

α = learning rate
$n = 2$ (no of data points)

* Before Iteration

Initial values $m = -1$, $b = 1$

Iteration 1: $\hat{y}_i = mx_i + b$

points (x, y)	$\hat{y} = mx + b$	$y - \hat{y}$
$(1,3)$	$-1(1) + 1 = 0$	$3 - 0 = 3$
$(3,6)$	$-1(\times 3) + 1 = -2$	$6 - -2 = 8$

By using mean squared error (MSE)

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

$$= \frac{-2}{2}\left[(1)(3) + (3)(8)\right] = -1\left[3 + 24\right] = -27$$

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

$$= -1 \left[3+8\right] = -11$$

Lets update m and b

~~$m = m - \alpha \cdot \partial$~~

$$m = m - \alpha \cdot \frac{\partial J}{\partial m} = -1 - 0.1(-27) = -1 + 2.7 = 1.7$$

$$b = b - \alpha \cdot \frac{\partial J}{\partial m} = 1 - 0.1(-11) = 1 + 1.1 = 2.1$$

$m = 1.7$, $b = 2.1$

Iteration 2:

1. New $m = 1.7$, $b = 2.1$

Point (x, y)	$y' = 1.7x + 2.1$	$y - y'$
$(1, 3)$	$1.7 \times 1 + 2.1 = 3.8$	$3 - 3.8 = -0.8$
$(3, 6)$	$1.7 \times 3 + 2.1 = 7.2$	$6 - 7.2 = -1.2$

2. Gradients:

$$\frac{\partial J}{\partial m} = -1\left[(1)(-0.8) + (3)(-1.2)\right] = -1(-0.8 - 3.6) = 4.4$$

$$\frac{\partial J}{\partial b} = -1(-0.8 - 1.2) = 2.0$$

3. Update:

$$m = 1.7 - 0.1(4.4) = 1.7 - 0.44 = 1.26$$

$$b = 2.1 - 0.1(2.0) = 2.1 - 0.2 = 1.9$$

Iteration 3.

New $m = 1.26$, $b = 1.9$

points (x,y)	$\hat{y} = 1.26x + 1.9$	$y - \hat{y}$
(1,3)	$1.26 + 1.9 = 3.16$	$3 - 3.16 = -0.16$
(3,6)	$3.78 + 1.9 = 5.68$	$6 - 5.68 = 0.32$

$$\frac{\partial J}{\partial m} = -1[(1)(-0.16) + (3)(0.32)] = -1[-0.16 + 0.96] =$$
$$= -0.8$$

$$\frac{\partial J}{\partial m} = -1[-0.16 + 0.32] = -0.16$$

Updated. $m = 1.26 - 0.1(-0.8) = 1.26 + 0.08 = 1.34$

$b = 1.9 - 0.1(-0.16) = 1.9 + 0.016 = 1.916$

$m = 1.34$, $b = 1.916$

iteration 4:

1: New $m = 1.34$, $b = 1.916$

Point (x, y)	$\hat{y}$	$y - \hat{y}$
$(1, 3)$	$1.34 + 1.916 = 3.256$	$3 - 3.256 = -0.256$
$(3, 6)$	$4.02 + 1.916 = 5.936$	$6 - 5.936 = 0.064$

2: gradient:

$$\frac{\partial J}{\partial m} = -1\left[(1)(-0.256) + (3)(0.064)\right]$$

$$= -1\left[-0.256 + 0.192\right]$$

$$= 0.064$$

$$\frac{\partial J}{\partial b} = -1\left[-0.256 + 0.064\right] = 0.192$$

3: update:

$$m = 1.34 - 0.1(0.064)$$

$$= 1.34 - 0.0064$$

$$= 1.3336$$

$$b = 1.916 - 0.1(0.192)$$

$$= 1.916 - 0.0192$$

$$= 1.8968$$

$m = 1.3336$ & $b = 1.8968$

Observations

* m Increased from 1 → 1.7 → 1.26 → 1.34 → 1.3336.

* b Increased from 1 → 2.1 → 1.9 → 1.916 → 1.896 8

* The updated are getting smaller, meaning we are approaching the minimum of the cost function.

* The predicted values are getting closer to the true "y" values meaning the error is reducing.

By conclusion: Yes, the parameters m and b are moving toward values that reduces the prediction error gradient descent is working.