

Gradient descent

initial values

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

*	1	3
y	3	6

lets find \hat{y} using m and b

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

$$= (-1)(1) + 1 = 0$$

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

expected y | predicted y

3
6

0
-2

lets find the new value of m and b

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

$$\text{using the } MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Suppose MSE is J

then the derivative differentiation of MSB
with respect to m is

$$\frac{\partial J}{\partial m} = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

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

$$\frac{\partial J}{\partial m} = \frac{1}{n} \sum (y_i - mx_i - b)^2$$

using the chain rule where

$$u = y_i - mx_i - b$$

now we have

$$\frac{\partial J}{\partial m} = \frac{1}{n} \sum \frac{\partial}{\partial m} (u^2)$$

$$\text{Also } \frac{\partial}{\partial m} = \frac{1}{n} \sum \frac{\partial}{\partial m} (u^2)$$

$$= \frac{1}{n} \sum \frac{\partial J}{\partial u} \cdot \frac{\partial u}{\partial m}$$

$$\frac{\partial J}{\partial u} \cdot \frac{\partial}{\partial u} (u^2) = 2u$$

$$\frac{\partial u}{\partial m} = \frac{\partial}{\partial m} (y_i - mx_i - b)$$

m is the changing value others are constant

bringing everything together we get

$$\frac{\partial J}{\partial m} = \frac{1}{n} \sum 2(y_i - mx_i - b)(-x_i)$$

$$= \boxed{-\frac{2}{n} (y_i - \hat{y}_i) x_i}$$

now that we have the formula of $\frac{\partial J}{\partial m}$
 let's find M_{new}

$$\frac{\partial J}{\partial m} = \frac{2}{2} ((3-0)(1) + (6-(-2))(3))$$

$$(\frac{2}{2}) = \frac{1(3+24)}{2} = \underline{\underline{-27}}$$

$$M_{new} = \frac{1-0.1(-27)}{n}$$

let's differentiate past formula again

with respect to b

$$\frac{\partial J}{\partial b} = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

$$= \frac{1}{n} \sum (y_i - (mx_i + b))^2$$

using chain rule we have

$$\frac{\partial J}{\partial b} \leftarrow \left(\frac{\partial J}{\partial u} \cdot \frac{\partial u}{\partial b} \right) \text{ when } u = y_i - mx_i - b$$

$$\frac{\partial J}{\partial u} = \frac{\partial}{\partial u} (u^2) = 2u$$

$$\frac{\partial u}{\partial b} = \frac{\partial}{\partial b} (y_i - mx_i - b)$$

because b is the only changing value the rest are just constants

$$\frac{\partial J}{\partial u} = \frac{1}{n} \sum_{i=1}^n 2(y_i - mx_i - b) (-1)$$

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

$$= -\frac{2}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

replacing with our values of y_i, \hat{y}_i , we get

$$\frac{\partial J}{\partial u} = -\frac{2}{n} \sum_{i=1}^n ((3-0) + (6-(-2)))$$

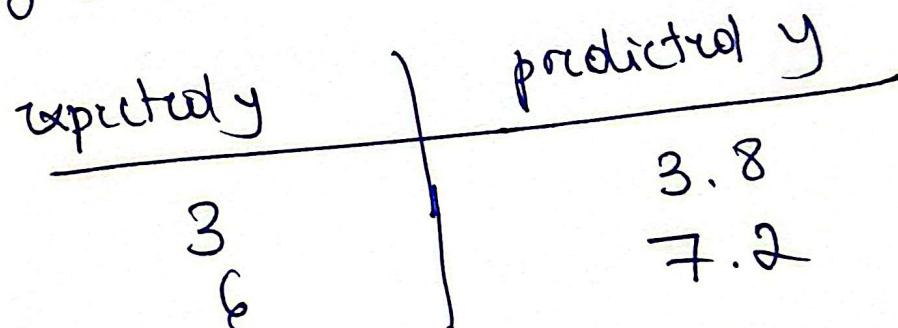
$$= -1 (11) = \underline{\underline{-11}}$$

$$b_{\text{new}} = b_{\text{old}} - \alpha \frac{\partial J}{\partial b}$$
$$= 1 - (0.1)(11) = \underline{\underline{2.1}}$$

Now let's find the new value of \hat{y} using the new m and b

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

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



Iteration 2:

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

$$= -\frac{2}{9} \sum [(3-3.8)(1) + (6-7.2)(3)]$$

$$= -1 (-0.8 - 3.6)$$

$$= 4.4$$

$$M_{new} = M_{old} - \alpha \frac{\partial J}{\partial w}$$

$$= +1.7 - 0.1(4.4) = 1.26$$

$$\frac{\partial J}{\partial b} = -\frac{2}{9} [(3-3.8) + (6-7.2)] = 2/1$$

$$b_{new} = b_{old} - \alpha \frac{\partial J}{\partial b} = 2-1 - 0.1(2) = 1.9/1$$

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

$$\hat{J}_2 = (1.26)(3) + 1.9 = 5.68$$

Expected Y	Predicted Y
3	3.16
6	5.68

Iteration 3 :

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

$$= -\frac{2}{2} ((3 - 3.16)(1) + (6 - 5.68)(3))$$

$$= -1 (-0.16 + 0.96) = \underline{-0.8}$$

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

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

$$= -1 ((3 - 3.16) + (6 - 5.68))$$

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

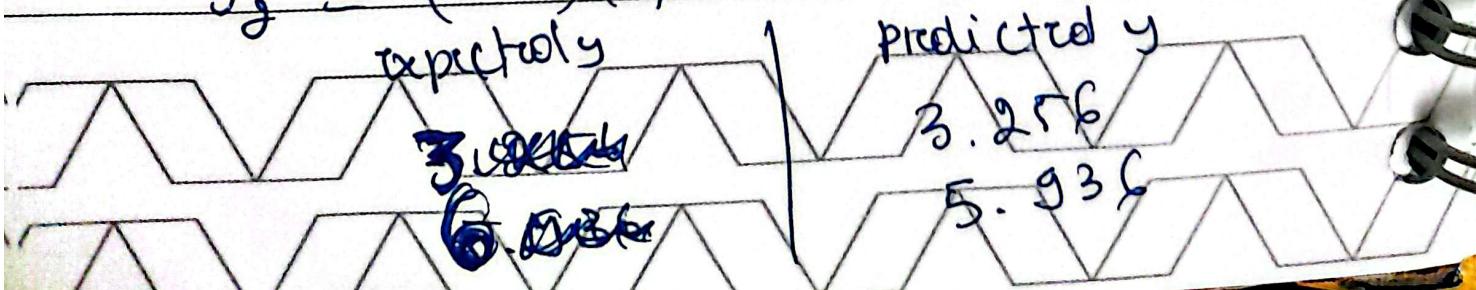
$$b_{\text{new}} = b_{\text{old}} \times \frac{\partial J}{\partial b}$$

$$= 1.9 - 0.1(-0.16)$$

$$= 1.916$$

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

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



Iteration 4

$$\frac{\partial J}{\partial m} = -\frac{2}{n} \sum (y_i - \hat{y}_i) = 2x^i$$
$$= -1 ((3-3.256)(1)+(6-5.934)(8))$$
$$= -1 (-0.256 + 0.196)$$
$$= 0.064$$

$$M_{\text{new}} = M_{\text{old}} - \alpha \frac{\partial J}{\partial m}$$
$$= 1.39 - 0.1(0.064) = 1.3356$$

$$\frac{\partial J}{\partial b} = -\frac{2}{n} \sum (y_i - \hat{y}_i)$$
$$= -1 ((3-3.256) + (6-5.934))$$
$$= -1 (-0.256 + 0.064)$$
$$= 0.192$$

$$b_{\text{new}} = b_{\text{old}} - \alpha \frac{\partial J}{\partial b}$$
$$= 1.916 - 0.1(0.192)$$
$$= 1.8968$$

Find \hat{y} using now m and b

$$\hat{y}_1 = (1.333)(1) + 1.8968$$

$$= 3.2304$$

$$\hat{y}_2 = (1.333)(5) + 1.8968$$

$$= 5.8976$$

expected		predicted \hat{y}
3		3.2304
6		5.8976

Conclusion

Conclusion

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

In Iteration one, there is a huge jump since error was high but the rest is low because prediction improved. So with more iterations we can get the optimal value for b and m .