

Goal: minimize the cost function $J(\theta) = \min_{w,b} J(w,b)$

We need to have some function $J(w,b) \rightarrow$

For linear regression or any function

$$\min_{w_1, \dots, w_n, b} J(w_1, w_2, \dots, w_n, b)$$

Outline:

- Start with some w, b (set $w=0, b=0$)

- keep changing w, b to reduce $J(w,b)$

Until we settle at or near a minimum

may have > 1 minimum

Implementation of Gradient descent algorithm

We want to repeat until it converge

Correct way to: Simultaneously update w and b

$$w = w - \alpha \frac{d}{dw} J(w,b)$$

$$b = b - \alpha \frac{d}{db} J(w,b)$$

Simultaneously update w and b

$$\text{temp_}w = w - \alpha \frac{d}{dw} J(w,b)$$

$$\text{temp_}b = b - \alpha \frac{d}{db} J(w,b)$$

$$w = \text{temp_}w$$

$$b = \text{temp_}b$$

Where:

α : Learning rate (0,1)

$\frac{d}{dw} J(w,b)$: Derivative term of the cost function J

Intuition

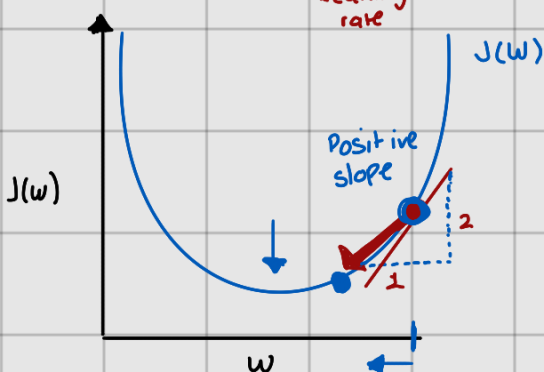
repeat until convergence {

$$\text{Params} \begin{cases} \underline{w} = w - \alpha \frac{\partial}{\partial w} J(w,b) \\ \underline{b} = b - \alpha \frac{\partial}{\partial b} J(w,b) \end{cases}$$

Partial Derivative

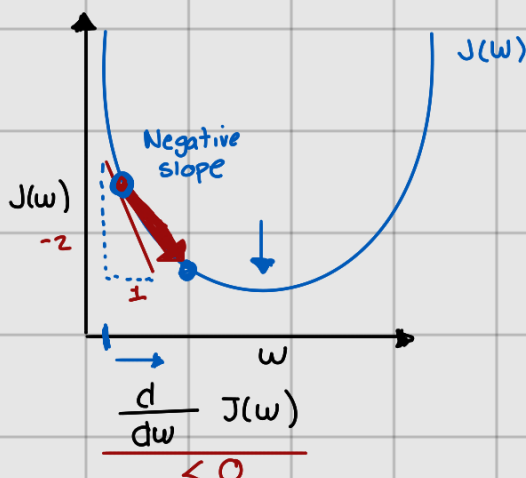
Learning rate

$$J(w) \\ w = w - \alpha \frac{d}{dw} J(w)$$



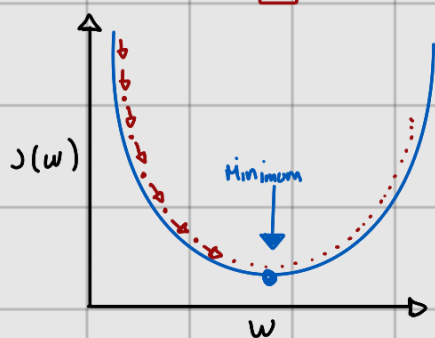
$$w = w - \alpha \frac{d}{dw} J(w) \\ > 0$$

$$w = w - \alpha \cdot (\text{positive number})$$

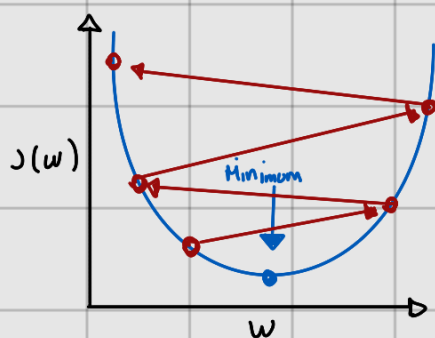


$$w = w - \alpha \cdot (\text{negative number})$$

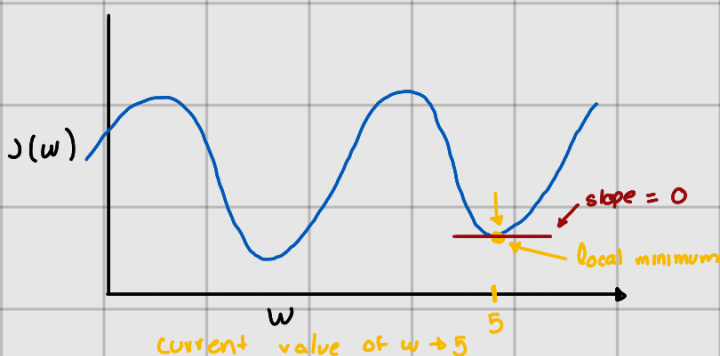
$$w = w - \boxed{\alpha} \frac{d}{dw} J(w)$$



- If α is too small, gradient may be slow



- If α is too large, gradient descent may:
 - Overshoot, never reach minimum
 - Fail to converge, diverge



If the slope is 0, then

$$w = w - \alpha \frac{d}{dw} J(w)$$

$= 0$

$$w = w - \alpha \cdot 0$$

$$w = w$$

$$w = 5$$

$$5 = 5$$

its the same
current value of w

Near a local minimum

- Derivative become smaller
- Update steps become smaller

Can reach minimum without decreasing
learning rate.