Gradient descent algorithm

repeat until convergence {

learning rate

derivative

$$\begin{cases} w = w - \alpha \boxed{\dfrac{\partial}{\partial w} J(w,b)} \\ b = b - \alpha \dfrac{\partial}{\partial b} J(w,b) \end{cases}$$

$J(w)$

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

$$\min_{w} J(w)$$

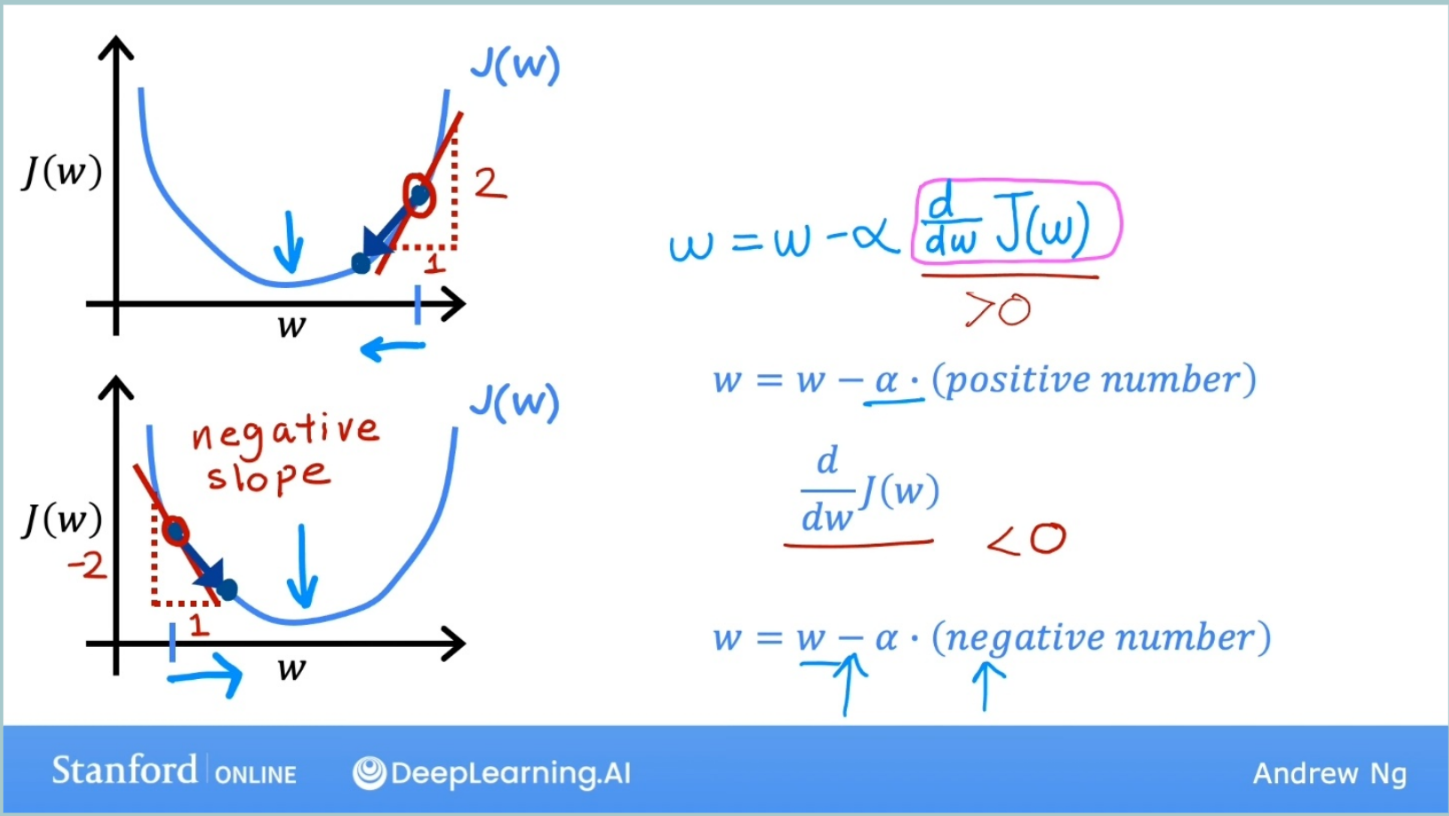
Stanford ONLINE   DeepLearning.AI   Andrew Ng

→ Here is the gradient descent formulas.

→ We have a partial derivative term here with muliply by α that is learning rate.

→ Now, to understand we use the simplified version of those formula.

→ That is $w = w - \alpha \frac{\partial}{\partial w} J(w)$, $b = 0$



→ If we pick any point from the graph, the gradient descent will gradually decrease that point toward local minima.

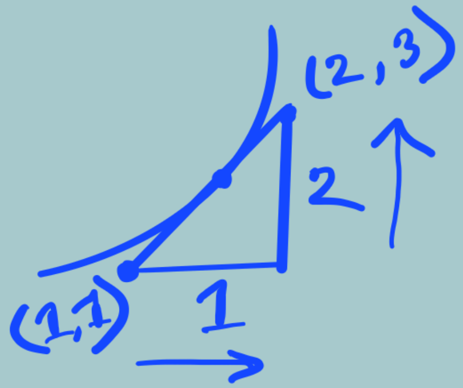
→ Here the derivative term is,

$$\frac{dJ(w)}{dw}$$

this means changes in y axes over changes in x axes.

→ And we know it is a slope of that point.

→ So, for the first point the slope is positive



→ If we consider this example we can see how the slope is positive.

→ Here, slope $= \dfrac{3-1}{2-1} \left(\dfrac{y_2 - y_1}{x_2 - x_1}\right)$

$= 2$ (positive)

→ If the derivative term is positive than the w will decrease as well as the point will decrease toward local minima.

→ same will happen for the second point.

→ It is a negative slope, so the w value will increase and the point will decrease toward local minima.