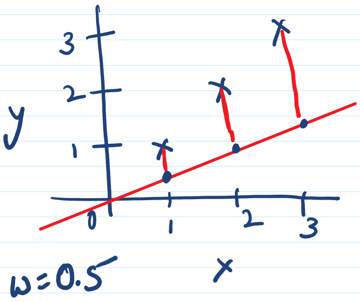
$f(x)$	$J(w)$, $b=0$
function of x	function of w



$$J(w) = \frac{1}{2m} \sum_{i=0}^{m} (\hat{y}_i - y_i)^2$$

$$J(1) = \frac{1}{2m} (0^2 + 0^2 + 0^2)$$

$$= 0$$

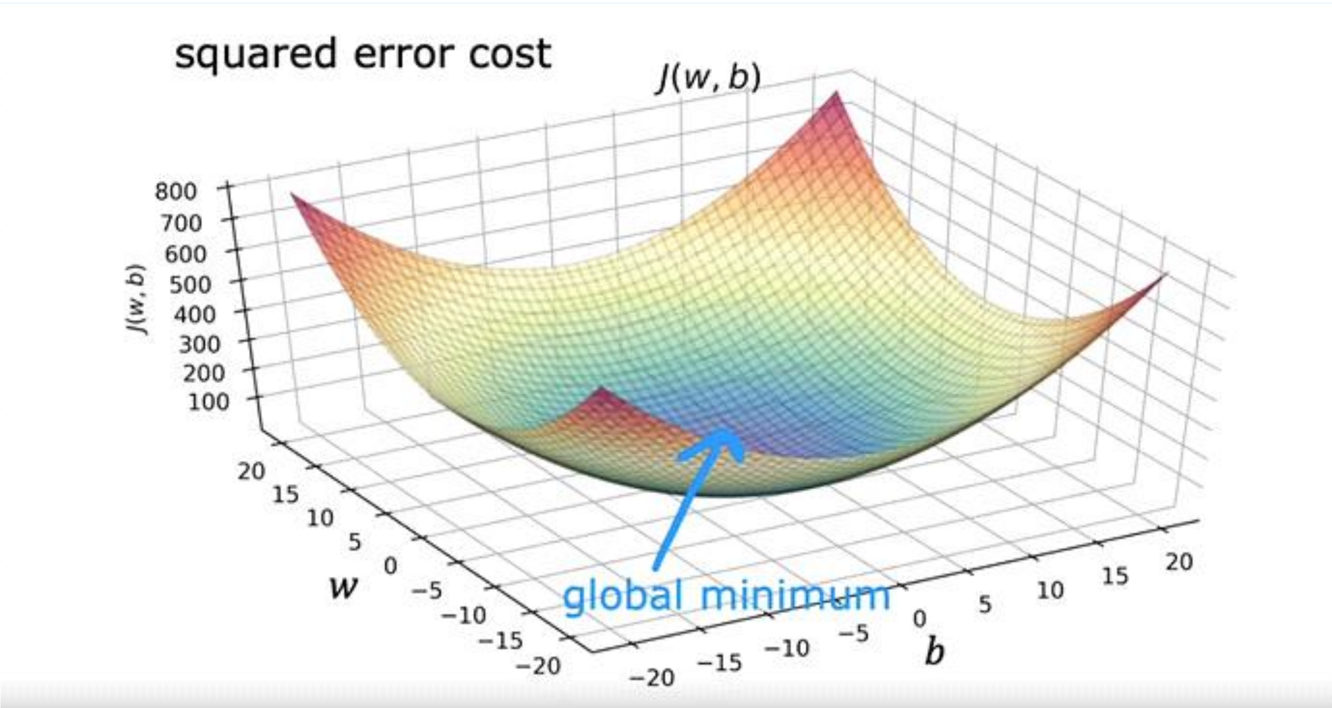
$$J(0.5) = \frac{1}{2m} \left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2 \right]$$

$$= \frac{1}{2 \times 3} [3.5]$$

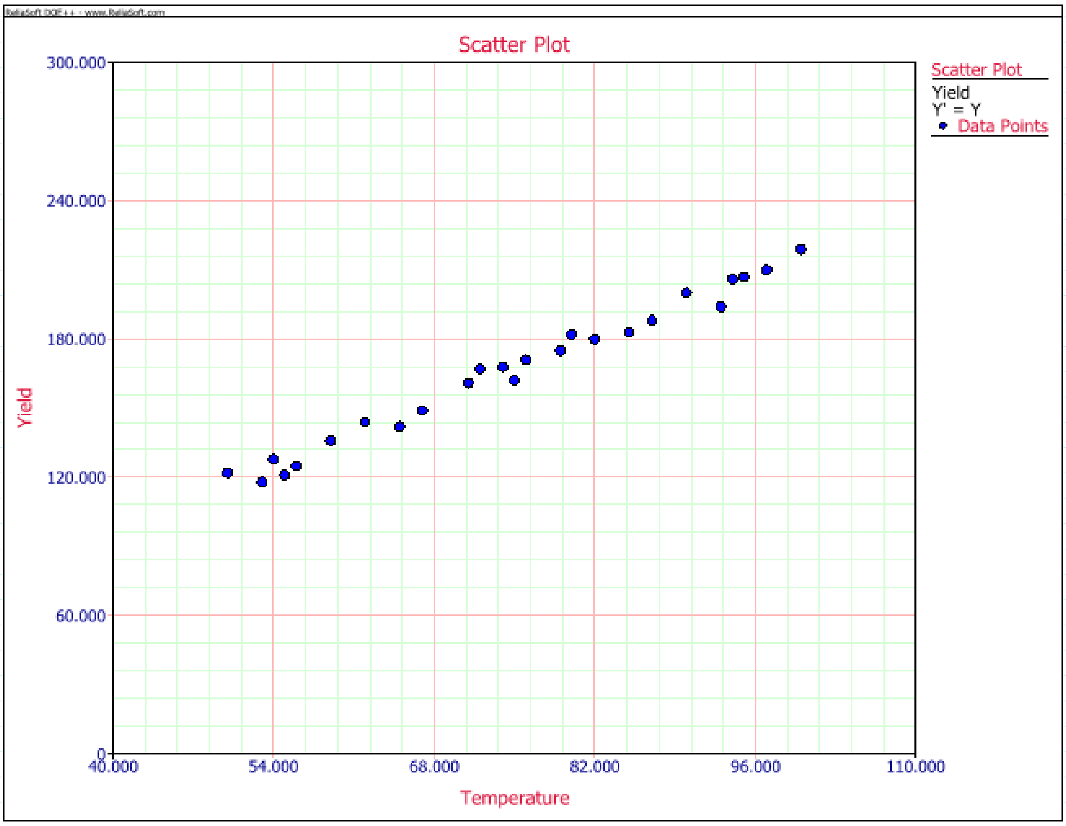
$$= 0.58$$

$$J(0) =$$

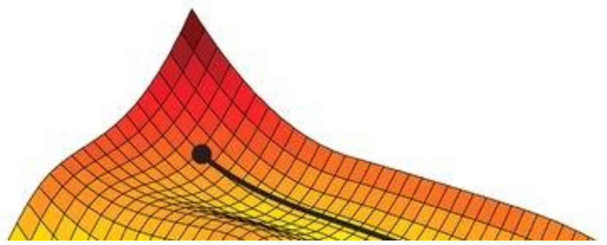
But $J(w, b)$ is a two parameters function.

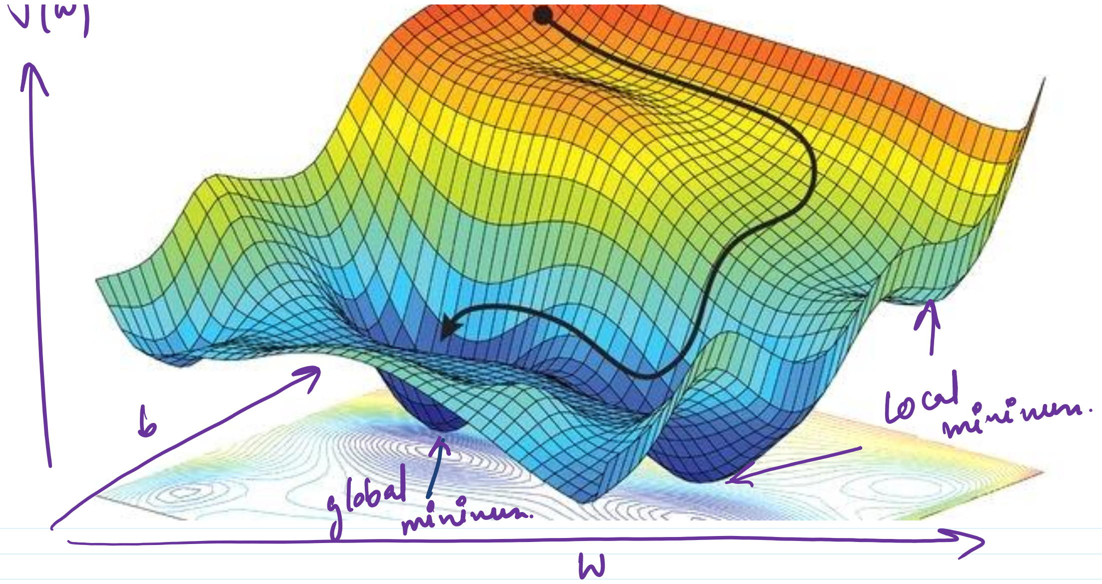


Linear Regression



J(w)





Neural Network :– $J(w_1, w_2, w_3, \ldots, b)$

(Not a squared error cost function)

Goal :– For function $J(w,b)$

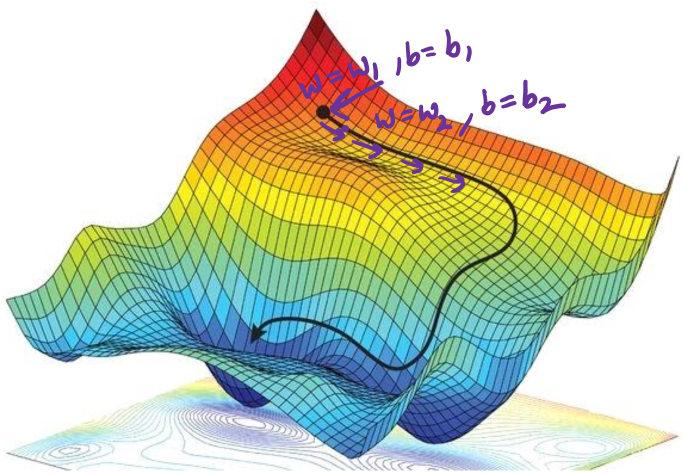
we want $\min\limits_{w,b} J(w,b)$

Outline :–

Start with some initial values of w,b.

(Let w=0, b=0)

Keep changing w,b to reduce J(w,b) until we find minimum or nearest min.



Gradient Descent Algorithm :–

Repeat until convergences

$$\{$$

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

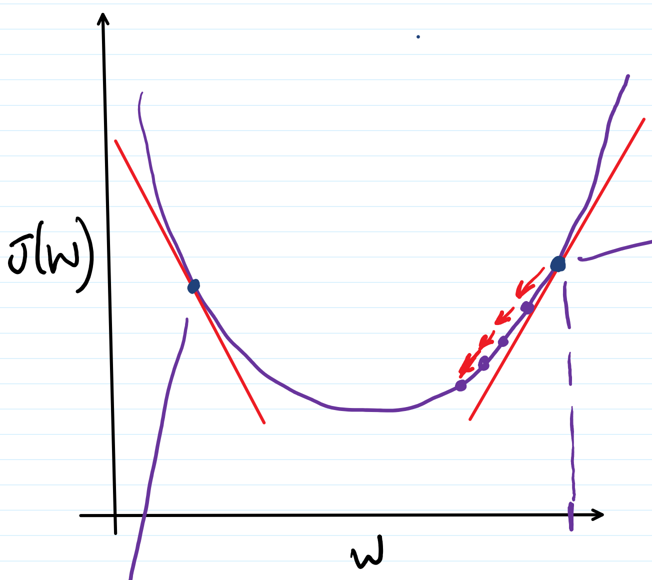
$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$\}$$

$\alpha \rightarrow$ learning rate

$\frac{\partial}{\partial w} J(w,b) \rightarrow$ derivative of $J(w)$

$\frac{\partial}{\partial b} J(w,b) \rightarrow$ derivative of $J(b)$



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

$$w = w - \underline{\alpha.}\,(+ve)$$

$$w = w - (+ve)$$

$w \downarrow$ decreasing

$$w = w - \alpha \underbrace{\frac{d}{dw} J(w)}_{<0}$$

$$w = w - \alpha.(-ve)$$

$= w - (-ve)$

$w \uparrow$ increasing

IIly for "b".

Choosing <u>"α"</u>.

Too small

Too large.