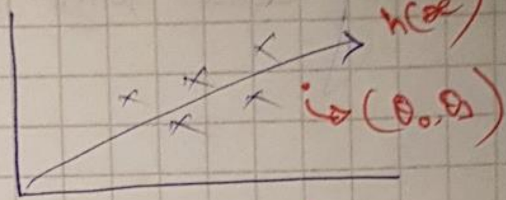
Unsupervised : → no data label.



Hypothesis :

→ draw the line based on given data set.

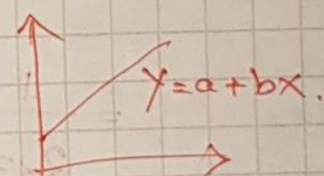
Cost function : (also called squared error func.)

→ calculate the accuracy of our hypothesis function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i \right)^2$$

Here,
x = input
y = actual output.

Hypothesis : $h_\theta(x) = \theta_0 + \theta_1 x$



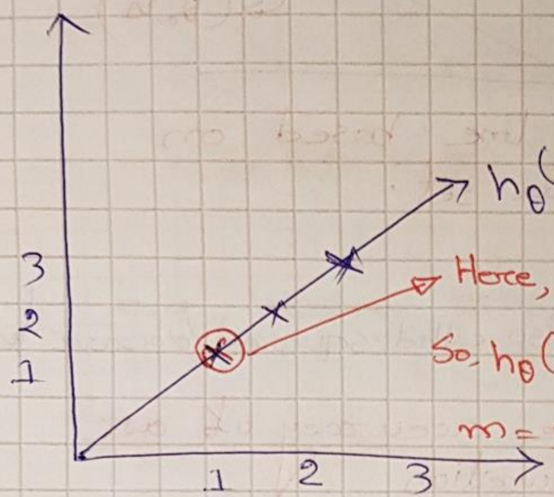
Parameters : θ_0, θ_1

cost function :

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i \right)^2$$

Goal :
minimize cost func. $J(\theta_0, \theta_1)$

calculating cost func. from hypothesis:



$h_\theta(x) = \theta_0 + \theta_1(x)$

Here, $\theta_0 = 0$, $\theta_1 = 1$

So, $h_\theta(x) = y_i$ (cause in this line x and y value same)

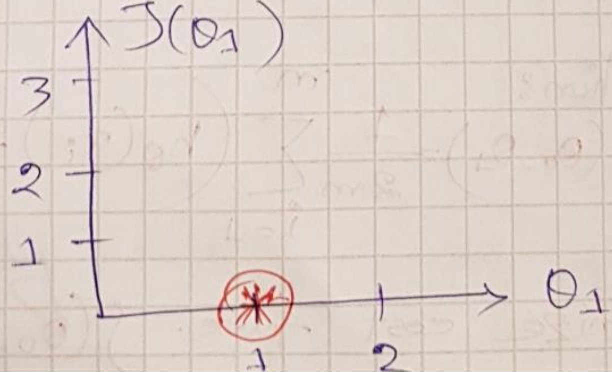
$m = 3$ = data points

Cost func for $\theta_1 = 1$:

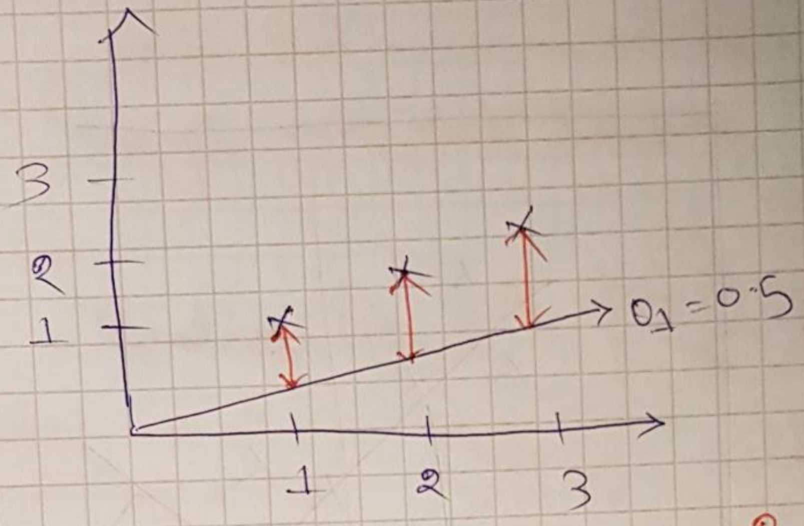
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2$$

$$= \frac{1}{2\times3} \sum_{i=1}^{3} \left(\theta_1(x_i) - y_i\right)^2$$

$$= \frac{1}{6} \left(0^2 + 0^2 + 0^2\right) = 0 = J(1)$$

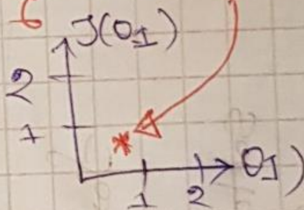


Cost function when $\theta_1 = 0.5$:

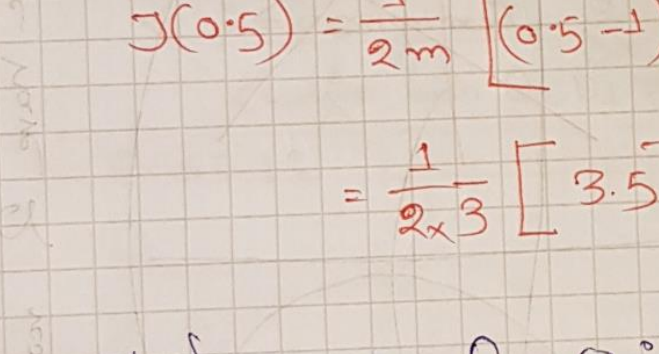


$$J(0.5) = \frac{1}{2m}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$$= \frac{1}{2\times 3}\left[3.5\right] = \frac{3.5}{6} \approx 0.58$$



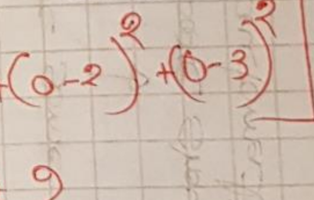
Cost func when $\theta_1 = 0$:



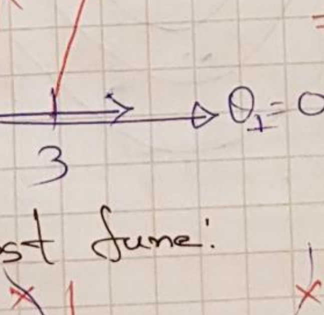
$$J(0) = \frac{1}{2\times 3}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

$$= \frac{1+4+9}{6}$$

$$= 14/6 = 2.33$$

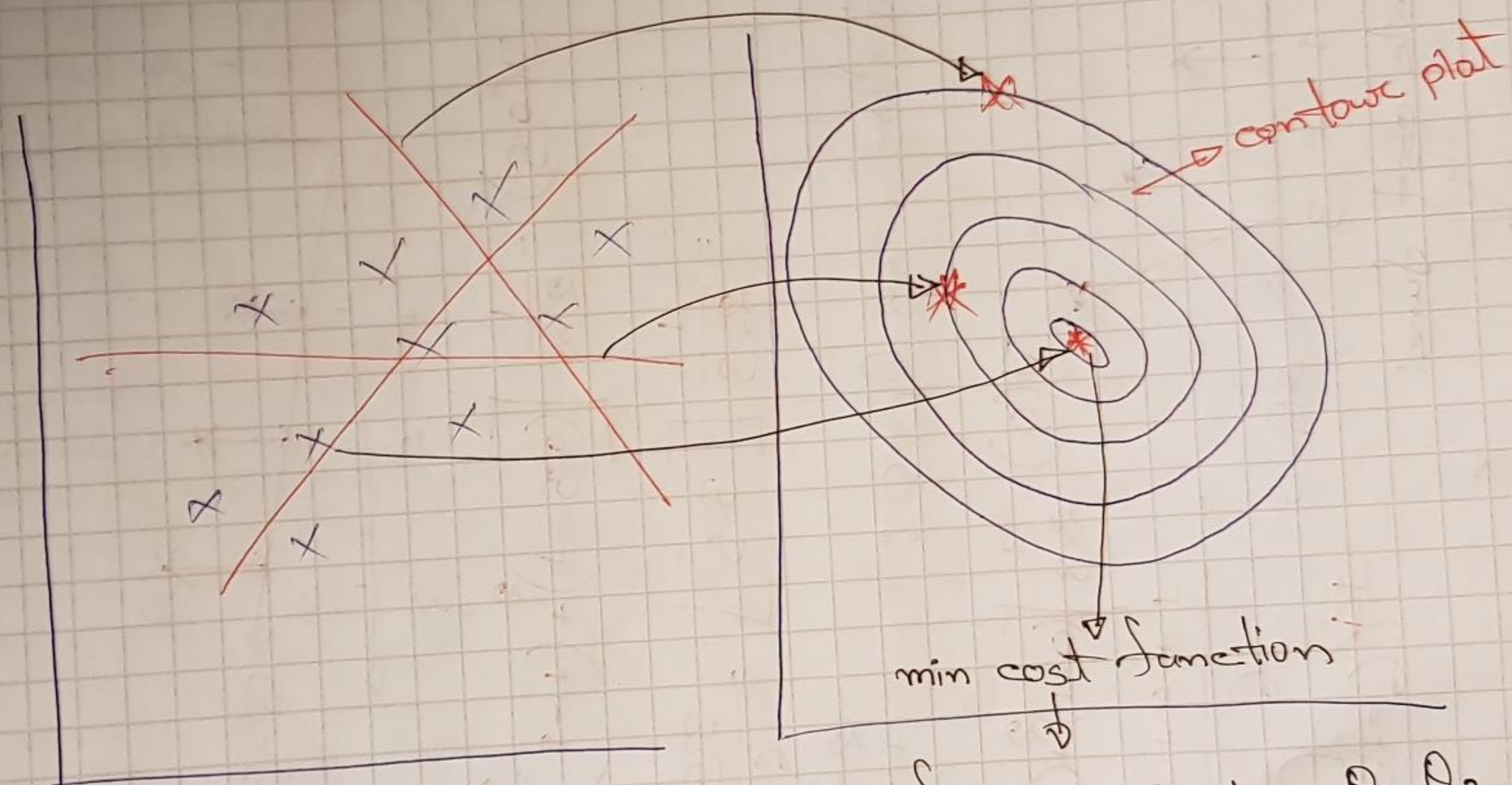


Ultimate cost func:



min. cost func. when $\theta_1 = 1$.

Cost function when we have 2 variable : θ_1, θ_2



—— ⇒ different ~~cost~~ ~~for~~ hypothesis

↓

for some value θ_1, θ_2

↓

Software should find it - θ_1, θ_2 automatically

Gradient Descent:

→ can find minimum position

→ get min. cost function

Equation: $\theta_j := \theta_j - \boxed{\alpha} \cdot \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$

small α = small step
large α = large step

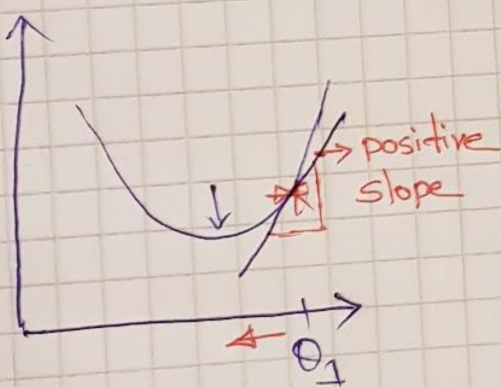
learning rate

slope / tangent

$\langle$ $:=$ assignment, $\frac{\partial}{\partial x}$ = partial derivative $\rangle$

$\langle$update both θ_0, θ_1 simultaneously$\rangle$

CASE 1:



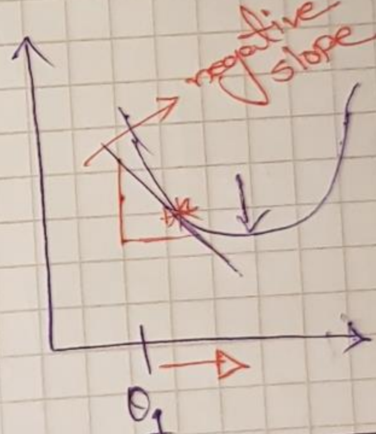
$\theta_1 := \theta_1 - \alpha \boxed{\frac{d}{d\theta_1} J(\theta_1)}$

positive: > 0

$\theta_1 := \theta_1 - \alpha \boxed{\text{+ve number}}$

= Smaller value of θ_1

CASE 2:



$\theta_1 := \theta_1 - \alpha \boxed{\frac{d}{d\theta_1} J(\theta_1)}$

neg → ≤ 0

$\theta_1 := \theta_1 - \alpha(\text{-ve value})$

$= \theta_1 + (\text{+ve val})$

= greater θ_1