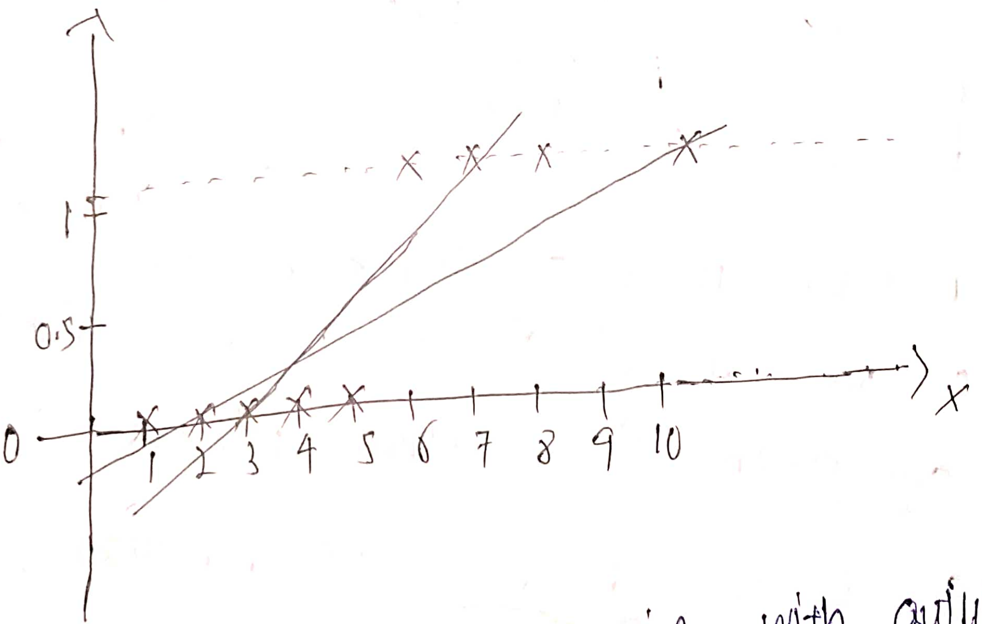
Logistic Regression

Dataset

study hours	O/P (Pass/fail) 1 0
2	fail
3	fail
4	fail
5	fail
6	pass
7	Pass
8	Pass
9	fail.

(i) Can we solve this problem with Linear Regression.



In case of Linear regression with outliers line shifts its position. so the thresold fails to achive.

$Y \le 0.5 = 0$, $y > 0.5 = 1$

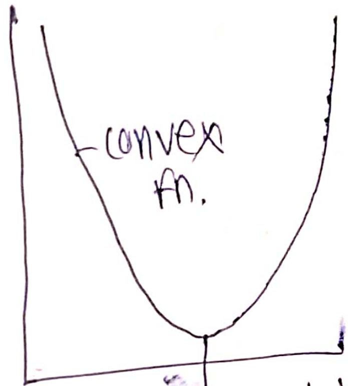
To fix this we use Logistic Regression.

On top of liner line we add Sigmoid function.

$$h_\theta(x) = \text{Sigmoid}(\theta_0 + \theta_1 x)$$

$$= \underbrace{\frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}}_{0 \text{ to } 1} \quad \left[\sigma(\text{sigmoid}) = \frac{1}{1+e^{-x}} \right]$$

But here MSE cost function is generate a non-convex function



But we need a convex function so we will not stuck at local minima. and reach to global minima :

So we use <u>Log loss cost function</u>.

* Log Loss Cost function

$$\text{cost}(h_\theta(x)^{(i)}, y^{(i)}) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

[Never local minima]

$$\boxed{\text{cost}(h_\theta(x)^{(i)}, y^{(i)}) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))}$$

By putting value of y (1 or 0) we will get above equations.

↓

minimize cost function $J(\theta_0, \theta_1)$ by changing θ_0, θ_1 using convergence algorithm.

Repeat until convergence
{

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

}