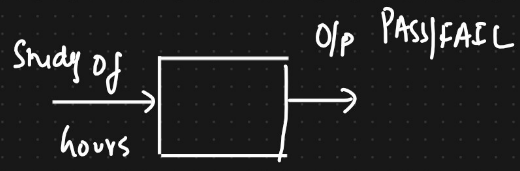
Logistic Regression (Binary classification) ←

Dataset

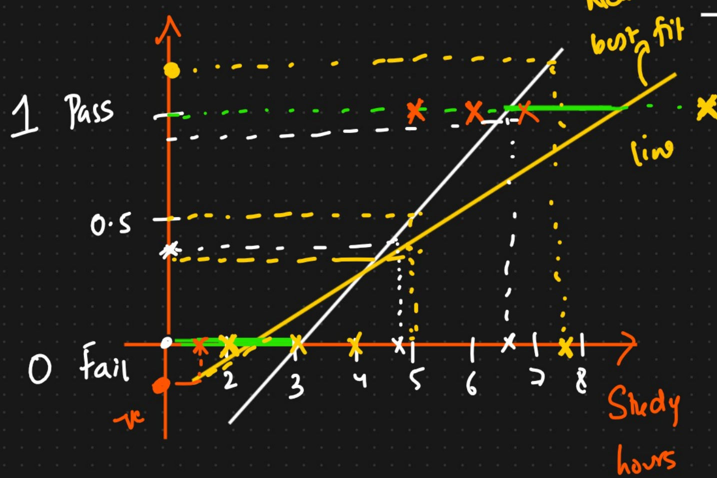
Study hours	Pass/Fail O/p {Binary categories}
2	Fail
3	Fail
4	Fail
5	Pass
6	Pass
7	Pass
12	Pass



Logistic Regression

0 to 1

Linear Regression



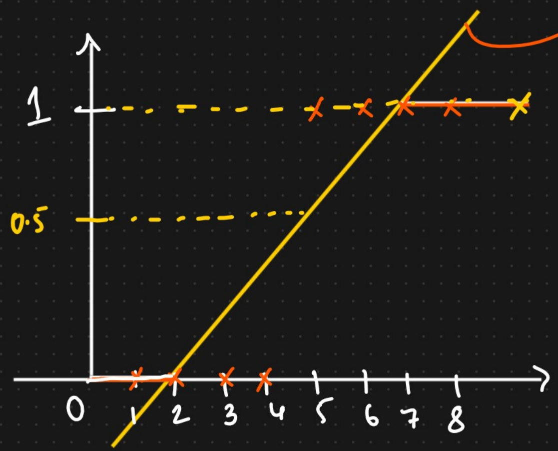
$$\left\{\begin{array}{l} \leq 0.5 \Rightarrow 0 \\ > 0.5 \Rightarrow 1 \end{array}\right\}$$

$$\left\{\begin{array}{l} > 1 \text{ and} \\ < 0 \end{array}\right\}$$

Why we cannot use Linear Regression for Classification?

1. Outlier {Best fit line Change}
2. > 1 and < 0 {Squash line}

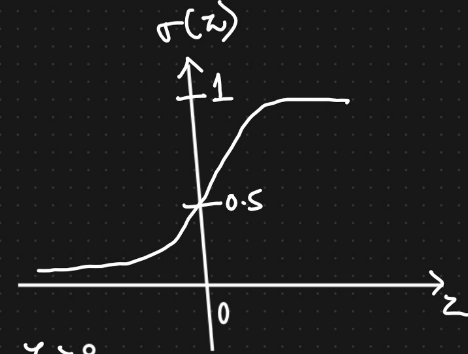
How Logistic Regression Solves Classification Problem



$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$$

$\Downarrow$

Best fit line



$z \geq 0$

$\sigma(z) \geq 0.5$

Sigmoid Activation

$$\frac{1}{1+e^{-z}}$$

$z = \theta_0 + \theta_1 x_1$

$\Uparrow$

$$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1) \qquad \sigma = \frac{1}{1+e^{-z}}$$

$$= \sigma(z)$$

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-z}}} \Rightarrow$$

Logistic Regression hypothesis

$$z = \theta_0 + \theta_1 x_1$$

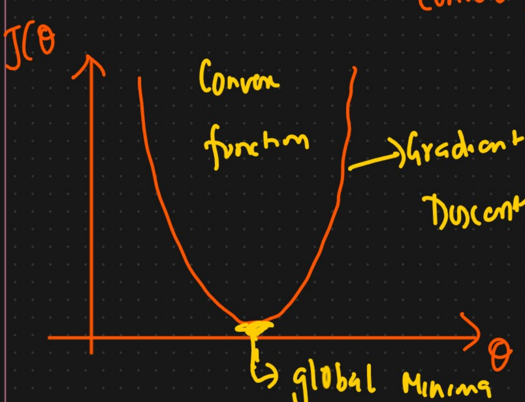
Linear Regression Cost function

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

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

$\Downarrow$

Convex function



Logistic Regression Cost function

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

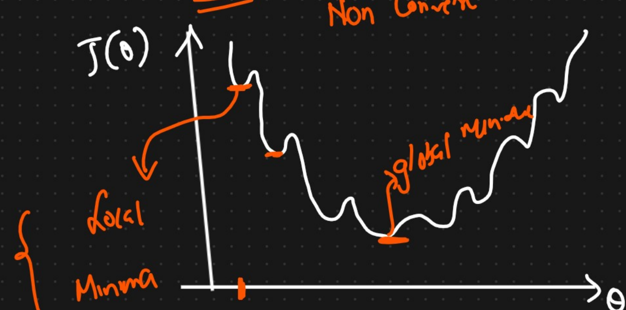
Non Convex function

$\Downarrow$

$$h_\theta(x) = \frac{1}{1+e^{-z}} \qquad z = \theta_0 + \theta_1 x_1$$

Sigmoid

Non Convex



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

$h_\theta(x)^{(i)} = \frac{1}{1+e^{-z}}$ $z = \theta_0 + \theta_1 x_1$

$\Downarrow$

Lets Denote $\mathrm{Cost}(h_\theta(x)^{(i)}, y^{(i)})$

$\Downarrow$

{ Log Loss }

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

Convex function

$\Downarrow$

$$\mathrm{Cost}(h_\theta(x)^{(i)}, y^{(i)}) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

$$\boxed{J(\theta_0,\theta_1) = -\frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)}\log(h_\theta(x)^{(i)}) - (1-y^{(i)})\log(1-h_\theta(x)^{(i)})\right)}$$

Minimize Cost function $J(\theta_0,\theta_1)$ by changing θ_0 & θ_1

Convergence Algorithm

Repeat

$\{$ $j=0$ and 1

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

$\}$