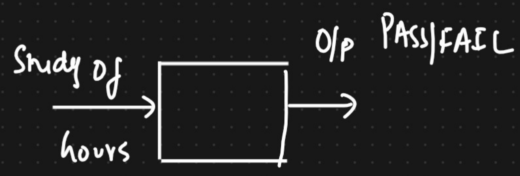
Logistic Regression (Binary classification)

★ Classification algorithm, not a regression problem

- O/P feature will be categorical data, and that will be binary.

Dataset

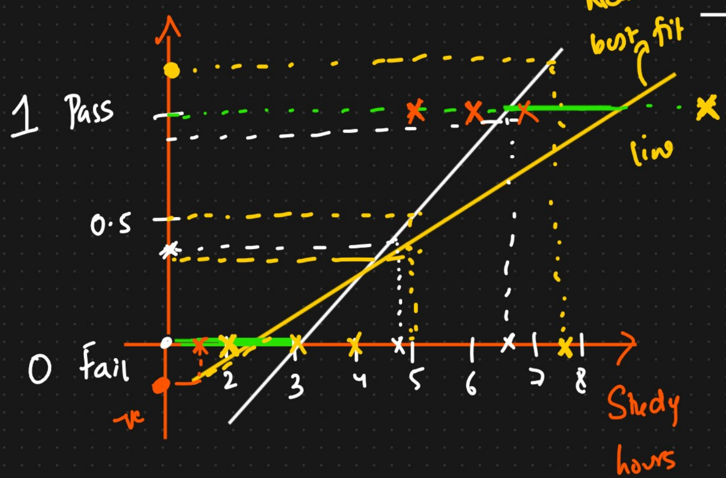
Study hours	Pass/Fail O/P {Binary categories}
2	Fail
3	Fail
4	Fail
5	Pass
6	Pass
7	Pass
12	Pass



Logistic Regression

0 to 1

Can Linear Regression Classification Problems.



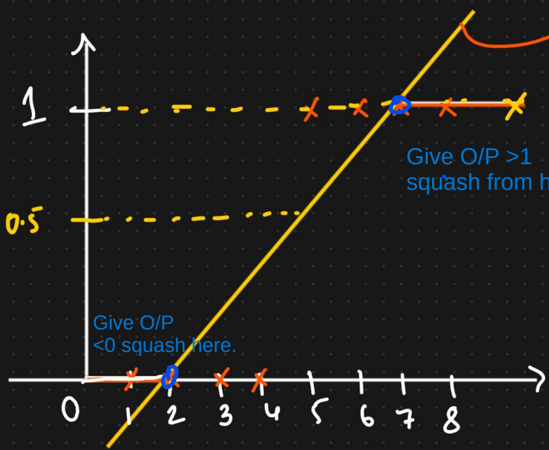
$\left\{ \begin{array}{l} \leq 0.5 \Rightarrow 0 \\ > 0.5 \Rightarrow 1 \end{array} \right\}$

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

Why we cannot use Linear Regression for Classification?

1. Outlier {Best fit line Change}
2. > 1 and < 0 {Squash line} And we can't squash the line in linear regression.

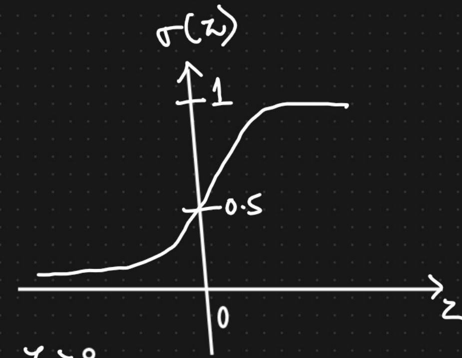
How Logistic Regression Solves Classification Problem



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

$\Downarrow$

Best fit line



$z \geq 0$

$\sigma(z) \geq 0.5$

Sigmoid Activation

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

Mathematical function used in machine learning, to map input values to a range between 0 and 1. It is commonly used for binary classification problems.

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

$\Uparrow$

$$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1)$$

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

$$= \sigma(z)$$

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

$\Rightarrow$ Logistic Regression hypothesis

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

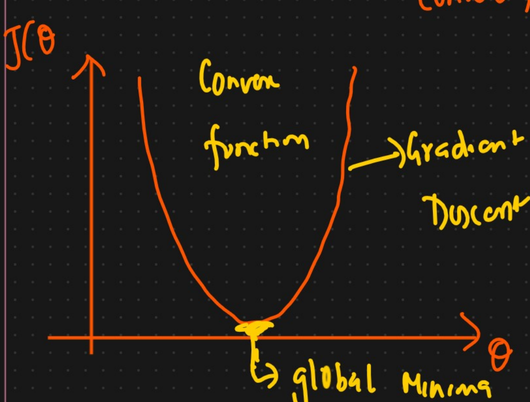
Linear Regression Cost function

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

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

$\Downarrow$

Convex function



Logistic Regression Cost function

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

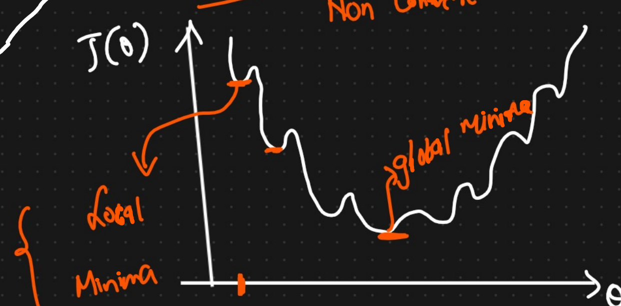
Non Convex function

$\Downarrow$

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

Sigmoid

Non Convex



- At local minima slope=0, so very difficult to come at global minima.

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

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

$\Downarrow$

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

$\Downarrow$

{ Log Loss }

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

- Mean Squared Error(MSE) cannot be used as cost function for Logistic Regression due to non-convex.

  So, we use log loss.

  $\downarrow$ Also Called

  Binary Cross-Entropy

- Log loss Cost function used in logistic regression to measure how well the model's predictions match the actual labels.
  - Log loss quantifies the difference between predicted probabilities and actual binary class labels (0 or 1).

Convex function

$\Downarrow$

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

By using log loss, it will be convex function.

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

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

Convergence Algorithm

Repeat

$j: 0$ and 1

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