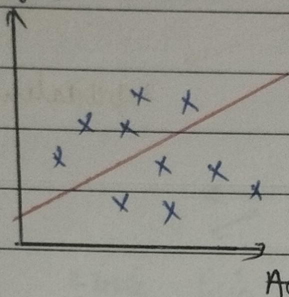
Logistic Regression :- Classification

Logistic Regression also called Logit regression is commonly used to estimate the probablity that an instance belongs to a particular class (eg., what is the probablity that this email is spam?)



— This is our best fit line

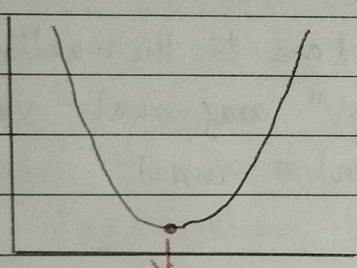
formula :- $h_\theta(x) = \theta_0 + \theta_1 x_1$

or

$\hat{y} = mx + c$

Cost function :- $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)})^2$ or $(\hat{y}_i - y)^2$

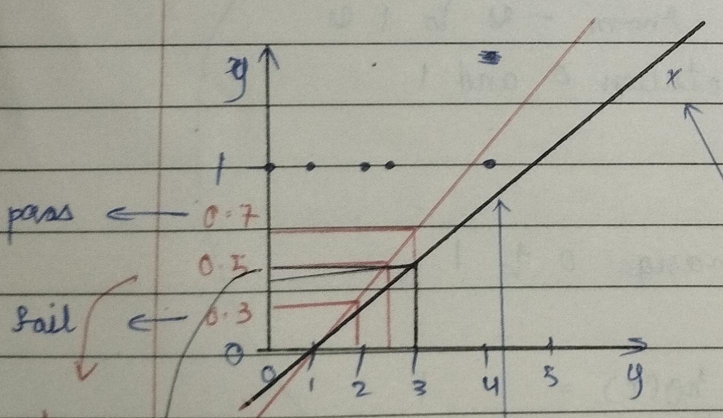
We, need to reduce our cost function $J(\theta_0, \theta_1)$ reaches global minima ⟶ which indicates best-fit line



global minima

Study hours	Pass/Fail
4	Pass 1
2	Fail 0
0	Fail 0
1	Fail 0
2.5	Pass 1

Logistic Regression ⟶ Binary Classification



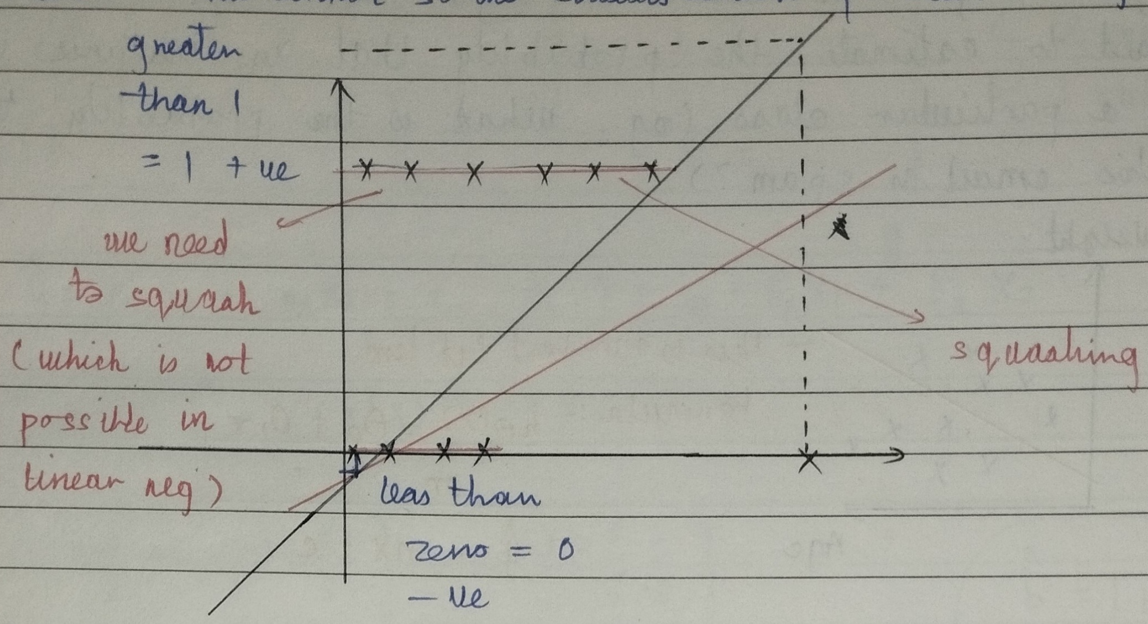
∴ Can we able to solve ~~linear reg~~ classification problem using linear regression

above 0.5 pass and below 0.5 fail

this indicates fail

next best fit line will be ⟵ If I have one outlier then the best fit line will change its slope and intercept to cover most of the data points

1. Outlier :- We cannot solve Classification problem using Regression



1. Outlier
2. Unable to squash >1 or <0
   ↳ in linear regression

In logistic regression, squashing refers to the action of the sigmoid function "compressing" any real-valued input into a fixed range between 0 and 1

$z = b_0 + b_1x_1 + b_2x_2 + \dots\dots + b_nx_n \Rightarrow h_\theta(x) = \theta_0 + \theta_1x_1$

- z :- can be any real number, from $-\infty$ to $+\infty$
- But probabilities must be between 0 and 1

The solution - sigmoid
"squashes" any z into the range 0 to 1

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

Hypothesis of Logistic Regression $\Rightarrow$ $\boxed{h_\theta(x) = \dfrac{1}{1+e^{-(\theta_0+\theta_1x_1)}}}$

Hypothesis of Logistic Regression

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

· Why we use sigmoid function

→ Linear Regression $\hat{y} = mx + c$

$= 2(2) + 2$

$= 4 + 2$

$\hat{y} = 6$

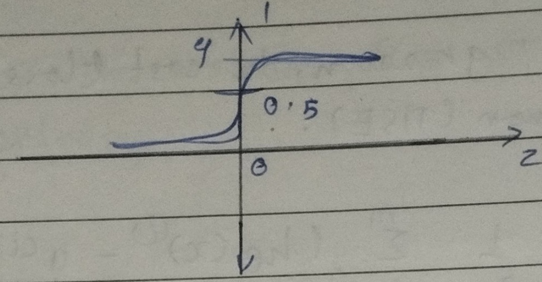
since this value is not in the range between 0 to 1 :- which is typically required in classification for this reason we use sigmoid function which squashes this values between the range 0 to 1

Sigmoid function takes any real-valued number (which could be output from a linear equation like $y = mx + c$) and "squashes" it into the range between 0 and 1. This makes the output suitable for representing probablities in classification tasks, ensuring the predictions always stay within meaningful probability limits

Sigmoid - Logistic Regression

Sigmoid

$$g(z) = \frac{1}{1+e^{-z}}$$



if $z \geq 0 \longrightarrow$ $z \leq 0$

$g(z) \geq 0.5 \longrightarrow g(z) \leq 0.5$

* Cost - function

- Linear Regression

$\Rightarrow J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)})^2$:- m represents total no of (training examples) or data points in your dataset

- Logistic Regression

$\Rightarrow J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)})^2$ Eg 67 rows with 60 columns for input & 5 for output $\rightarrow$ m is 67 (rows)

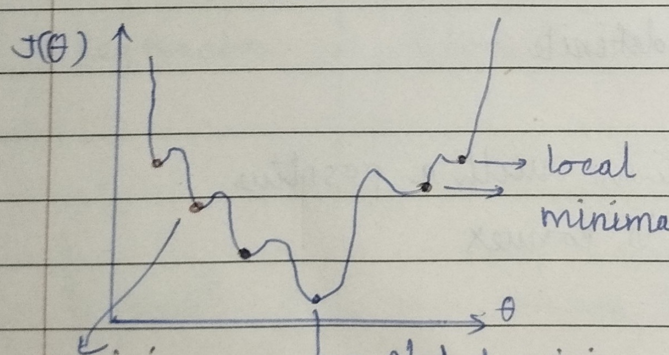
m = number of rows

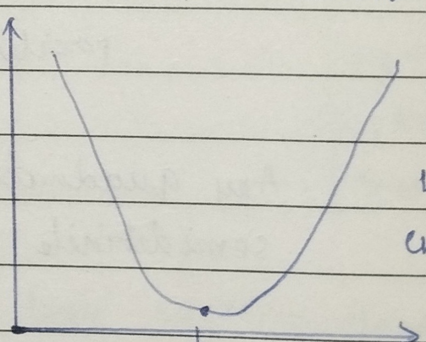
$h_\theta(x) = \frac{1}{1+e^{-(\theta^T x)}}$ } This is a non-convex function

$\rightarrow$ sigmoid function

$\rightarrow$ Linear Regression $\rightarrow$ Convex Function $\downarrow$ cost function used to create convex function

Non-Convex





local minima where slope = 0

(if any value comes over local minima $\longrightarrow$ it will be stuck their only because the slope value is 0 and because of this is couldn't able to reach global minima

Page No.
Date

Convex Cost function - Linear Regression

- Cost Function:

For linear regression, the cost (loss) function is the Mean Squared Error (MSE):

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

where $h_\theta(x) = \theta^T x$

- Why convex:
- This function is a quadratic function with respect to the parameter (θ)
- Quadratic functions always have a single global minima and no local minima
- Graphically, it looks like a "bowl" - any optimization algorithm (like gradient descent) will always move toward that unique lowest point, no matter where it starts

- Reason: squaring any difference always yields a parabolic surface, which is the mathematical definition of a convex function

The Hessian matrix is $\frac{1}{m} X^T X$, which is

positive semidefinite

↓

Any quadratic function with a positive semidefinite Hessian is convex

Convex cost function - Logistic Regression

- Cost Function:

For logistic regression, the cost function used is called log loss or binary cross - entropy

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

- Why Convex
  - Despite involving exponentials and logarithms because of the sigmoid function, the cost function is also convex with respect to the model parameters (θ)
  - Convexity means there is only one global minima, and gradient descent can reliably find the optimal parameters
  - The function's surface does not have local minima or maxima — only a single bowl - shaped minimum.

Model	Typical cost Function	Convexity Type	Why Convex
Linear Regression	Mean Squared Error (MSE)	Convex	Quadratic, single global minimum
Logistic Regression	Binary cross - entropy (Log loss)	Convex	log - sum exponential form, convex

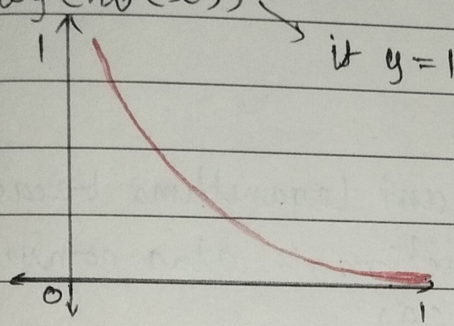
In short: Both are convex because their surface have only one minimum point, making optimization straightforward and reliable.

Cost function - Logistic Regression

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

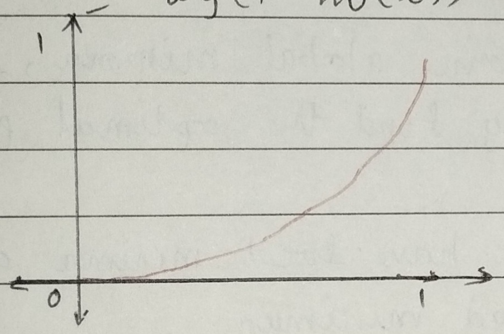
if $y=1$: cost is $-\log(h_\theta(x))$
if $y=0$: cost is $-\log(1-h_\theta(x))$

$-\log(h_\theta(x))$ → if $y=1$



{ if $y=1$ and $h_\theta(x)=1$ cost $=0$ }

$-\log(1-h_\theta(x))$ → if $y=0$



{ if $y=0$ and $h_\theta(x)=0$ cost $=0$ }

Convergence Theorem or Repeat

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

This tells when we will reach global minima