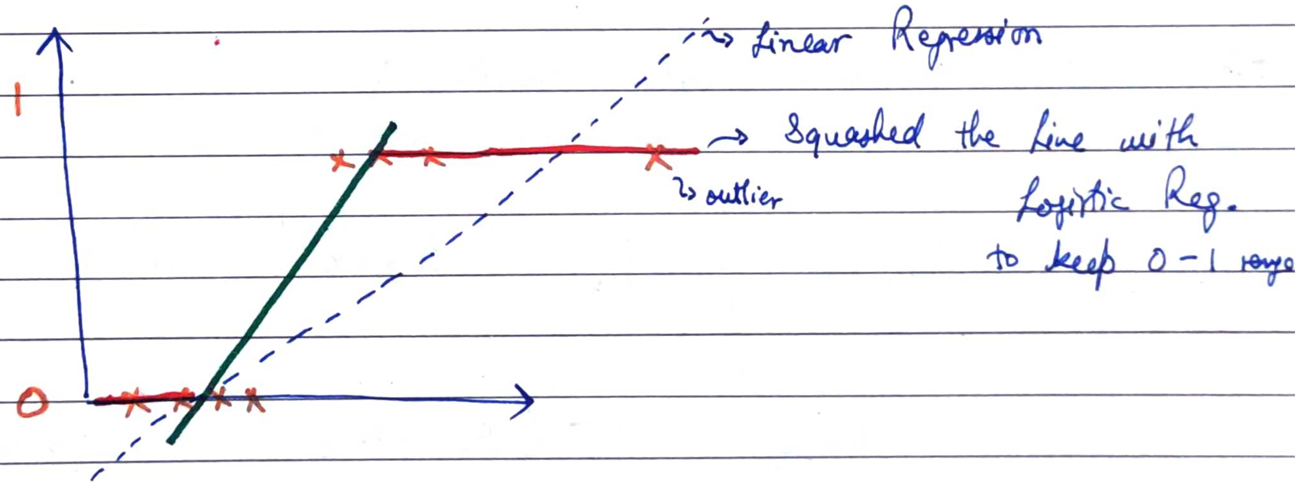
LOGISTIC REGRESSION

Used for :- Classification Problem

Since Linear Regression is prone to outliers, it changes the result in classification problems. That's why we need other Algorithms.

- Note :- Outlier is important & shouldn't be handled everywhere

Here we don't keep on extending Best fit line, rather we **SQUASH** (cut) it.



*** Squashing is Done by :- Sigmoid Activation

$h\theta(x) = \theta_0 + \theta_1 x$ (Best fit line) on top of BF line, we apply

↳ Sigmoid Activation function

↳ Squashing happen ⇒ O/P ⇒ 0 to 1

1. First, we create Best fit line

$$z = h\theta(x) = \theta_0 + \theta_1 x$$

2. Apply Sigmoid function (Squash)

$$\Rightarrow \frac{1}{1+e^{-z}} \Rightarrow \underline{0 \text{ to } 1}$$

$z = \theta_0 + \theta_1 x$

→ Linear Regression Cost Function :

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

→ Logistic Regression Cost Function :

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

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

$\sigma \rightarrow$ Sigmoid Activation

$$= \sigma(z)$$

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

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

$$\boxed{h\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}}$$

- Output of this will always between 0 & 1
  - ↳ we use threshold!

- Threshold segregates the output in this way.

  if output is $\leq 0.5 = 0$

  $> 0.5 = 1$

  So any output is 0.35, it becomes 0

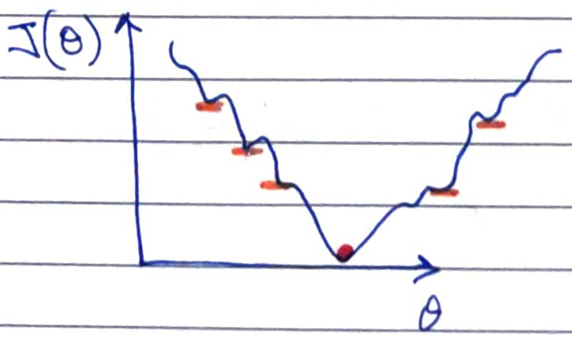
  — — is 0.8, it becomes 1

! Problem :-

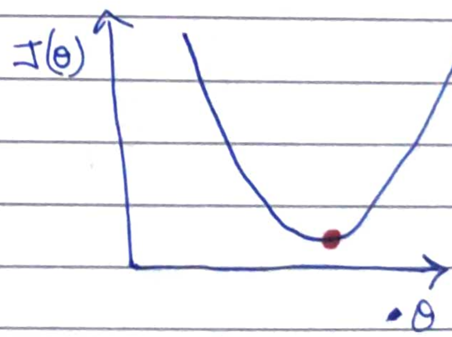
- Since Cost function lead to Convex function 

  ↳ Applying Sigmoid creates Non-Convex function!

Non Convex Function



Convex Function



How to fix? ... → Change Cost function

* Log Loss Cost Function

(we reduce value of cost function)

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

$y \rightarrow$ Truth value

*

$$\boxed{\text{Cost}\left(h\theta(x)^{(i)}, y^{(i)}\right) = -y \log\left(h\theta(x)\right) - (1-y)\log\left(1-h\theta(x)\right)}$$

$\downarrow$

Never Get Local Minima

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

$\Rightarrow$ Convergence Algorithm

(we repeat all the Convergence)

Repeat Convergence

$J = 0 \,\&\, 1$

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

Recap →

→ We can't solve Classification Problem with Logistic Regression because of 2 reasons :-

1. Outlier change Best fit line

→ So we try to Squash Best fit line so it never goes beyond 0 & 1

→ Squashing is done by Sigmoid Activation function.

→ Output of Sigmoid is b/w 0 to 1

→ Problem that arise is, Sigmoid creates a Non-Convex function

→ Log loss Cost function is used to fix Convex function

* PERFORMANCE METRICS

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F - Beta Score

1. Confusion Matrix

We Create 2×2 Matrix to find Right output Count

	Actual y: 1	0
Predict $\hat{y}$: 1	3	2
0	1	1

Dataset

f_1	f_2	y o/p	$\hat{y}$ Prediction	
–	–	0	1	×
–	–	1	1	✓
–	–	0	0	✓
–	–	1	1	✓
–	–	1	1	✓
–	–	0	1	×
–	–	1	0	×

→ Confusion Matrix

Actual

	1	0
1	True Positive	False Positive
0	False Negative	True Negative

Predict

Positive = 1
Negative = 0
True = Match
False = No Match

2. Accuracy

TP & TN are right Predictions.

$$\boxed{Acc = \frac{TP + TN}{TP + FP + FN + TN}}$$

$$\text{eg.} = \frac{3+1}{3+2+1+1} = \frac{4}{7} = 57\% \text{ Accuracy}$$

3. Precision

Let say we have Datset of Binary Classification

$$1000 \text{ Datapoints} \begin{cases} 900 \rightarrow 1 \\ 100 \rightarrow 0 \end{cases} \Bigg\} \text{ Imbalanced Dataset}$$

Now if I create a Dumb Model & ask for output as 1.

↳ I will get 90% Accuracy.

! It's BAD because I have an Imbalanced data & still am getting 90% Accuracy.

It means we can't rely on Accuracy only. X

That's why we Require PRECISION.

$$\boxed{\text{Precision} \Rightarrow \frac{TP}{TP+FP}}$$

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

Out of all the Actual values, how many are Correctly Predicted.

We focus on FALSE POSITIVE & we try to Reduce it ↓

Q In what Problem statement, False Positive is important?

A eg Our Mail is categorized as Spam & Ham.

0 1

if our Actual value is Spam & Predicted as Ham, it goes into False Negative. Not Harmful. 10

But if our Ham mail goes into Spam folder then its a problem, False Positive. 01

↳ That's why we reduce False Positive.

⇒ Now there are cases where False Negative is important

↓

4. Recall

$$\boxed{\text{Recall} \Rightarrow \frac{TP}{TP + FN}}$$

Out of all Predicted values, how many are correctly predicted

eg. Person went for diagnosis of diabetes.

He had diabetes but resulted as Non-Diabetic. Its a case of False Negative. We try to reduce it.

Let say Tomorrow stock market is going to crash.

2 Parties get affected in different way....

→ Consumer → FN ↓↓

→ Company → FP ↓↓

	1	0
1	TP	FP
0	FN	TN

0 = No crash
1 = Crash

Scenario 1,

Tomorrow Market fails (1) but Model say No (0).

So this impact in consumer money loss!

Scenario 2,

Tomorrow market No crash (0) but Model say Crash (1).

It won't affect consumers anyhow. They will withdraw.
But company doesn't want them to withdraw

~~Let say for company Model says No Crash but Crashes~~

In such case **FP** and **FN** both can be importants.

So we use... →

(5) F-Beta Score

$$\boxed{\frac{(1+\beta^2)\ \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}}$$

#1 If FP & FN are both important

$\beta = 1$

$$\text{F1 Score} = 2\,\frac{P \times R}{P + R}$$

#2 If FP is more important than FN

$\beta = 0.5$

$$\text{F 0.5 score} = \frac{1 + 0.25\ \ P \times R}{0.25 \times P + R}$$

#3 If FN is more important than FP

$\beta = 2$

$$\text{F2 score} = \frac{1+4\ \ P \times R}{4 \times P + R}$$