16-OCT-2022

Logistic Regression

- It is used to solve classification problem.

eg IIT JEE

Study hours	Play hours	O/p (Pass or Fail)
1	8	Fail
2	7	Fail
3	7	Fail
6	3	Pass
outlier → 1	4	Pass

Classification into Pass or Fail

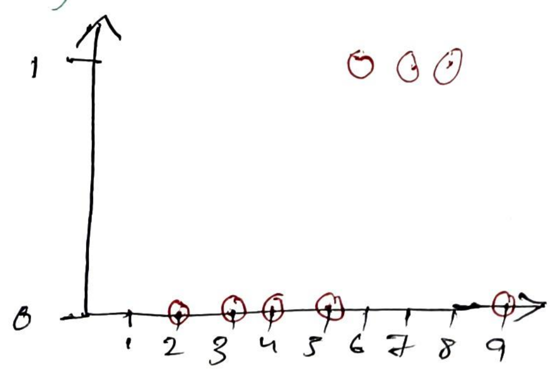
eg whether a person will buy or not some products during sale.

eg Dataset MPSC

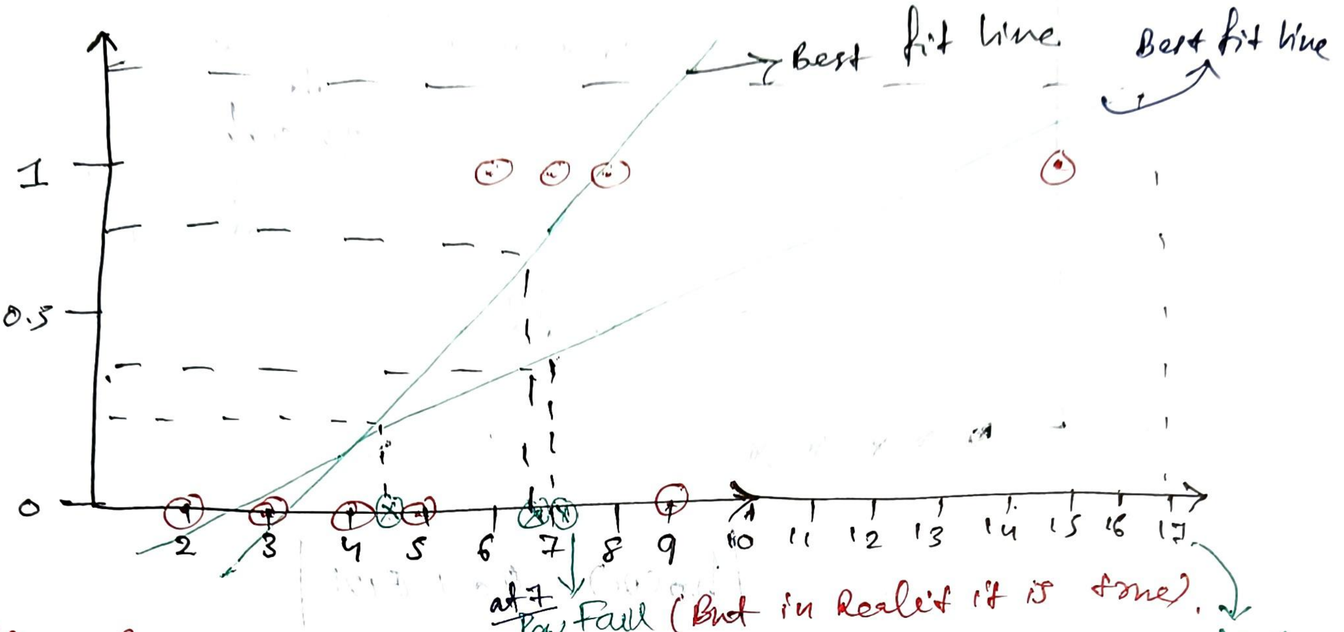
Study Hours	O/p (Pass / Fail)
2	Fail
3	Fail
4	Fail
5	Fail
6	Pass
7	Pass
8	Pass
outliers → 9	Fail

classification into Pass or Fail

Pass=1
Fail=0



Que Can we solve this problem statement using Regression?
Ans



Using Regression

0.5 → Threshold

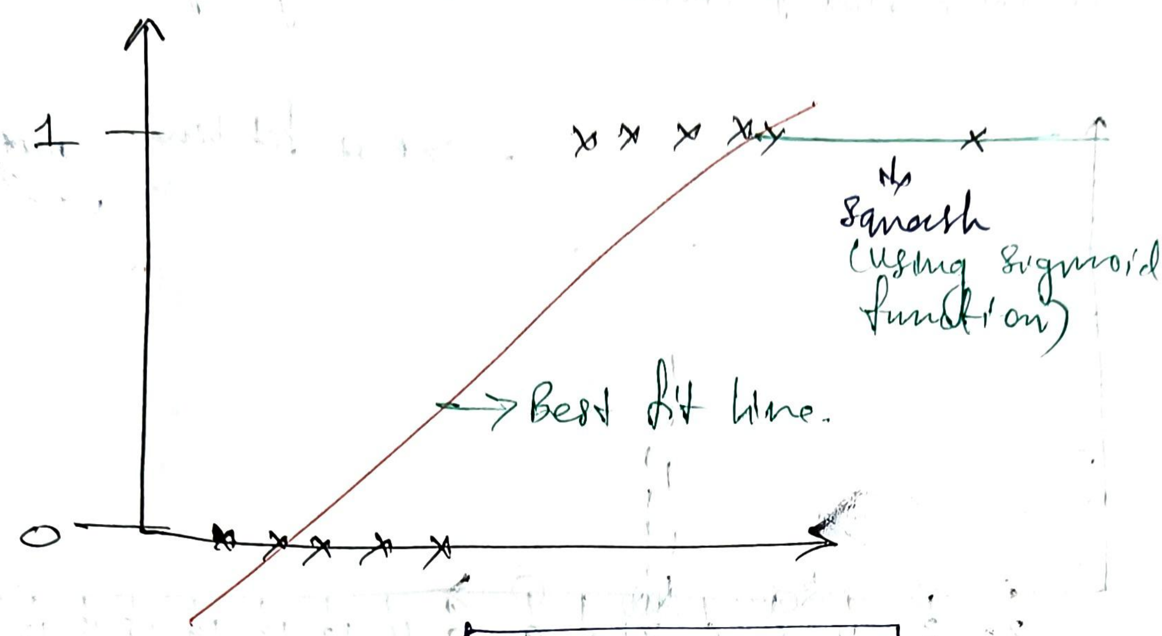
$y \leq 0.5 = 0$

$y > 0.5 = 1$

Why we use logistic Regression when we can solve classification problem using linear Regression?

Ans
- One to outlier best fit line gets change. and result will be wrong.
- We cannot remove outlier always.
- Threshold can't be changed, once fixed.
- In logistic, we squash the line. we will not change the line.

> Squash means "Kat Kar do"



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

Apply sigmoid activation and its value is will be ranging b/w 0 and 1.

Sigmoid function

$$\boxed{\text{Sigmoid function} = \frac{1}{1+e^{-z}}} \rightarrow \text{b/w } 0 \text{ to } 1.$$

- we can squash the line using sigmoid function.

Linear Regression Cost function

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

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

→ Gradient Descent
convex function
One Global Minima.

Logistic Regression Cost function

Steps!

① Create best fit line
② Apply squashing using sigmoid activation

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

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

$\downarrow$

Sigmoid Activation

Let $z = \theta_0 + \theta_1 x$

$h_\theta(x) = \sigma(z)$

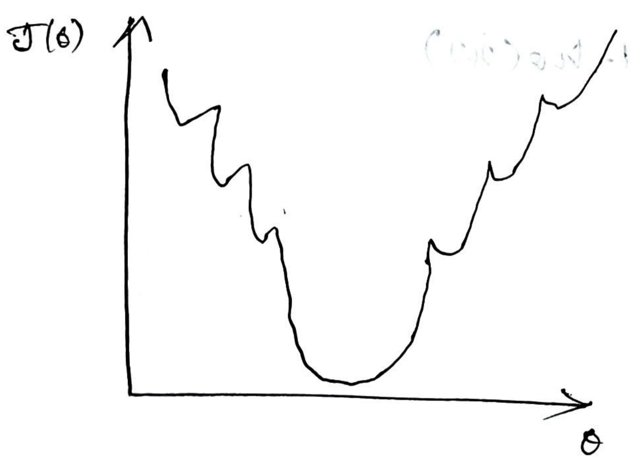
$h_\theta(x) = \dfrac{1}{1+e^{-z}}$

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

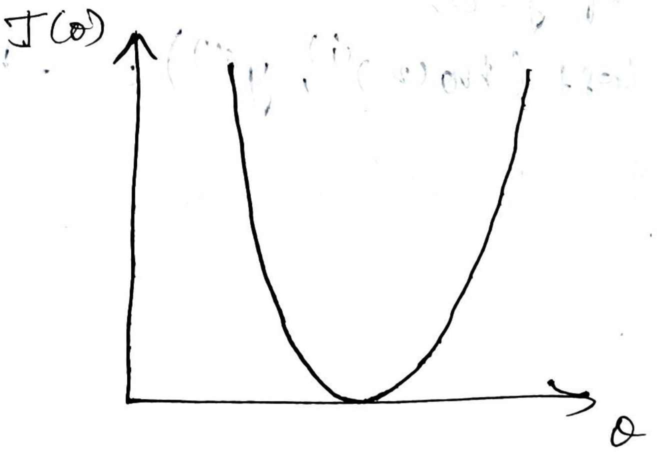
One problem!

- After applying sigmoid function, cost function will become non-convex function and have a chances to get local minima.

Non-convex function



convex function



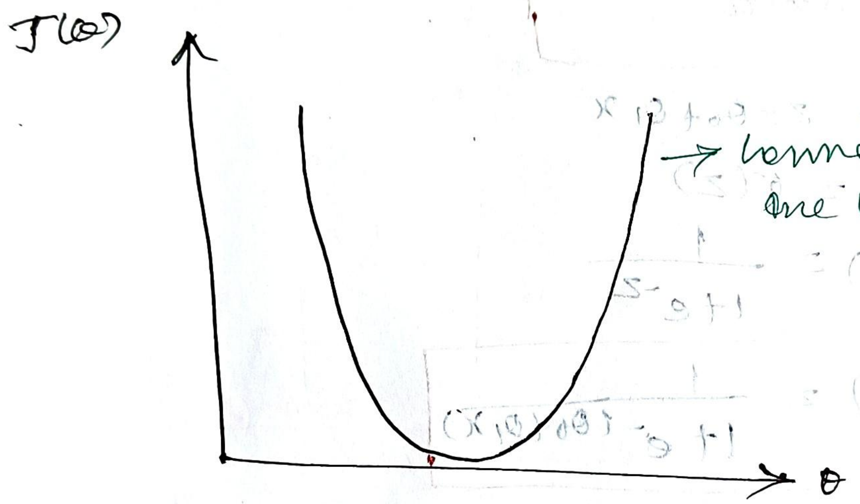
[Note: change the cost function to solve the convexity problem.]

Log loss cost function (No convexity issue)

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

convex function

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



→ convex function
one global minima.

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

if $y=1$,

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

if $y=0$,

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

y = Truth value.

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

Repeat convergence,

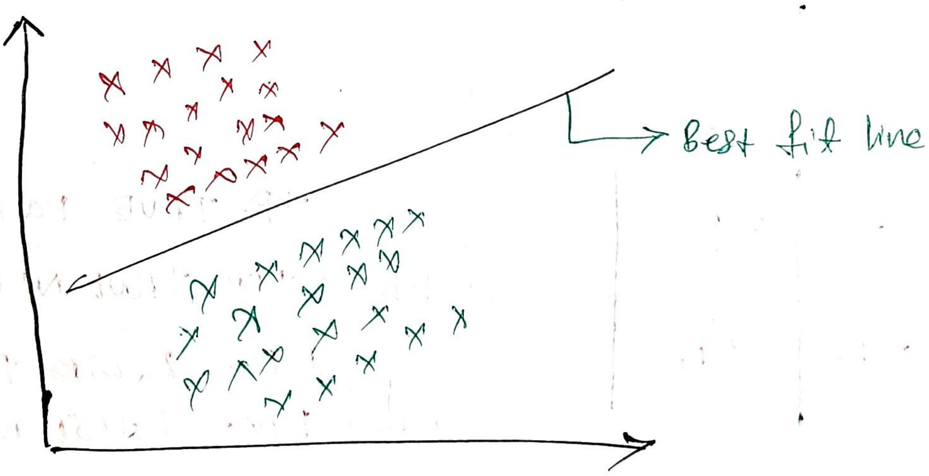
{

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_j}$$, for $\theta_0 = 0$ and $\theta_1 = 0$

}

→ By default take threshold = 0.5.
Using ROC and AUC curve, we can define threshold.

Performance Metrics



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

Dataset

Model Prediction

f_1	f_2	y o/p	$\hat{y}$ (Prediction after training)
—	—	0	1
—	—	1	1
—	—	0	0
—	—	1	1
—	—	1	1
—	—	0	1
—	—	1	0

Confusion Matrix

← Actual value (y)

Predicted value ($\hat{y}$) ↑	1	0
1	1, 2, 3, (3)	1, 2 (2)
0	1 (1)	1 (1)

1 → correct Prediction

0 → wrong Prediction

← Actual (y)

Predicted ($\hat{y}$) ↑	1	0
1	TP	FP
0	FN	TN

Correct match
- TP → TRUE POSITIVE
- TN → TRUE NEGATIVE

Wrong match
- FP → FALSE POSITIVE
- FN → FALSE NEGATIVE

Accuracy

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

eg $Acc = \frac{3+1}{3+2+1+1} = \frac{4}{7} = 57\%$

Dataset → binary classification
↓
1000 datapoints
output

900	100
(1)	(0)

→ Imbalanced Dataset

Dumb Model → 1 → will get 90% accuracy.

→ If the accuracy is 90%, it is not sufficient. Model is not good. To overcome this problem we can use recall and precision.

Precision

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

← Actual

Predicted ↑	1	0
1	TP	FP
0	FN	TN

→ Out of all the actual values, how many are correctly predicted.

→ Our aim is to reduce false positive (FP).

eg ~~when FP is important~~

① Mail → Spam or ~~not Ham~~ Harm

Actual → Spam and Predicted → Spam → Good.

Actual → Spam and Predicted → Harm → not Good

Actual → Harm and ~~Sp Actual~~ Actual → spam → Critical Problem

↓

Focus on reducing FP.

② Diabetes or not diabetes.

Actual → Diabetes and Predicted Not diabetes → Critical

↓

Reduce FN

Recall

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

→ Out of all the predicted values, how many are corrected predicted.

→ our aim is to reduce FN.

eg Tommorow the stock market is going to crash.

Two points

- consumer → FN↓
- company → FP↓ (can take certain decisions)

← Actual

	1	0
1	TP	FP
0	FN	TN

↑ Predicted

FP → company (action: sell shares at discounted price, offer etc)

FN ↓ consumer

F-Beta Score

$$\boxed{\text{F-Beta Score} = (1+\beta^2)\frac{\text{Presion} * \text{Recall}}{(\beta^2 \text{Precision} + \text{Recall})}}$$

① if FP and FN, both are important,

$\beta = 1$

$$\boxed{F1 \text{ Score} = \frac{(1+1)\ P \times R}{(1^2)P + R}} = \frac{2PR}{P+R}$$

② if FP is more important than FN,

$\beta = 0.5$

$$\boxed{F\text{-}0.5 \text{ Score} = \frac{(1+0.25)\ P \times R}{(0.25)P + R}}$$

③ If FP is less important than FN,

$\beta = 2$

$$\boxed{F\text{-}2 \text{ Score} = \frac{(1+4)\ P \times R}{4P + R}}$$