+ Hare Krishna +

Logistic Regression:- (classification problem)

IIT JEE Example

Study	Play hrs	o/p (Pass/Fail)
1	8	Fail
2	7	Fail
3	7	Fail
6	3	Pass
7	$\boxed{2}$	Pass
6	$\boxed{4}$	$\boxed{\text{Pass}}$
5	$\boxed{3}$	$\boxed{\text{Pass}}$

$\boxed{}$ → outliers.

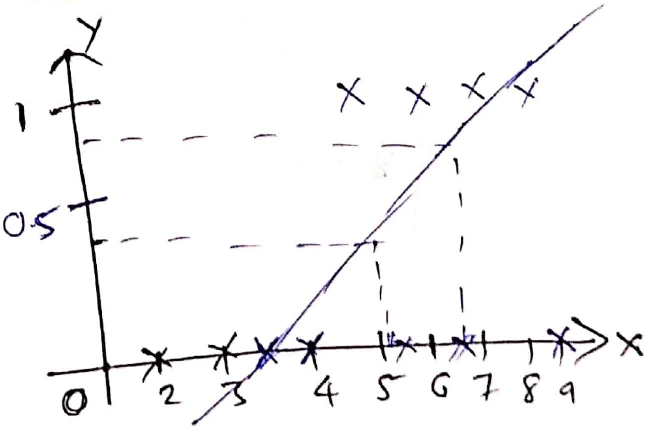
Data Set

UPSC

1 ⇒ Pass
0 ⇒ fail

Study hours	o/p (Pass/Fail)
2	Fail
3	Fail
4	Fail
5	Pass
6	Pass
7	Pass
8	Pass
9.	fail

① Can we Slove this Program Using regression.



Regression $0.5 \Rightarrow$ Threshold

$\boxed{y \le 0.5 = 0}$

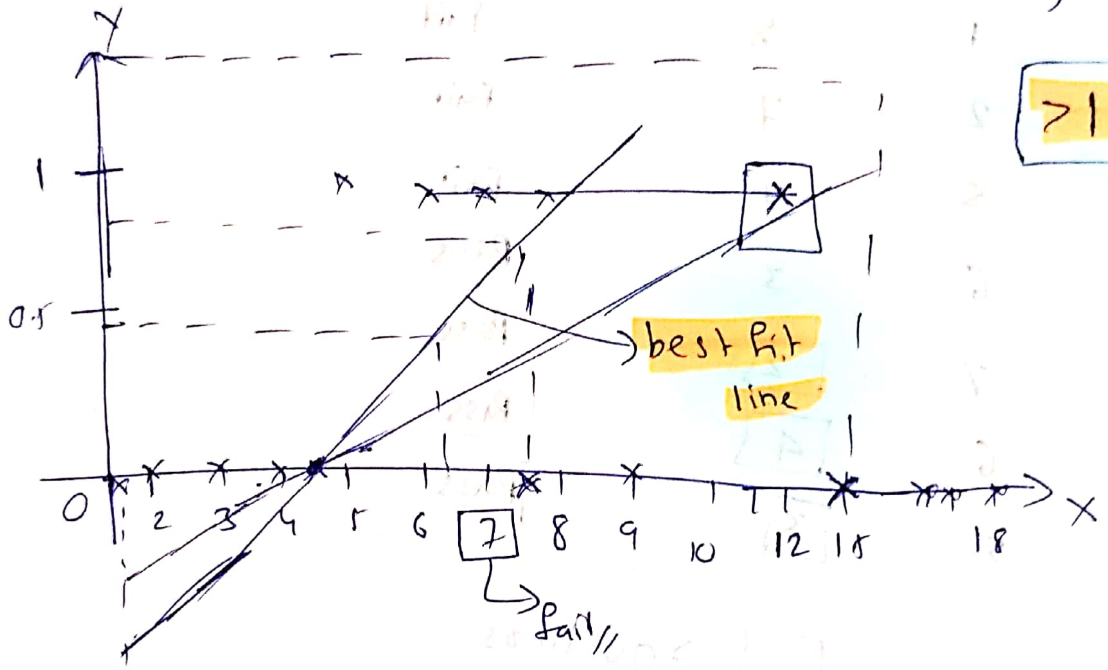
$\boxed{y > 0.5 = 1}$

Regression

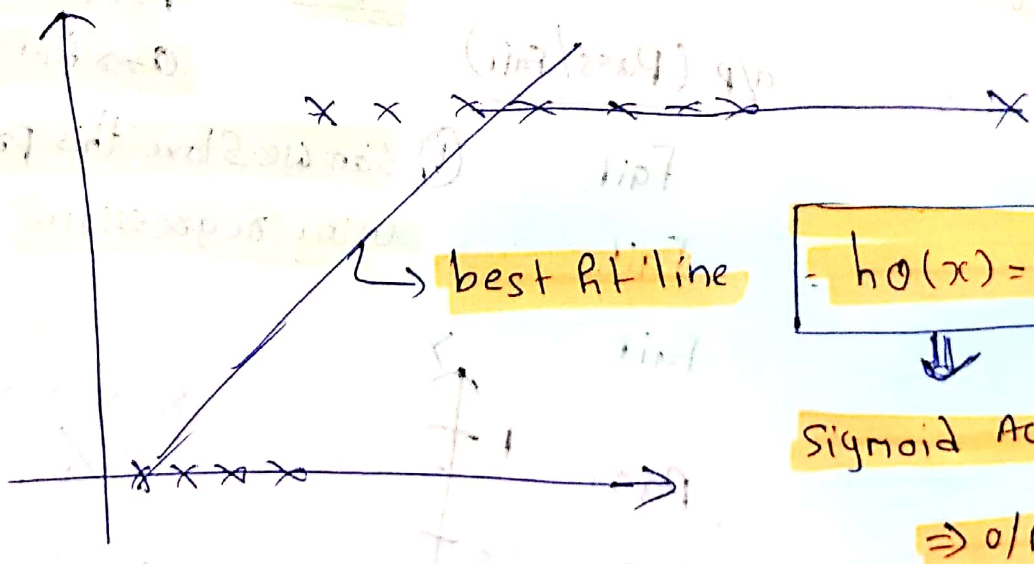
$y \le 0.5 = 0$

$y > 0.5 = 1$

$\boxed{>1 \ \& <0}$



Sigmoid Activation



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

$\Downarrow$

Sigmoid Activation

$\Rightarrow$ o/p $= 0$ to 1

① $Z = h_\theta(x) = \theta_0 + \theta_1 x$

Sigmoid fn $= \boxed{\frac{1}{1+e^{-z}}} \Rightarrow$ 0 to 1

$Z = \theta_0 + \theta_1 x$

Linear Regression Cost function

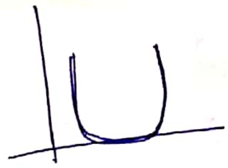
$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2$$

MSE

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

$\boxed{\text{Convex Function}}$

1 global minima



① Create a best fix line

② Squashing → Sigmoid function

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

$\boxed{Z = \theta_0 + \theta_1 x}$

Logistic Regression Cost function:-

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

best fit line.

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

Sigmoid Activation

$= \sigma(z)$

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

$$= \sigma(z)$$

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

$$\boxed{z = \theta_0 + \theta_1 x}$$

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

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}} \Rightarrow 0 \text{ to } 1$$

⇒ out will be b/w (0 to 1)

$\leq 0.5 \Rightarrow 0 \Rightarrow$ fail

$> 0.5 \Rightarrow 1 \Rightarrow$ pass

0 ⇒ fail

1 ⇒ Pass

$\boxed{\text{Threshold} - 0.5}$

let

⇒ 0.35 ⇒ 0

0.25 ⇒ 0

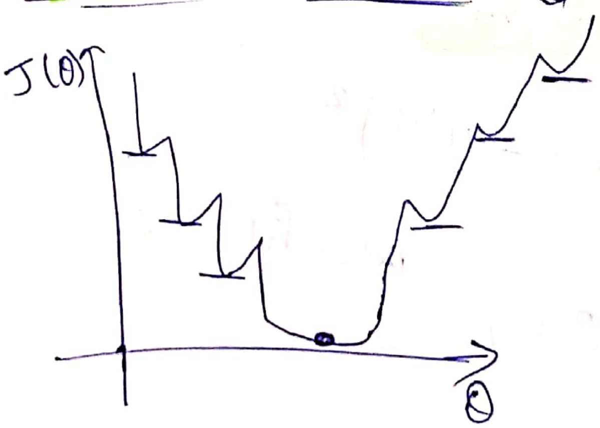
0.95 ⇒ 1

0.5 ⇒ 0

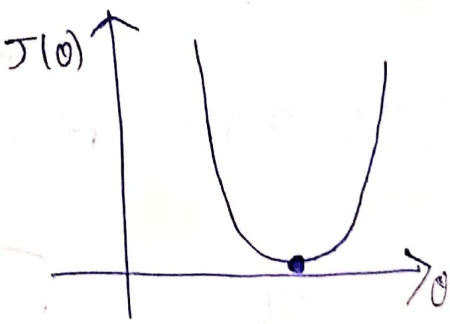
0.54 ⇒ 1

1 ⇒ 0.7

Non - convex function

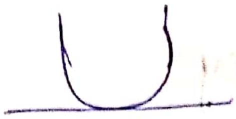


Convex function



log loss Cost Function

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



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

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

→ convex function

→ Never get local minima

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

Converging algorithm

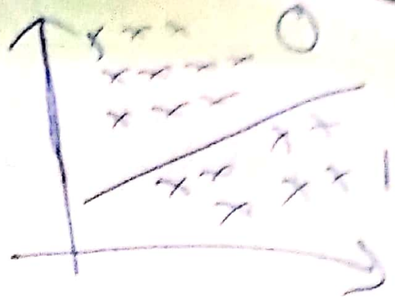
Repeat converging

$j = 0$ and 1

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

$\boxed{\text{Threshold} = 0.5}$

Performance metrics



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

Data set

f_1	f_2	o/p (y)	$\hat{y}$
–	–	0	1
–	–	1	1
–	–	0	1
–	–	1	0
–	–	0	1
–	–	1	1
–	–	0	0

$\hat{y}$ → model predict

$\hat{y} \Rightarrow y_{cap}$

Confusion matrix:

y ⇒ Actual value

	1	0
1	2	3
0	1	1

⇓ Predicted

→ Confusion matrix

⇒ Actual

	1	0
1	TP	FP
0	FP	TN

⇓ Predict

TP ⇒ True prediction (Positive)

TN ⇒ True negative

FP ⇒ False Prediction (Positive)

FN ⇒ False Negative

Acctual =

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} = \frac{2+1}{2+3+1+1} = \frac{3}{7}$$

Dataset: Binary Classification

→ 1000 data points → 900 → 1, → 100 → 0 } Imbalanced Dataset

Accuracy

Dum model → 1 ⇒ 90% Accuracy ⇒ → × sufficient

(+) Precision = $\boxed{\dfrac{TP}{TP+FP}}$

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

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

Problem statement:

Mail → Spam or Ham

i am getting a mail which is spam (1) and model predicted as spam(1) then it is TP (true possitive)

while ⇒ ~~may~~ i am getting an important (1) and model is predicting as spam (1) then it is FP ⇒ (flase possitive)

so to reduce the false possitive

so we have to focus on False Possitive,,

Model → diabetes or not diabetes

Recall: $\boxed{\dfrac{TP}{TP+FN}}$ ⇒ out of all the predicted values how many are correctly predicted.

ex:-

Tomorrow the stock market is going to crash

⇓

- → consumer → FN ↓↓
- → companies → FP ↓↓

	1	0 ⇒ Actual
1	TP	FP
0	FN	TN

⇓

predicted

then we use

(+) F - Beta Score:

$$\boxed{\frac{(1+\beta^2)\ Precision * Recall}{(\beta^2)\ precison + Recall}}$$

(1) If FP and FN are both important

$\beta = 1$

$$F1\ Score = 2\ \frac{P * R}{P + R}$$

Scanned with OKEN Scanner

(2) if FP is more important than FN

$\beta = 0.5$

$$F_{0.5} \text{ score} = \frac{(1+0.25)\ P*R}{(0.25)(P+R)}$$

(3) If FN >> FP

$$F_2 \text{ Score} = \frac{(1+4)\ P*R}{(4*P+R)}$$