ML Day-4

23/04/23

Logistic Regression :→ We use it to solve classification problem.

Suppose a person is trying to clear IIT JEE exam.

Dataset

Study (hrs)	Play (hrs)	O/P (Pass/Fail)
1	8	Fail
2	7	Fail
3	7	Fail
6	3	Pass
1	4	Pass

We are saying O/P on the basis of our observation.

→ outliers

→ Can we make the m/c's do the same thing? that by seeing the study hours it should be able to predict whether ~~model~~ he will pass or fail.

→ Logistic Regression solves classification problem like whether the person will pass or not? whether the is going to buy a product or not?

Example

Dataset: UPSC Exam

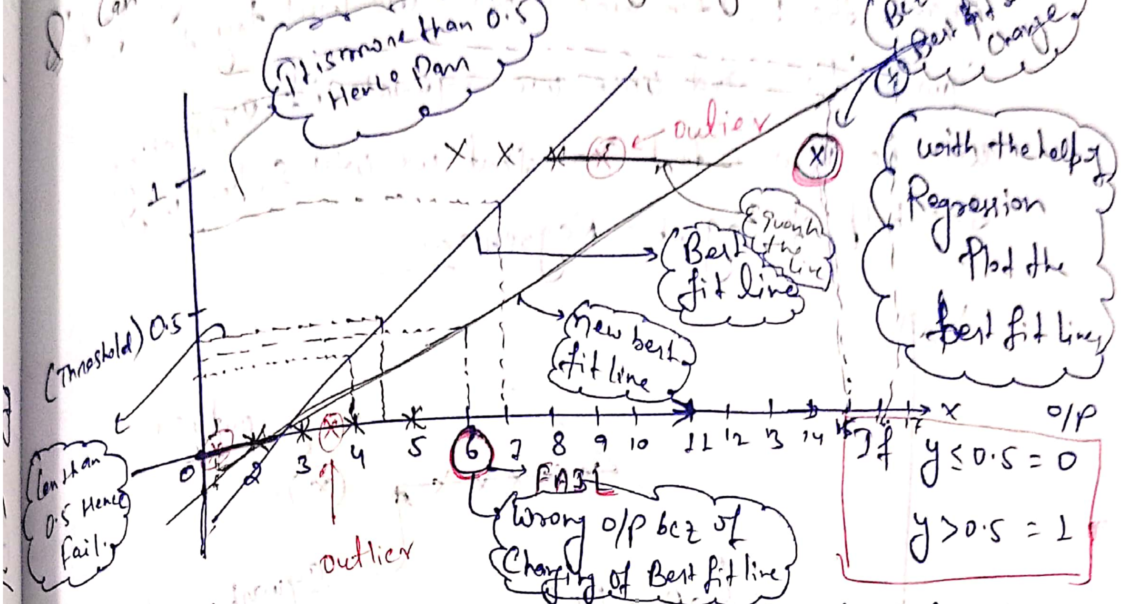
Study Hours	O/P (Pass/fail)
2	FAIL
3	FAIL
4	FAIL
5	FAIL
6	PASS
7	PASS
8	PASS
9	FAIL

→ outliers

1 0

(1) Can we solve this problem using Regression?

Q. Can we solve this problem using Regression?



So here we can see that we are able to solve the problem with the help of Linear Regression. then what is the need of Logistic Regression?

Ans

Problem 1:- Suppose one person studies for 15 hr. And when we try to plot the line on graph we can see that bcz of outlier best fit line will change and bcz of that it gives wrong o/p at different point.

when you check for 6 hrs it comes below 0.5 which means fail. Hence Regression doesn't work properly in this case.

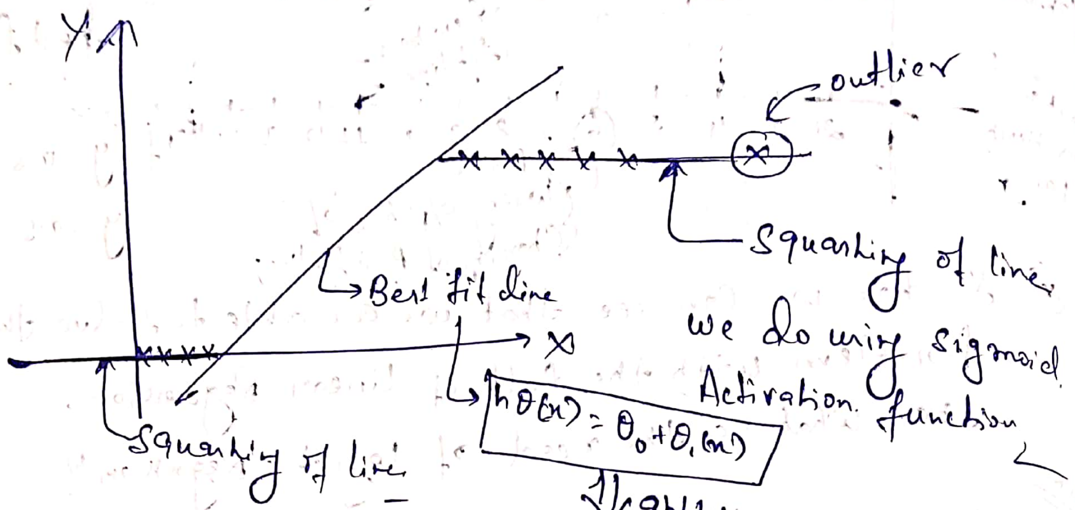
Problem 2:- for 17 hr it will go above 1

for 1 hr it will go below 0

so $\boxed{>1 \text{ and } <0 \text{ Can't be handled}}$

Hence, Rather than going above 1 and Below 0 we should focus on squash the line, and it is only possible bcz of logistic Regression.

(*) MAIN AIM :→ Because of outlier we are not going to change the Best fit line instead we will try to Squash it. and once we Squash it everything will be b/w 0 to 1 and that we will solve with the help of Logistic Regression.



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

⇓ apply

$\boxed{\text{Sigmoid Activation } f^n}$ ⇒ O/P = 0 to 1.

(*) Over the Best fit line which is $h\theta(x) = \theta_0 + \theta_1(x)$ we apply Sigmoid Activation function and with the help of that we squash the line since it's ~~also~~ O/P always lies b/w 0 to 1.

Steps:

(i) $z = h\theta(x) = \theta_0 + \theta_1 x$ i.e create a best fit line.

(ii) $\boxed{\text{Sigmoid function} = \dfrac{1}{1+e^{-z}}}$ ⇒ it lies b/w 0 to 1.

where, $(z = \theta_0 + \theta_1 x)$

i.e Squashing with the help of Sigmoid function.

Linear Regression Cost function

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

$\Downarrow$

MSE

$\Downarrow$

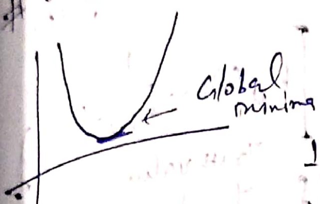
Convex function

$\Downarrow$

1 Global minima

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

Convex function



Logistic Regression Cost function

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

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

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

Sigmoid function $= \sigma(z) = \dfrac{1}{1+e^{-z}}$ o/p (0 to 1)

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

(*) Apply Sigmoid function over $h_\theta(x)$

O/P	
If it is 0.35	$\Rightarrow$ 0
0.25	$\Rightarrow$ 0
0.96	$\Rightarrow$ 1
0.5	$\Rightarrow$ 0 $\rightarrow \leq 0.5$
0.59	$\Rightarrow$ 1
0.75	$\Rightarrow$ 1 $\rightarrow > 0.5$

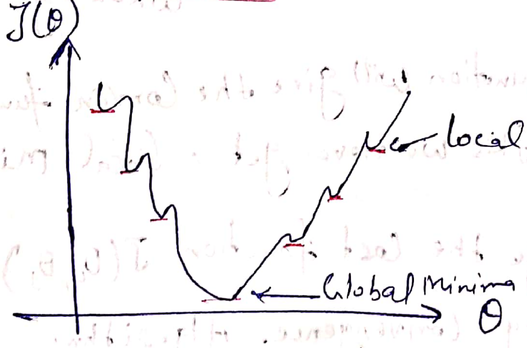
Threshold $\leq 0.5 \Rightarrow 0 \Rightarrow$ fail

$> 0.5 \Rightarrow 1 \Rightarrow$ Pass

(*) By doing this we solve outlier problem also by squashing it.

Problem with Sigmoid function :→

(*) The Problem with $\boxed{h_\theta(x) = \dfrac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$ is that it leads to non convex function.



(updation of coefficient may get stuck at local minima)

(*) A convex function has many local minima. So there are chances that it may get stuck at local minima and never reach to global minima.
Bcz of this we are not able to get efficient result in backend.

Q. How to fix the Problem of non Convex Problem?

Sol

Change the Cost function.

Hence instead of using Sigmoid function we use Some other cost function and i.e called Log loss Cost function, which creates a Convex function.

V.V.I

Log. Loss Cost function: → Here $\boxed{h_\theta(x) = \dfrac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$

Truth value

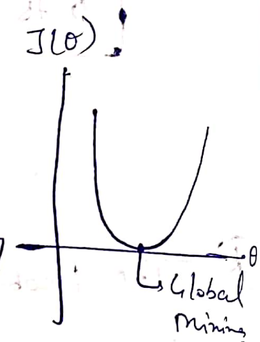
$$\text{Cost}(h_\theta(x)^i, y^i) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1-h_\theta(x)) & \text{if } y = 0 \end{cases} \Rightarrow$$

Aim is to reduce the cost f for that we use convergence Algo.

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

When $y = 1$

$$\boxed{\text{Cost}(h_\theta(x)^1) = -\log h_\theta(x)} \Rightarrow$$



If $y = 0$ $\boxed{\text{Cost}(h_\theta(x)^i) = -\log(1-h_\theta(x))}$

Threshold = 0.5

where y is Truth value

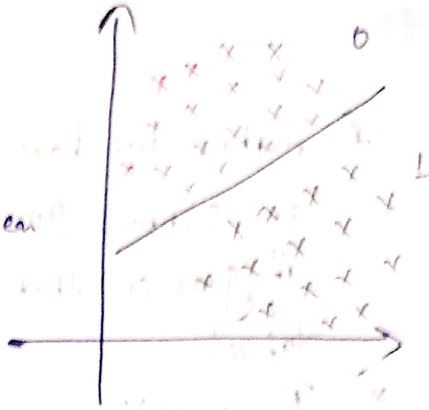
(*) And this cost function will give the Convex function, and through this we never get a local minima.

(*) We minimize the cost function $J(\theta_0, \theta_1)$ by changing θ_0, θ_1 using Convergence Algorithm.

Repeat Convergence $j = 0 \& 1$

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

Performance Metrics :→



(i) with the help of logistic Regression basically we are trying to draw a linear line which will divide the data points in two categories. i.e 0th and 1st Category.

(ii) Now to check whether the Model is performing good or not we use different techniques.

(i) Confusion Matrix.
(ii) Accuracy
(iii) Precision.
(iv) Recall
(v) F-Beta Score.

Dataset

f_1	f_2	O/P	$\hat{y}$	
–	–	0	1	← Wrong Predicted
–	–	1	1	Right Pred
–	–	0	0	Right Pred
–	–	1	1	Right Pred
–	–	1	1	Right Pred
–	–	0	1	← wrong Pred.
–	–	1	0	← Wrong Pred.

$\hat{y}$ ← Model Prediction

V.V.I

(i) Confusion Matrix

	Y (Actual value) 1	0
$\hat{y}$ (Predicted value) 1	3	2
0	1	1

⇓ Confusion Matrix

⇒

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

Right Prediction.

1,1 → True
0,0 → True
1 → Positive
0 → Negative

(*) $$\boxed{\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}} = \frac{3+1}{3+2+1+1} = \boxed{4/7} \approx \boxed{57\%}$$

(*) It means my model is able to make 57% accurate Prediction.

Need of Precision, Recall and F-Beta Score

- Suppose we have a dataset having 1000 data points, and suppose 900 times it gives, and only 100 times it gives 0. that means we have an Imbalance dataset.
- If we apply a Dumb Model over dataset, which gives O/P as only 1.. then the accuracy will be → 90%.
- Hence we can't be only dependent on Accuracy and that's we use Precision, Recall and F-Beta score.

DATASET → BINARY CLASSIFICATION

↳ 1000 datapoints → 900 ~~times~~ → 1 ; 100 ~~tim~~ → 0 } Imbalance Dataset

Dumb Model → 1 ⇒ 90% Accuracy → Accuracy ✗ Sufficient

↳ Hence only Accuracy is not sufficient enough to say that model is good or bad.

- And that's why we use Precision, Recall and F-Beta score

✱ Precision :→

Out of all the Actual values how many are correctly Predicted

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

Correctly Predicted (TP) ; all Actual values (TP+FP)

	Actual 1	Actual 0
Predicted 1	TP	FP
Predicted 0	FN	TN

Correctly Predicted 1 ; All Actual values

- This Scenario is called as Precision.
- In Case of precision we mainly focus on $\boxed{FP}$ ↓↓↓ and we try to reduce it.

Example:- Cases where False Positive (FP) is Very Very Important.

Problem statement :-

① Mail → Spam Or Ham (1 = Spam, 0 = Ham)

1 → spam
FP → 0 → ham

Predicted \ Actual	1	0
1	TP	FP ↓↓
0	FN	TN

TP → mail is "spam" and Predicted SPAM → Right Prediction

$\boxed{FP}$ → " " "ham" " " "SPAM" → Harmful, Wrong Prediction

FN → " " "SPAM" " " 'HAM' → not Harmful (Can be deleted)

TN → " " "HAM" " " "HAM" → Right Prediction

↓↓ $\boxed{FP}$ → The mail is a "HAM" but Predicted as "SPAM".
~~It Can~~ So user may miss the very important information and may create a big loss ...
Hence We need to try to reduce the $\boxed{FP}$ ↓↓↓.

② Model → (Diabetes ✓) (1) Or (No Diabetes ✗) (0)

~~TP → Person~~

Predicted \ Actual	1	0
1	TP	FP
0	FN ↓↓	TN

	Actual	Predicted	
TP	✓	✓	(Diabetes) Predicted
FP	✗	✓	→ Person doesn't have diabetes but Predicted diabetes. Person may go for second opinion. So not Harmful.
↓↓ $\boxed{FN}$	✓	✗	
TN	✗	✗	(No Diabetes)

FN → Here Person has diabetes but it Predicted No diabetes bcz of this after a period of time patient's health may degrade and He/she may die.
Hence it is Very Critical

⇐ Critical ; Not Critical

② **Recall** :→ Out of all the predicted values how many are correctly predicted.

$$\text{Recall} = \frac{\overbrace{TP}^{\text{Correctly Predicted}}}{\underbrace{TP+FN}_{\text{all Predicted values}}}$$

	Actual 1	Actual 0
Predicted 1	TP	FP
Predicted 0	$\boxed{FN}$ ↓↓↓	TN

Problem stmt' :→

Tommorow the stock Market is going to crash.

1 → Crash (✓)

0 → No crash (X)

	Actual	Predicted	
TP	✓	✓	
FP	X	✓	→ (for companies)
FN	✓	X	→ (for consumer)
TN	X	X	

	Actual 1	Actual 0
Predicted 1	TP	$\boxed{FP}$ ↓↓↓ → Reduce for Companies
Predicted 0	$\boxed{FN}$ ↓↓↓ → Reduce for Consumer	TN

Model for → Consumer → FN ↓↓↓

Model for → Companies → FP ↓↓↓

For Consumer

↓↓↓ $\boxed{FN}$ → Actual → Crashing; Predicted → No crash } Consumer money Gone ⇒ Huge loss for Consumer

⇓

Investors will take money and go.

FP → Actual → No crash; Predicted → Crash } will not Impact the consumer ⇒ They will just sell and go.

For Companies

↓↓↓ $\boxed{FP}$ → Actual → No crash; Predicted → Crash } Huge loss for Company ⇒ So that Company can take appropriate action, provide good offer, Sell the Share at discount

(*) In some scenarios FP and FN both can be important and we may need to reduce both. In that case we use F-Beta Score.

F-Beta Score :→

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

Case 1

(i) If FP and FN both are important

then $\boxed{\beta = 1}$

P: Precision
R = Recall

$$\boxed{F1\ Score = 2\left(\frac{P*R}{P+R}\right)}$$

Case 2

(ii) If FP is more important than FN

then, $\boxed{\beta = 0.5}$

$$\boxed{F\text{-}0.5\ Score = 1.25 \frac{(P*R)}{(0.25P+R)}}$$

Case 3

(iii) If FN is important than FP

then $\boxed{\beta = 2}$

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