Memo No. ________

Date / /

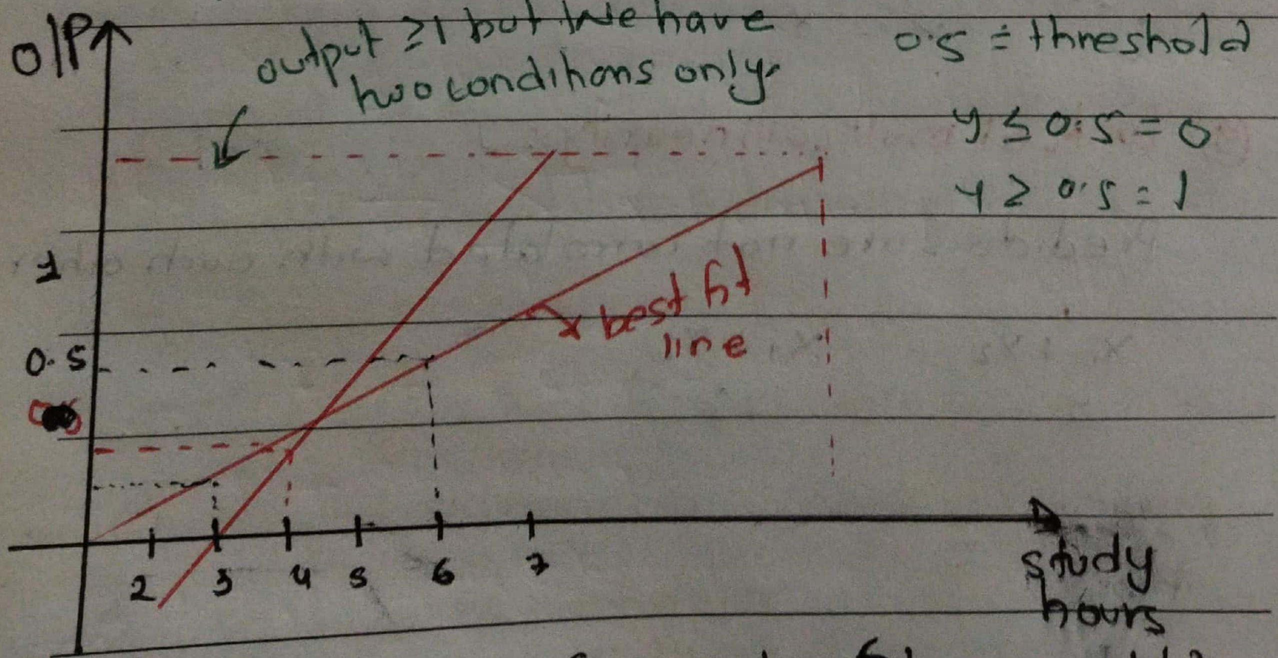
Logistic Regression:

- Classification problem

Eg: Dataset be:

	Study hour	Playhour	o/p
	2	8	fail
	3	7	fail
	6	3	pass
Outliers →	1	4	pass

Here, we cannot perform regression, we need to perform classification into pass / fail.



Even reading 6 hours, model is predicting fail.

Memo No. ____

Date / /

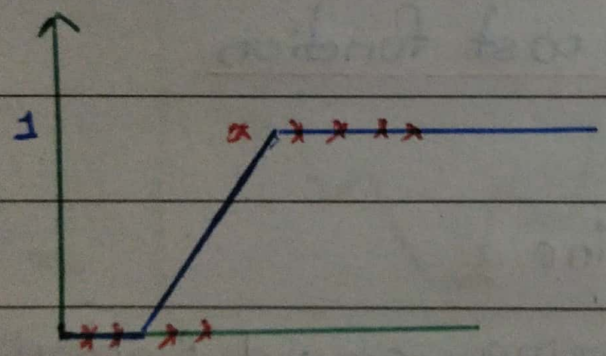
Why we use logistic regression when we can solve classification problem using linear regression?

→ Due to outliers best fit line gets changed and results will be wrong.

⇒ We cannot remove outliers always.

⇒ Threshold can't be changed, once fixed.

→ In logistic, we squash ('cut') the line, we will not change the line.



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

⇓

Sigmoid function ⇒ output will range betn 0 and 1.

Sigmoid Function:

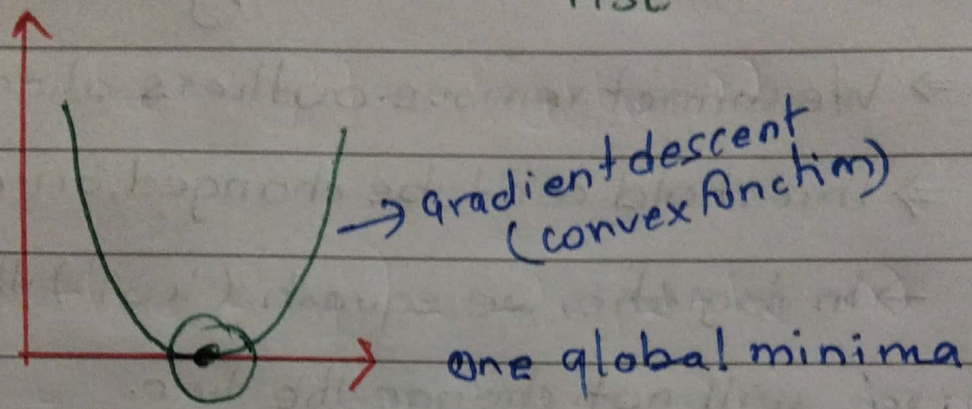
$$\boxed{\text{Sigmoid function} = \frac{1}{1+e^{-z}}} \rightarrow \text{bet}^n\ 0 \text{ to } 1$$

① Create a best fit line

② Squashing → sigmoid function

Linear Regression Cost function:

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



Logistic Regression cost function

steps:

1) create a best fit line

2) Apply squashing using sigmoid function.

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

sigmoid function

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

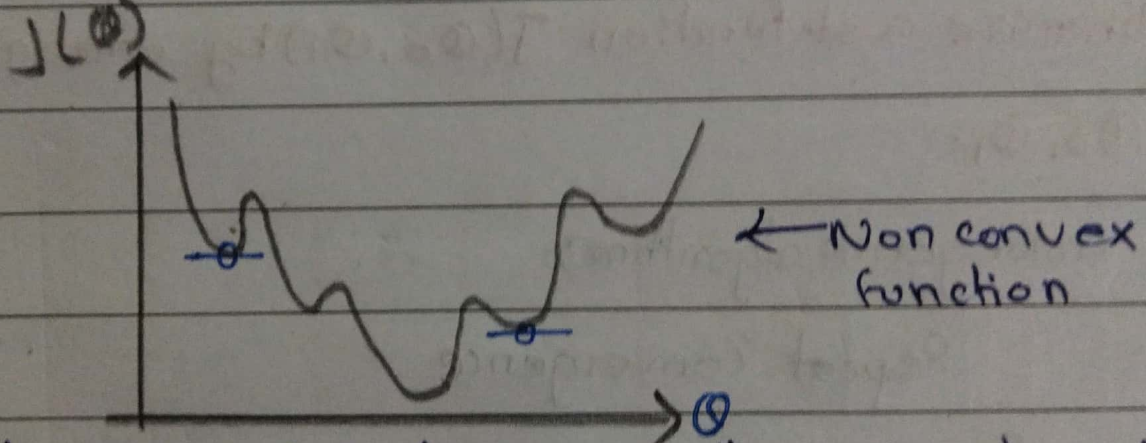
let $z = \theta_0 + \theta_1 x$

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

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

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

But after applying sigmoid function, cost function will become non convex function and have a chance to get local minima.



• change the cost function to solve the convexity problem.

Logloss function:

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

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

Convex function

This will never give local minima

Memo No.

Date / /

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

If $y=1$,

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

If $y=0$

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

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

Convergence algorithm:

Repeat Convergence

{

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

}

for $j=0$ and $j=1$

By default take threshold = 0.5

Using ROC and TOC, we can define threshold.

Performance Matrics

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

dataset

f_1	f_2	o/p	$\hat{y}$
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0

Confusion Matrix

$\hat{y}$ \ y	1	0
1	TP	FP
0	FN	TN

TP : True positive } correct match
TN : True negative }

FP : False positive } wrong match
FN : False negative }

Confusion matrix for the previous dataset:

	1	0 ← actual(y)
1	3	2
0	1	1

Accuracy

$$\text{Accuracy} = \frac{\overbrace{TP+TN}^{\text{correctly predicted}}}{\underbrace{TP+FP+FN+TN}_{\text{total dataset.}}}$$

Accuracy for above dataset $= \frac{3+1}{7} = 4/7 = 57\%$

Mo Tu We Th Fr Sa Su

Memo No.

Date / /

Imbalance dataset

1000 datapoint (Binary classification)

900(1) 100(0)

dumb model → If our model predict 1 everytime, we still get 90% accuracy.

So, 90% accuracy don't ensure the model goodness. To overcome this problem we use Recall and Precision.

Precision

we have to reduce false positive

$\hat{y}$ \ y	1	0
1	TP	FP
0	FN	TN

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

So, precision deals with correct true prediction among all true prediction.

Hence, precision help us to visualize the relability of the machine learning model in classifying the model as "Positive."

Memo No.
Date / /

Example:

1) Mail → spam or not (ham)

TP: actual → spam; predicted → spam ⇒ Good

FP: Actual → ham; Predicted → spam ⇒ bad.

~~It is true positive rate.~~

Recall

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

Recall is calculated as the ratio betⁿ the number of positive samples correctly classified as positive to the total number of positive samples.

↑ the Recall ⇓ more positive sample detected.

We don't care about negative sample are correctly classified or not.

Memo No.

Date / /

Difference betn Precision and Recall in ML:

Precision	Recall
→ Helps to measure the ability to classify positive sample in the model.	it helps to measure how many positive sample were correctly classified by model

Using situation

a) If there is a requirement of classifying all possitive as well as negative sample as positive, whethere they are classified correctly or incorrectly then use Precision.

b) If our goal is detect only all positive sample, the use Recall.

Mo | Tu | We | Th | Fr | Sa | Su

Memo No.

Date / /

F-beta Score

F-beta score = $\dfrac{(1+\beta^2) * Precision * Recall}{\beta^2 . Precision + Recall}$

① If FP and FN are both important:

$\beta = 1$

$$F_1 \text{ score} = \frac{2 * P \times R}{P + R}$$

② If FP is more important than FN.

$\beta = 0.5$

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

③ If FN >> FP, FP is less important than FN.

$\beta = 2$

$$F_2 \text{ score} = \frac{(1+4) \times P \times R}{4 \times P + R}$$

Memo No.

Date / /

Issue with High Dimensionality:

→ when a categorical variable has high number of label, it leads to quasi complete seperation

when the dependent variable seperates an independent variable or combination of several independent variables to a certain degree.

→ It can affect the convergence of the model and can lead to incorrect decision.

Solution :- Collapsing categories based on reduction in chi-square.