Logistic Regression:-

- It comes under supervised Learning
- Logistic regression is classification algorithm used to predict the probability of categorical target variable.

Assumption of logistic Regression

- No or Little multicollinearity
- The output should be binary [(0/1), (True/false), (diabatic/ nondiabetic)]
- It requires huge amount of data.

Types of classification

① Binary classification: Binary classification algorithms are used when target column has two categories

Ex:- diabetic or non diabetic.

② Multi classification:-

Multi classification algorithm used when we have two or more categories in target column.

Ex:- Iris.

How logistic Regression works?

Marks	Result
20	Fail (0)
30	Fail
45	pass (1)
78	pass
92	pass
50	?

Linear Regression

- Fit a line

$m = ?$ $\Rightarrow$ $m = 0.5$

$c = ?$ $\quad c = 20$

$y = 0.5x + 20$

$y = 0.5 \times 50 + 20$

$\boxed{y = 45}$

why Linear Regression Fails on classifi-cation data.

- It fits a straight line by finding slope and intercept then we use that line to make predictions we get continuous values which is violating the assuption of logistic Regression.

- Linear regression fails because it gives values other than 0 & 1.

Q. How to overcome the issue caused by linear Regression?

A. Sigmoid function:

sigmoid function is a mathematical function which transform continues data into a range of 0 to 1.

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

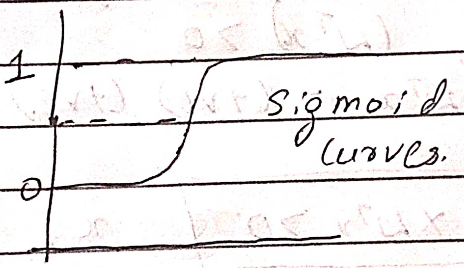
$z = mx + c.$

Marks	Result	$\hat{y}$	$h_\theta(x)$	$\hat{y}$
20	Fail	-100	0.2	0
30	Fail	-20	0.3	0
45	pass	100	0.6	1
78	pass	200	0.7	1
92	pass	300	0.9	1

Threshold:-

$$h_\theta(x) < 0.5, \quad y = 0$$
$$h_\theta(x) \geq 0.5, \quad y = 1$$

Graph:-



- $y = m_1 x_1 + m_2 x_2 + m_3 x_3 + b$
  $y = w^T x + c$

  $w^T = [m_1 \; m_2 \; m_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

- Distance b/w point and plane

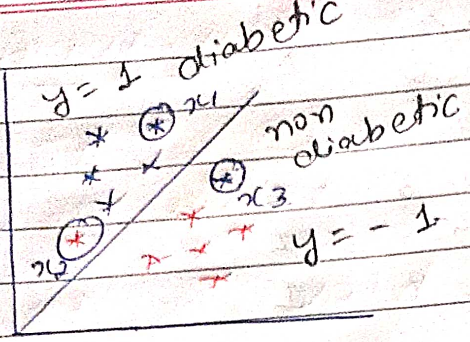
  $d = \frac{w^T x}{|w|}$, assume $|w| = 1 \Rightarrow$ unit vector

  $d = w^T x$

- The observation is correctly classified only if

  $$y \times w^T x > 0.$$

Ex:-



- The positive classes are denoted by y = 1
- the negative classes are denoted by y = -1
- The distance above the plane is always positive $w^T x > 0$

case - 1 :-

x_1 belongs to positive classes, so y = 1
x_1 is above the plane.
so, $(w^T x) > 0$.
$y \times w^T x = (+ve)(+ve) = (+ve)$

$\boxed{\therefore y \times w^T x > 0}$ x_1 is correctly classified

case - 2 :-

- x_2 is belong to negative, so $y = -1$
- x_2 is above the plane, $w^T x > 0$

$y \times w^T x = (-ve)(+ve) = -ve$
$y \times w^T x < 0 \rightarrow x_2$ is wrongly classified.

case - 3 :-

- x_3 is belong to positive, $y = 1$
- x_3 is below the plane, ~~w~~ $w^T x < 0$

$y \times w^T x = (1)(-1) = (-ve)(+ve) = -ve$

$y \times w^T x < 0 \rightarrow x_3$ is wrongly classified.

Evaluating the model:

(1) confusion matrix:

A confusion matrix is a table that summarizes the performance of a classification model by comparing predicted & actual class labels.

actual ↓	Prediction → Non diabetic (0) (-ve)	Diabetic (+ve) (1)
(-ve) Non diabetic 0	TN	FP
(+ve) diabetic 1	FN	TP

↓actual	Prediction → (-ve) 0	(+ve) 1	
(-ve) 0	TN	FP	→ Type 1
(+ve) 1	FN	TP	
	↓ Type 2		

correct classification = TP + TN

wrong classification = FP + FN.

actual	Pred.	
0	0.	TN
1	1	TP
0	1	FP
1	0	FN

(2) Accuracy:

- It is used to measure the strength of the model.
- It is valid only when data is balanced

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{Total prediction}}$$

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

(3) precision:

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

Out of positive prediction, what ~~pe~~ percentage is true positive.

(4) Recall:

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

out of actual positive, how many are true positives

(5) F_1 score:

It is the harmonic mean of Recall & precision.

$$F_1 Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

- Accuracy is good when data is balanced.
- F_1-Score is good when data is Imbalanced.

(6) ROC - AUC curve:-

It is one of the matric used to Evaluate binary classification problem

ROC → Reciever operator or charateristics

It is used to Evaluate binary classification model at various thresholds.

y	$\hat{y}$	$h_\theta(x) = 0.5$	0.2	0.7
0	0.2	0	0	0
1	0.4	0	1	0
0	0.3	0	1	0
1	0.7	1	1	1
1	0.8	1	1	1
1	0.9	1	1	1
0	0.5	1	1	0
1	0.6	1	1	0

TPR → True positive rate

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

FPR → False positive rate

$$FPR = \frac{FP}{FP+TN}$$

at 0.5

$$TPR = \frac{TP}{TP+FN} = \frac{4}{4+1} = \frac{4}{5} = \underline{\underline{0.8}}$$

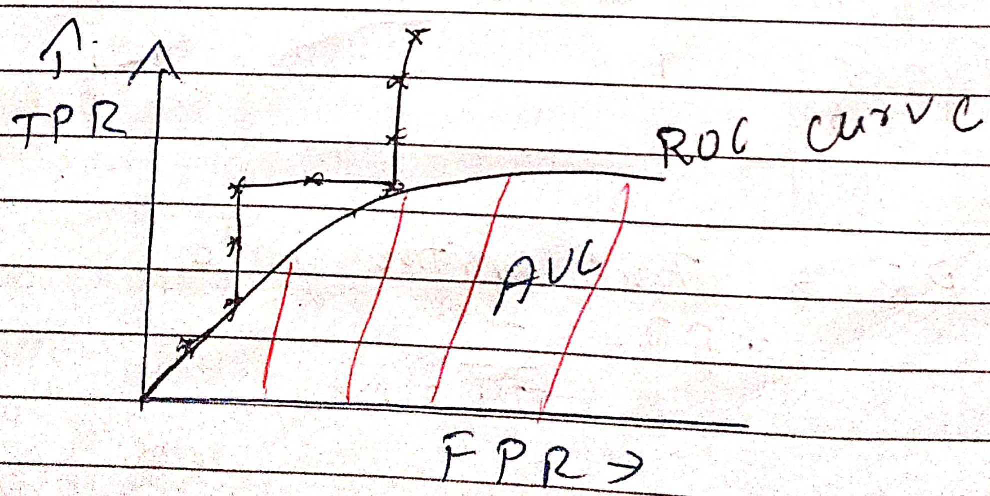
TP = [1 - 1] = 4 no 1-1 in table 4 so,
FN [10-0] = 1

~~FPR~~

$$FPR = \frac{FP}{FP+TN} = \frac{1}{1+2} = \frac{1}{3}$$

TN = [0 - 0] = 2
FP = [0 - 1] = 1

- ROC will find TPR and FPR for various thresholds
- Plot graph of TPR v/s FPR



- AUC → Area under the curve

- AUC tells about strength of the model by measuring area below the curve

Range of AUC = [0, 1]

- more the area below the curve then model is good

- If area below the curve is less, than model is bad.