Logistic Regression :-

→ It comes under supervised learning [both x & y].

→ It is used to solve classification problems [target is categorical]

eg: Diabetic data, cancerous data etc..,

→ It is used to predict the probability of categorical dependent variable.

Assumption of Logistic Regression.

→ The target variable is binary i.e. Yes / No, True / False, 0/1,...

→ No or little multicollinearity

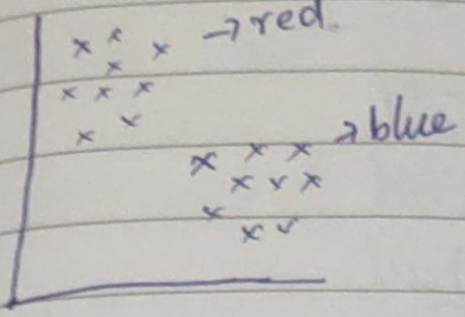
→ Outliers should be handled.

→ Sample size should be sufficiently large.

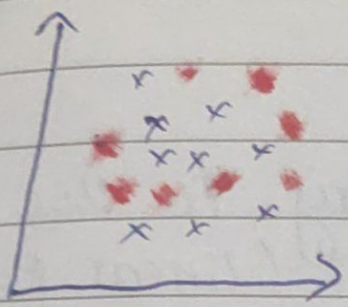
Date .....

→ Data is linearly seperable.

Linearly seperable → There should by be clear seperation among the classes.



Non-Linear → overlapping of classes.



How Logistic Regression works?

Pblm statement :-

I want to predict whether student will pass or not in exam.

Date .....

DATA :-

No. of hours studied (x)	Result (y)
1	Fail
4	Fail
3	Fail
8	Pass
10	Pass
12	Pass.
5	?

What happens if we try to use LR? (Linear Regression)

→ Linear regression will fit a straight line which seperates two classes.

$$y = mx + c$$

→ If we use Linear regression to make predictions it will ~~make~~ use straight line and give prediction other 0 & 1.

Date .....

→ If data has outliers then line gets shifted, it results in misclassification.

Sigmoid Function :-

It is one of the mathematical functions that we use to transform continuous values into a probability values which lies within a range 0 to 1.

$h_\theta(x) = g(z) = \frac{1}{1+e^{-z}}$ → O/P will be [0,1]

where $z = mx + c$
$z = \beta_0 + \beta_1 x$ } → straight line eqn.

Set a threshold (For Predictions)

$h_\theta(x) \geq 0.5 \rightarrow y = 1$ / Pass

$h_\theta(x) < 0.5 \rightarrow y = 0$ / Fail

Date .....

x	Result	$\hat{y}$	$h_\theta(x)$	Prediction
2	Fail (0)	-50	0.2	Fail
3	Fail (0)	40	0.4	Fail
5	Fail (0)	-20	0.3	Fail
8	Pass (1)	70	0.6	Pass
10	Pass (1)	80	0.8	Pass
12	Pass (1)	100	0.9	Pass.

* How it works?

→ It will fit lines.

→ Make Predictions.

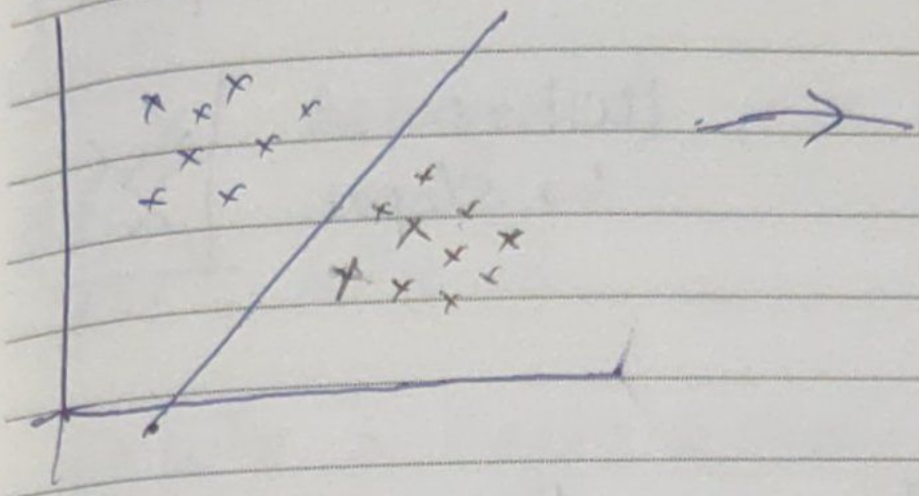
→ Transform continuous values using sigmoid function.

→ Set threshold.

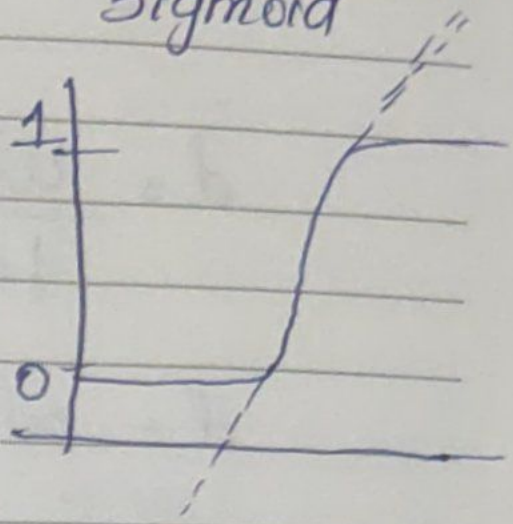
→ If threshold is ≥ 0.5 classify as pass and if threshold is less than (< 0.5) classify as fail.

Date .....

Linear Regression



Sigmoid



How classification works?

$$y = m_1x_1 + m_2x_2 + m_3x_3 + c$$

$$y = w^Tx + c$$

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

Let us say line passes through origin,

$c = 0$

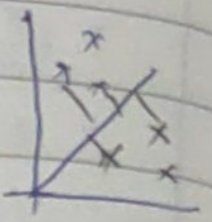
$$y = w^Tx$$

Date .....

How to find distance b/w point and plane?

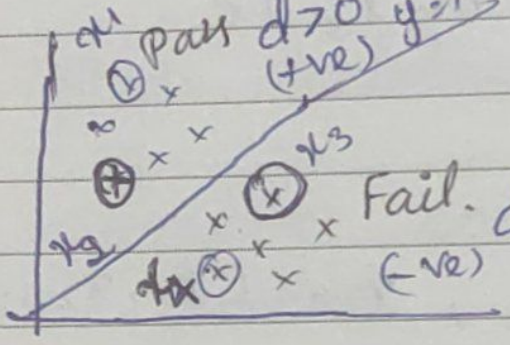
$$d = \frac{w^T x}{|w|}, \quad |w| = 1$$

↳ slopes.



$$d = w^T x$$

$y \times d = y \times w^T x \geq 0$ then we say the model has done correct classificat



→ Any point above the line / plane will have positive distance

→ Any point below the line / plane will have negative distance.

→ All (+ve) classes are denoted by 1.

Date .....

→ All (-ve) classes are denoted by -1.

Case ①

x_1 belongs to +ve class.

then,

$y=1$, $w^T x > 0$

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

$y \times w^T x > 0$, $\therefore x_1$ is correctly classified

Case 2

x_2 belongs -ve class.

then

$y=-1$, $w^T x > 0$

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

$y \times w^T x < 0$ so, x_2 is misclassified.

Case 3

x_3 belongs -ve class

$y=-1$, $w^T x < 0$

Date .....

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

$y \times w^T x \geq 0$

so, x_3 is correct classification.

Case 4:

x_4 belongs to +ve class

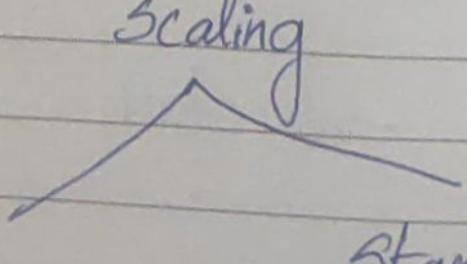
$y = 1, \quad w^T x < 0$

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

$y \times w^T x < 0$, x_4 is misclassified.

Scaling.

scaling is always applied on continuous numerical column.



Date .....

Normalization :-

→ Min max Scalar

It transforms all the values range into a scale of 0 to 1 range

$$\text{Minmax} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Standardization :-

→ Standard Scalar

$$z = \frac{x - \mu}{\sigma}$$

$\mu \rightarrow$ mean

$\sigma \rightarrow$ std.

It transforms all the values into a scale of z-value.

Imp:

Scaling is important to transform ~~characteristic~~ Continous data into certain scale.

Date .....

* Scaling is very important in distance based algorithm.

Metrics to Evaluate classification problems:

Different metrics are.

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall / Sensitivity
5. F1 score.

Note:
classification algms are evaluated based on misclassification that model has done.

Confusion Matrix:

Actual	Predicted 0	Predicted 1
0	TN	FP
1	FN	TP

0-0 → TN 0-1 → FP

0-1 → FP 1-1 → FN

Actual	Predicted Non diabetic (0)	Predicted Diabetic (1)
Non diabetic (0)	correct	wrong
Diabetic (1)	wrong	correct

Date .....

Ex:

TN → Actual is Non_diabetic and predicted as dia non_diabetic

TP → Actual is diabetic and predicted is diabetic.

FP → Actual is Non_diabetic and predicted is diabetic
↓(Type_1 error)

FN → Actual is diabetic and predicted is Non_diabetic.
↓
Type-2 error

→ Confusion Matrix gives information about no. of correct classification & misclassificat.

Correct classification = TP + TN

Mis classification = FN + FP.

Date .....

② Accuracy Score:-

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

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

→ Accuracy measures strength of model.

→ It talks about the percentage of correct predictions.

→ Accuracy score is best metric when data is balanced.

→ If data is not balanced then accuracy is not best metric; in that case, we use recall, precision and F1 score.

③ Precision

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

Date .....

Out of positive predictions, what percentage are truely positive:

ex: $\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$ → how many are 1 -1]

④ Recall

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

Out of all actual positive, what percentage is truely positive.

$\begin{bmatrix} 1 & -0 \\ 1 & -1 \end{bmatrix}$ → how many are 1 -1].

⑤ F1 score :-

→ F1 score is (average) of Precision and Recall. ~ harmonic mean

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

eg:

F1-score gives total positive predictions out of error.

Date .....

actual predicted.

eg

y	$\hat{y}$
0	0
1	1
0	0
1	1
0	1
1	0
0	1
0	1
1	0

① Confusion Matrix.

	Pred 0	Pred 1
Actual 0	2	3
Actual 1	2	2

TP → 2
TN → 2
FP → 3
FN → 2

② Accuracy := $\frac{TP+TN}{TP+TN+FP+FN} = \frac{2+2}{2+2+3+2} = 4/9$

$= 0.44$

③ Precision = $\frac{TP}{TP+FP} = \frac{2}{2+3} = 2/5 = 0.4$

④ Recall = $\frac{TP}{TP+FN} = \frac{2}{2+2} = 2/4 = 0.5$

Date .....

5) F1-score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$= \frac{2 \times 0.4 \times 0.5}{0.4 + 0.5} = 0.44$

Classification

Binary classification

→ If target columns has two categories/output

eg: Diabetic < Diabetic / Non diabetic

Logistic Regression works well with binary classification.

Multiclassification

→ If target column has three or more categories / output.

eg: Iris < Setosa / Versonica / Vaginica

Ratings < 1 / 2 / 3

→ It's Not work well with Logistic regression.

Date .....

ROC - AUC Curve :-

→ It is one of the metric used to evaluate Binary classification problems such as logistic regression.

ROC → [Receiver Operator characteristic]

Using this ROC, we evaluate our model at different thresholds.

$h_\theta(x) < 0.5 \rightarrow y = 0$

$h_\theta(x) \geq 0.5 \rightarrow y = 1$

→ threshold

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

Date .....

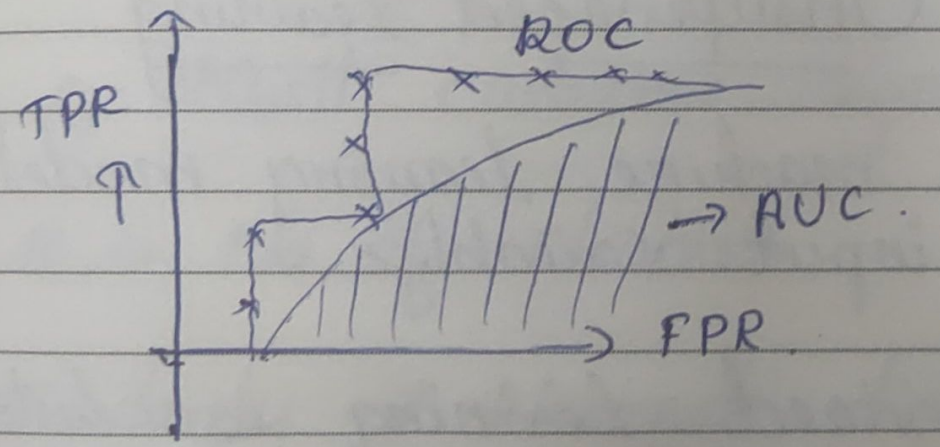
ROC will plot a graph of TPR Vs FPR

TPR → True Positive Rate.

$TPR \Rightarrow \frac{TP}{TP+FN} = \frac{2}{2+1} = 2/3$ → at 0.5 threshold

FPR → False Positive Rate.

$FPR = \frac{FP}{FP+TN} = \frac{1}{1+2} = \frac{1}{3}$ at 0.5 threshold



AUC [Area Under the curve]

→ AUC tells about the strength of the model by measuring area below the ROC curve.

Date .....

→ More the area below the ROC curv better is the model.

Range of AUC = [0,1]

AUC value near to '0' we say bad model

AUC Value near to '1' we say good model