(2) LOGISTIC REGRESSION

* log to sigmoid $\Rightarrow p = \frac{1}{1+e^{-x}}$

$$= \frac{1}{1+e^{-(mx+c)}}$$

$$= \frac{1}{1+\frac{1}{e^{(mx+c)}}}$$

Pro function $\Rightarrow$ $\boxed{P(x) = \frac{e^{(mx+c)}}{1+e^{(mx+c)}}}$ .... Equ for P(x) (for probability)

Negation of probability

$$1 - P(x) = 1 - \frac{e^{(mx+c)}}{1+e^{(mx+c)}}$$

$$= \frac{1+e^{(mx+c)} - e^{(mx+c)}}{1+e^{(mx+c)}}$$

$$= \frac{1}{1+e^{(mx+c)}}$$

odds value

Event happing / Event not happing $\therefore \frac{P(1)}{1-P(1)} = \frac{\frac{e^{(mx+c)}}{1+e^{(mx+c)}}}{\frac{1}{1+e^{(mx+c)}}}$

$$\frac{P(1)}{1-P(1)} = \frac{e^{(mx+c)}}{1+e^{(mx+c)}} \times \frac{1+e^{(mx+c)}}{1}$$

$\frac{P(x)}{1-P(x)}$ ⇒ where, $x \Rightarrow y = \frac{1}{1+e^{-(mx+c)}}$

$$\frac{P(x)}{1-P(x)} = e^{(mx+c)} \Rightarrow \log\left(\frac{P(x)}{1-P(x)}\right) = \log_e e^{(mx+c)}$$

$$\left[\log\left(\frac{Px}{1-P(x)}\right) = mx + c\right]$$ ..... log logit function

Prediction ⇒

$$y = \begin{cases} 0 & \text{if } p(x) < 0.5 \\ 1 & \text{if } p(x) >= 0.5 \end{cases}$$

↑ Threshold

Cost function ⇒

$\hat{y}$ ← is the probability

- logloss

target; $p(x)$ ↓ probability value; $1-p(x)$

Log loss = $J(\theta) = \frac{-1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(\hat{p}^{(i)}) + (1-y^{(i)})\log(1-\hat{p}^{(i)})\right]$

(Binary cross entropy)

↑ is the prob of 1 ↑ is the prob of 0

Note :-
- Loss means we are going to find out w.r.t only one record
- Cost means we are going to find out w.r.t Entire record

- Cost function

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

- Optimization

⇒ (Gradient Decent)

$m_{new} = w_{old} - \eta \frac{\partial L}{\partial m}$

$m_{new} = w_{old} - \eta \frac{\text{log loss}}{\partial m}$

odds ⇒ $\frac{\text{pr of happening}}{\text{pr of not happening}}$

→ (1) Calculation
→ (2) loss / cost
→ (3) optimize the per

Imp ⇒ odds (what is odds?)

$\frac{p(x)}{(1-p(x))}$ pr ← negation of pr

eg ⇒ $p(x) = 0.6$
$1-p(x) = 0.4$

Imp ⇒ ~~log git~~ logit function (what is logit function?)

⇒ log of odds

$$\log(\text{odd}) = \log\left(\frac{p(x)}{1-p(x)}\right)$$

Optimization ⇒ 1st approach is "Iterative Approach" (Gradient D.)
2nd approach is "Maximum likelihood Approach" maximum likelihood Estimate (MLE)
MLE (Probability)

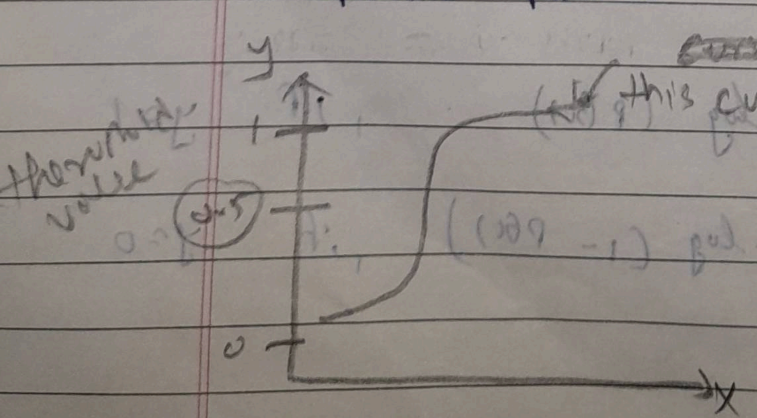
(mle optimize this parameter) $p(x) = \frac{e^{(mx+c)}}{1+e^{(mx+c)}}$

Optimize the parameter
→ MLE (probability based approach)
→ Iterative (Gradient D. approach)

In both we are going to find out the optimized parameter (m, c)

$$m_{new} = m_{old} - \eta \frac{\partial L}{\partial m}$$ (log loss)

This is iterative process until we are not getting a optimize parameter



● MLE ⇒ Goal is to get a best fit probabilistic curve

eg:

weight_value	obese / Not_obse	Prediction (Prob)
50 kg	No → 0	0.4
60 kg	No → 0	0.5
70 kg	ob → 1	0.6
75 kg	ob → 1	0.7
85 kg	No → 0	0.3
90 kg	ob → 1	0.8

main aim is to find out, the best pro curve for best pro value

$p(x) = \frac{e^{(mx+c)}}{1+e^{(mx+c)}}$ → threshold → 0.5

$\text{loss\_value} = -\frac{1}{m}\sum_{i=1}^{m}\left[y \log(p(x)) + (1-y)\log(1-p(x))\right]$

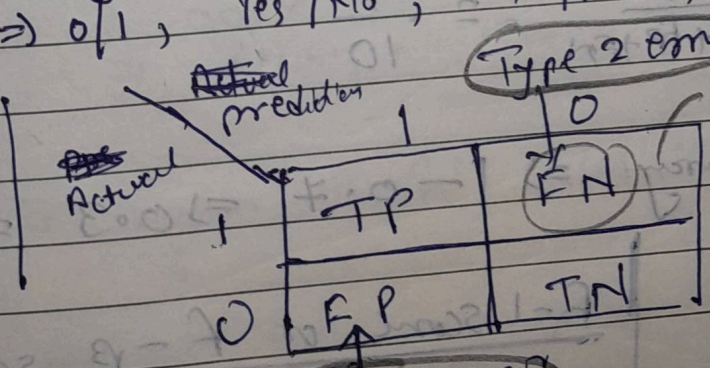
Gradient decent for optimizer ⇒ $m_{new} = m_{old} - \eta \frac{\partial L}{\partial m}$ (L: logloss)

$C_{new} = C_{old} - \eta \frac{\partial L}{\partial C}$ (L: logloss)

* Binary classification

⇒ 0/1, Yes/No, T/F, T/F [In confusion matrix we have 2 class]

Twin 2 class (0, 1)



Actual it is 1 but the answer which we got seeing inside prediction

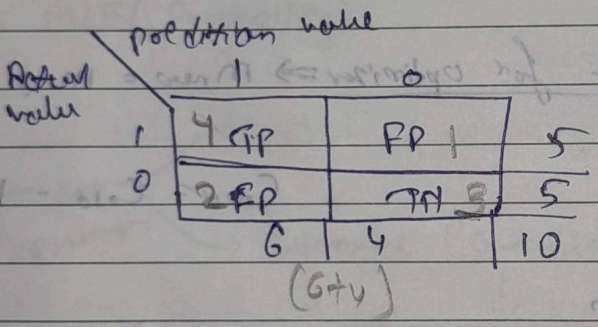
(TPR) Sensitivity ⇒ True Positive Rate ⇒ TP/P (Recall)

(FPR) Specificity ⇒ False Positive Rate ⇒ TN/N (Precision)

eg. Obese or Not Obese

eg: Threshold (0.5)

(weights) X	O/NO y	$\hat{y}$ (0.5)		$\hat{y}$ model
50	0	0.4	→	0 ⇒ TN
60	1	0.5	→	1 ⇒ TP
65	0	0.6	→	1 ⇒ FP
68	0	0.3	→	0 ⇒ TN
70	1	0.7	→	1 ⇒ TP
72	1	0.8	→	1 ⇒ TP
84	0	0.85	→	1 ⇒ FP
90	1	0.3	→	0 ⇒ FN
45	1	0.9	→	1 ⇒ TP
55	0	0.2	→	0 ⇒ TN



- Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} = \frac{7}{10} \times 100 = 70\%$

- Error $= 1 - \text{Accuracy} \Rightarrow 1 - 0.7 \Rightarrow 0.3 \Rightarrow 30\%$

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

- Precision: $\frac{TP}{TP+FP}$

F-1 score or F-β score

f1 score $= \frac{1}{\frac{1}{P}+\frac{1}{R}}$ ... Harmonic mean

$= \frac{1}{\frac{R+P}{PR}}$

$= \frac{2 \times PR}{P+R}$

ROC Curve (It is Probabilistic Curve)

because it represent ⇒ TPR and FPR ⇒ Sensitivity & 1 - Specificity

eg	X	Y (Actual value)	$\hat{y}$ (Predicted)	(Threshold) 0.5	Suppose (Threshold) 0.6	0.3
	5.5	0	0.2	0	0	0
	5.6	0	0.4	0	0	1
	5.8	1	0.6	1	1	1
	6.1	1	0.8	1	1	1
	6.2	0	0.85	1	1	1
	6.6	0	0.3	0	0	1
	6.7	1	0.1	0	0	0
	4.2	1	0.9	1	1	1
	6.4	0	0.95	1	1	1
	5.0	1	0.2	1	0	1

0.5 ⇒ Accuracy = $\frac{TP+TN}{All} = \frac{7}{10} = 70\%$

0.6 ⇒ Accuracy = $\frac{TP+TN}{All} = \frac{6}{10} = 60\%$

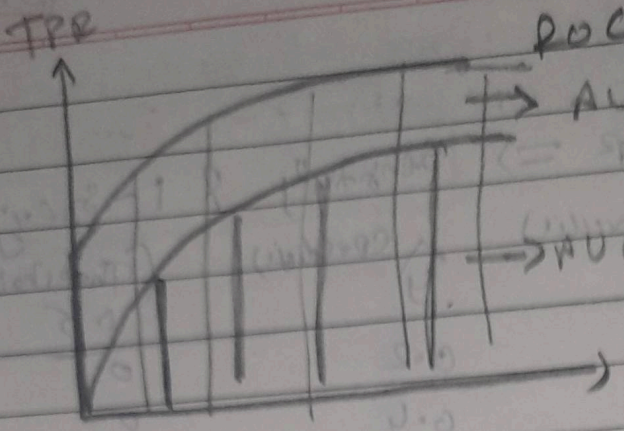
0.3 ⇒ Accuracy = $\frac{TP+TN}{All} = \frac{5}{10} = 50\%$

Imp ⇒ Changing Threshold value My Accuracy is also changing

Note:- But, we don't know which threshold we have to choose, for this only we are going to Calculate AUC Curve

- AUC (Area under Curve)

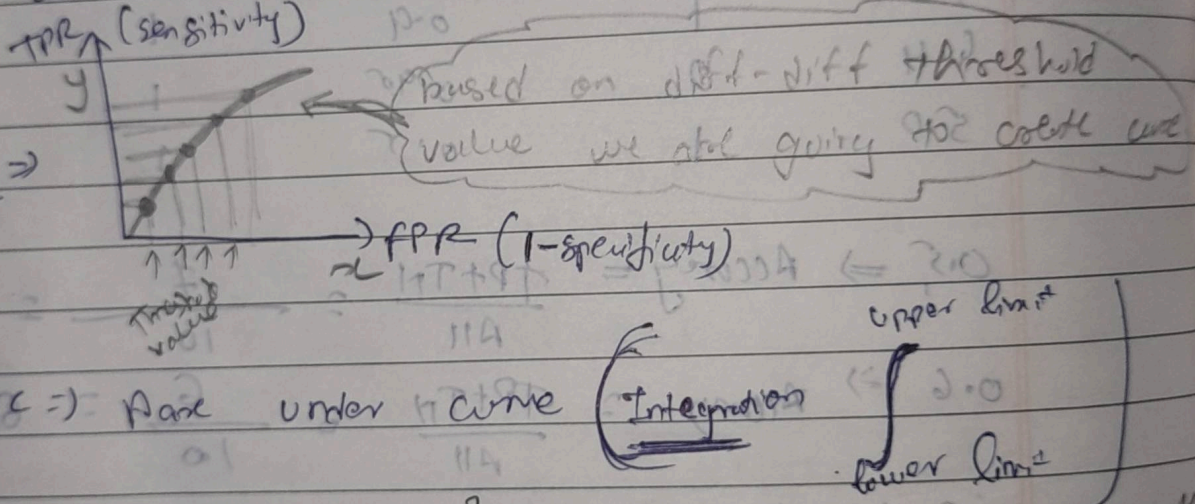
→ As much as AUC → As good as our model



→ As much as AUC → as good as our Model

AUC = 0.7 } 0.7 is good
AUC = 0.3 }

ROC ROC curve between TPR and FPR



Q) What is a Roc curve? Explain how a Roc curve works

⇒ AUC-Roc curve is a performance measurement for the classification problem at various threshold settings.

- Roc is a probability curve and
- AUC represents the degree or measures of separability

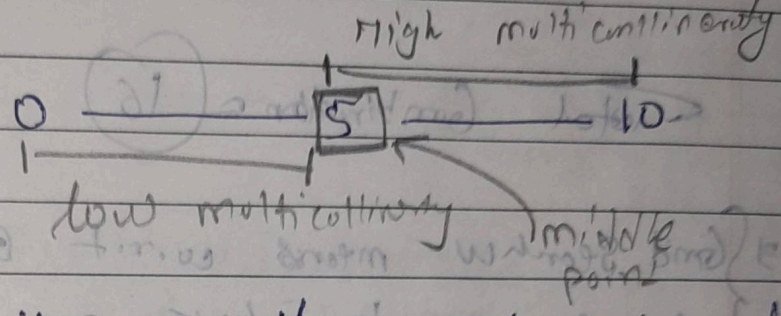
→ It tells how much model is capable of distinguishing between classes.

→ Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s

* Parameter
  Hyperparameter] — what is Difference between these

* tuning
  - Grid Search + CV
  - Random Search + CV

* VIF =>



- verbose = 1 ← Description of the model after fitting the model

> accuracy_score (y_test, y_pred)

- Confusion Matrix for 2 classes (binary classification) 0/1

A \ P	1	0
1	TP	FN
0	FP	TN

- Confusion Matrix for 3 class / 4 classes (multiclass classification)

A \ P	0	1	2
0	TN	FP	FP
1	FN	TP	FN
2	FN	FP	TP

Accuracy = (TN + TP + TP) / All (sum of Diagonal) / Total value

Error = 1 − Accuracy