Logistic Regression

→ Classification Problem

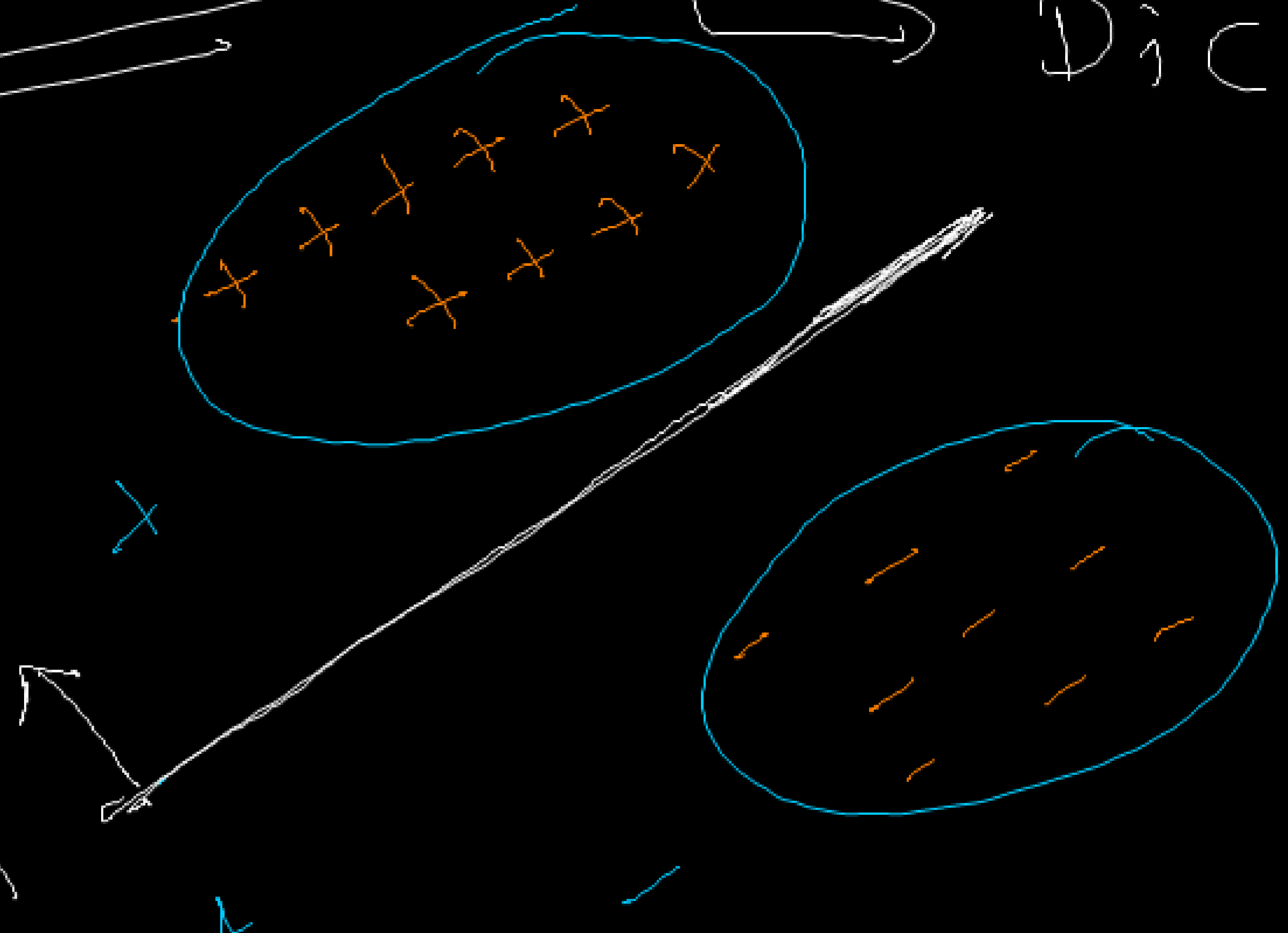
- Geometry
- Probabilistic
- Loss - function

Logit Regression

→ Binary nature
→ Categorical Prob
→ Dichotomous in nature

Simple, elegant

2D



if my data is linearly separable or almost linearly seperable

π = plane

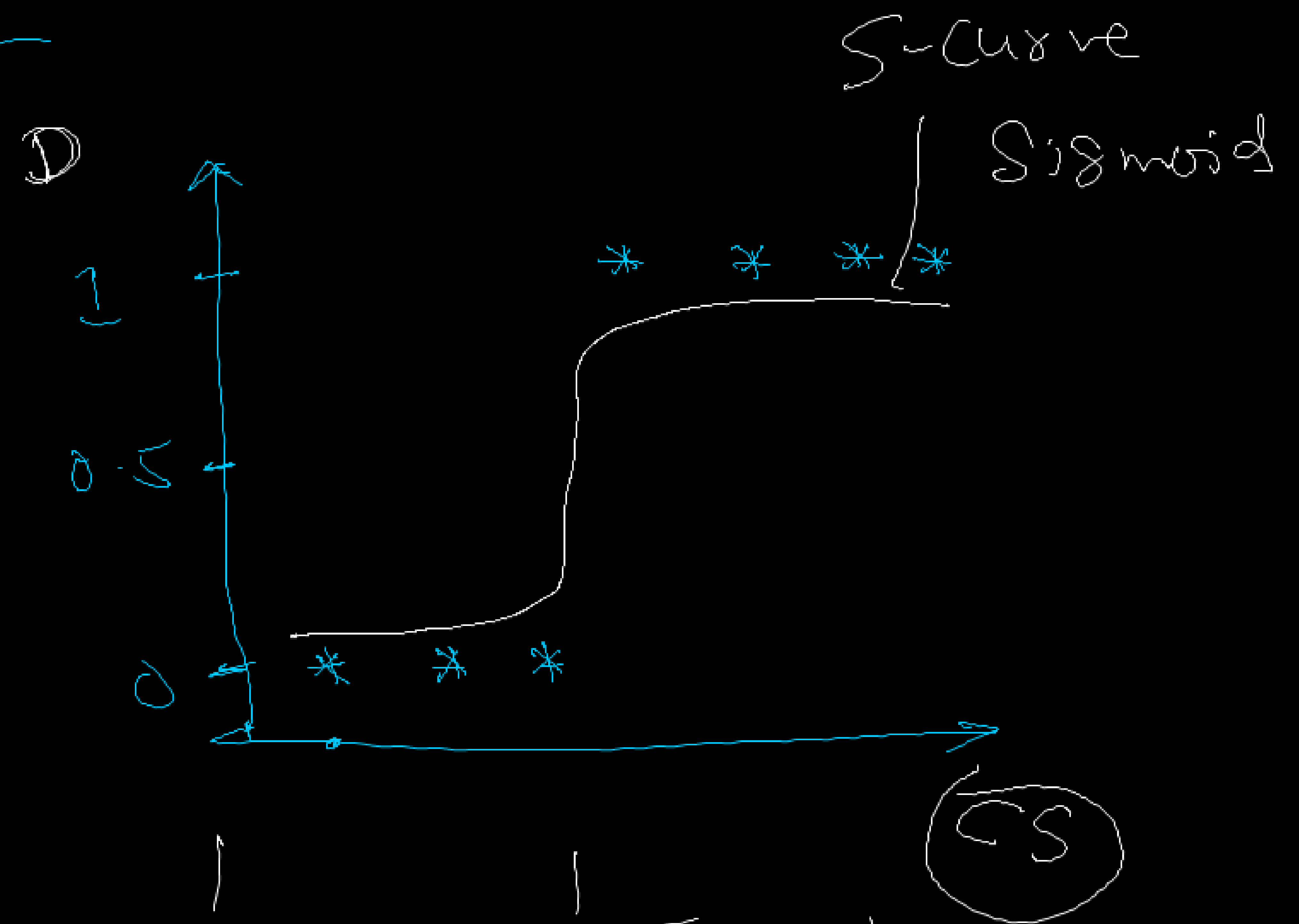
$\pi : w^T x + b = 0$

π : w , b

w → coefficient, b → intercept

Assumption :- Classes are almost/perfectly linearly seperation

Sigmoid = binary Class



D = Default = (1 & 0)

CS = credit score

Dep variable – Fraudulent or not

$$\boxed{P = \frac{e^{cs}}{1+e^{cs}}} = \boxed{\frac{1}{1+e^{-cs}}} = \frac{1}{1+\frac{1}{e^{cs}}} = \frac{1}{\frac{e^{cs}+1}{e^{cs}}}$$

$$y^{-1} = \frac{1}{y} \qquad y^{-2} = \frac{1}{y^2}$$

Mute Start Video Participants 69 Chat 9 Q&A New Share Pause Share Annotate Apps More

You are screen sharing Stop Share

$$\frac{P}{1} = \frac{e^{CS}}{1+e^{CS}}$$

Elementary maths

$$\Rightarrow P(1+e^{CS}) = e^{CS}$$

$$\Rightarrow P + P \cdot e^{CS} = e^{CS}$$

$$\Rightarrow P = e^{CS} - P \cdot e^{CS}$$

$$\Rightarrow P = e^{CS}(1-P)$$

$$\frac{\frac{1}{6}}{1-\frac{1}{6}} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}$$

$$\boxed{\frac{P}{1-P} = e^{CS}} \longrightarrow 1$$

$$\frac{P}{1-P} = \boxed{\text{ODDS}}$$ — Gambling / betting

CS = Single variable

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{——}$$

$$y = \beta_0 + \beta_1 x_1$$

Prob - Coin $\{H, T\}$ - $\frac{1}{2}$

$$\frac{\frac{1}{2}}{1-\frac{1}{2}} = \frac{\frac{1}{2}}{\frac{1}{2}}$$

$$\boxed{1:1}$$

$$\boxed{\frac{p}{1-p} = e^{cs}}$$

Probability $\begin{cases} 0 \\ 1 \end{cases}$

Regression Family

Case I :- $p = 0$

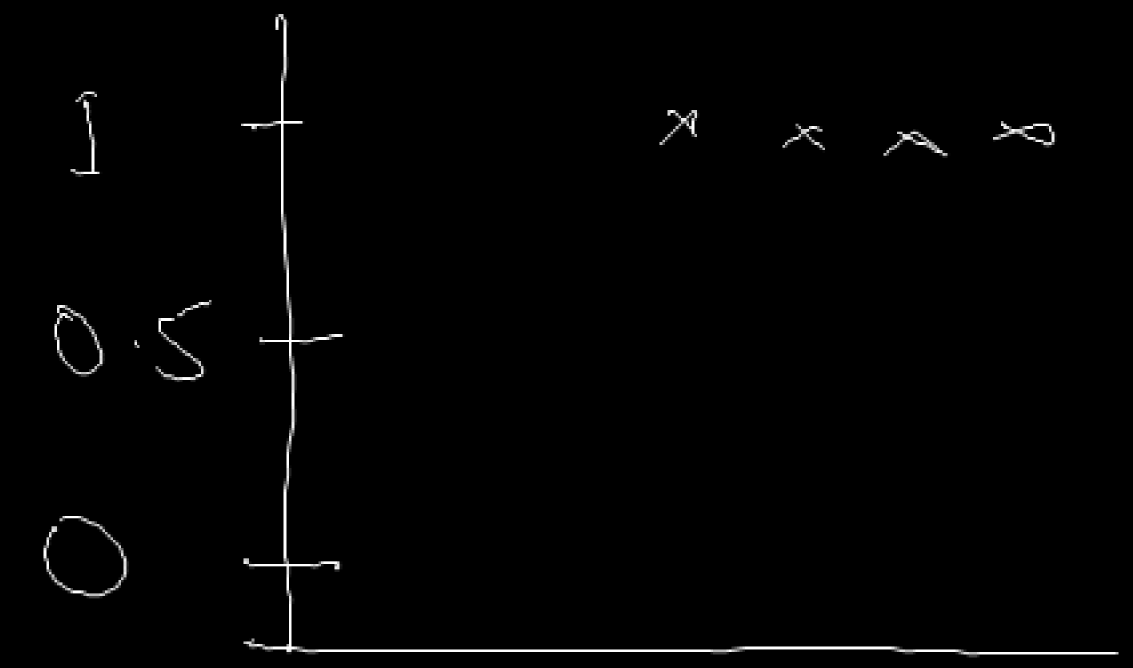
$$\frac{0}{1-0} = 0$$

$$\boxed{e^{cs} \begin{cases} 0 \\ +\infty \end{cases}} \rightarrow \begin{bmatrix} -\infty \\ +\infty \end{bmatrix}$$

Linear family $\begin{cases} -\infty \\ +\infty \end{cases}$

Case II :- $p = 1$

$$\frac{p}{1-p} = \frac{1}{1-1} = \frac{1}{0} = \infty$$



$$\frac{P}{1-P} = e^{cs}$$

$$\text{ODDS} = \frac{P}{1-P}$$

$\log_2$

base = $e = 2.7$

Log on both sides, we get (Natural log : $\log_e$

$$\log_e\left(\frac{P}{1-P}\right) = \log_e e^{cs} = cs$$

1

$$\ln\left(\frac{P}{1-P}\right) = cs$$

[where Natural log would be $\log_e$ or Ln]

Basic log function

$\log_a a = 1$

$\log_a a^2 = 2$

$\log_e e = 1$

$\log_e e^{cs} = cs$

Mute | Start Video | Participants 70 | Chat | Q&A | New Share | Pause Share | Annotate | Apps | More

You are screen sharing | Stop Share

$$\ln\left(\frac{p}{1-p}\right)$$

Case I: $p = 0$

$$\log\left(\frac{p}{1-p}\right) = \log\left(\frac{0}{1-0}\right) = \log(0) = -\infty$$

Case II :- $p = 1$

$$\log\left(\frac{p}{1-p}\right) = \log\left(\frac{1}{1-1}\right) = \log(\infty) = +\infty$$

Linear Reg
Regression

$-\infty$ $+\infty$

Log ODDS

$$\ln\left(\frac{P}{1-P}\right) = CS$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{---}$$

Linear

50000

DeP

Generalized linear Form

Idv

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{- - - -} + \text{Error}$$

Intercept Slope

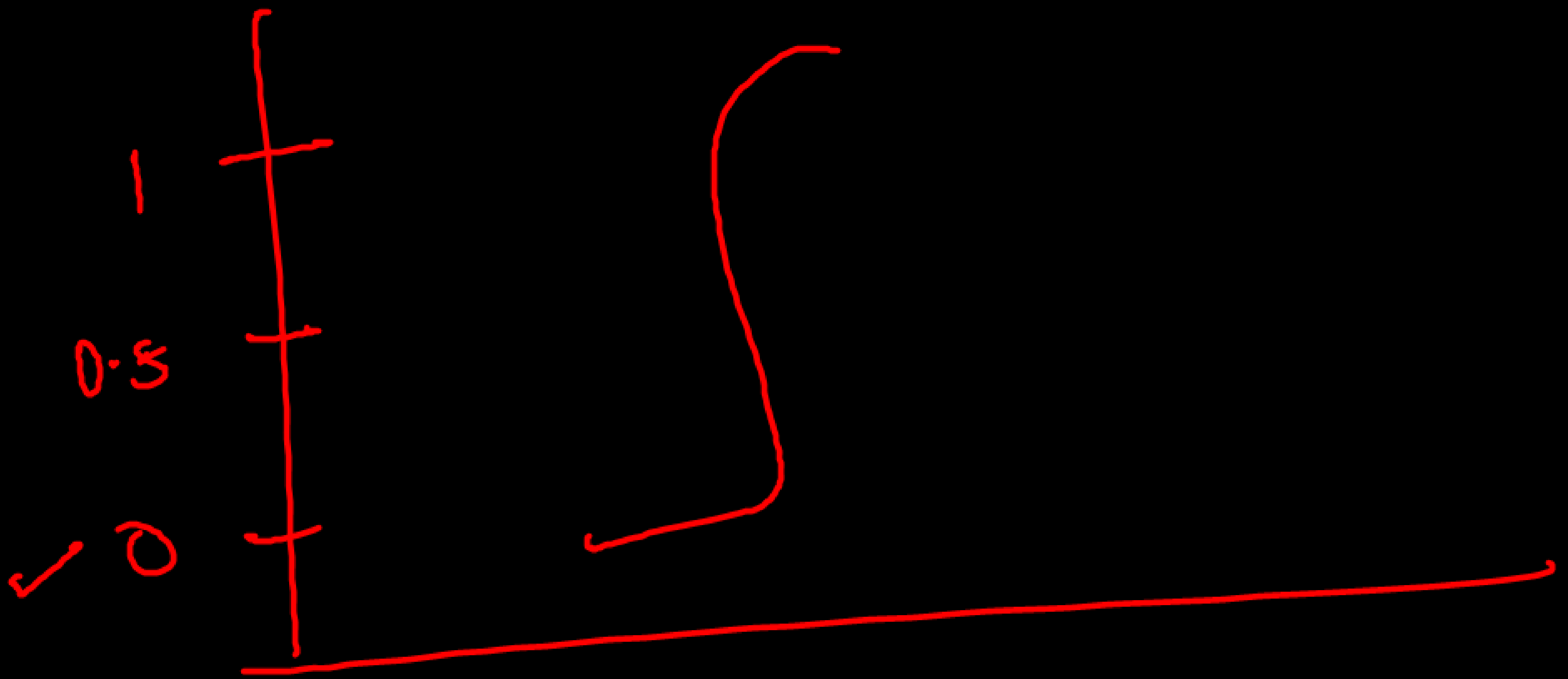
Formula

This is called Logistic or Logit Regression.

Where, $\ln\left(\frac{P}{1-P}\right)$ is generalized form of y.

You are screen sharing Stop Share

5000



-100K

You are screen sharing
Stop Share

Performance Prediction Of models

\# = Total no.

$$\text{Accuracy} = \frac{\#\ \text{correctly classified Points}}{\#\ \text{points given in Dataset}}$$

Dataset	Given dataset	predict	Total error
	60 = (+)ve	53 = (+)ve	7
	40 = (−)ve	35 = (−)ve	5
		88	12

Accuracy = 88% = $\frac{88}{100}$

Error = 12%

Case I :- Imbalance dataset

70% 30%

"Dumb" model

MLE = maximum likelihood Estimates

Not good measurement of Accuracy.

Case 2 :- Probability value

$m_1 \neq m_2$

$\hat{y}_1 = \hat{y}_2$ — Absolute

Given data		Pred prob value			
x	y	m_1	m_2	$\hat{y}_1$	$\hat{y}_2$
x_1	1	0.56	0.92	1	1
x_2	0	0.49	0.001	0	0
x_3	1	0.62	0.99	1	1
x_4	0	0.42	0.01	0	0

m_1 m_2 ✓

Ebi == Akshay

Accuracy ✓

m_2 is the better model than m_1

1 ← 0.99

0.5

0.52

Threshold

0.42

0 ← 0.01

Confusion matrix

Actual value →

Predicted value ↓

	0	1
0	TN	FN
1	FP	TP

N P

Trick

T P

① what is the predicted label

② are you correct?

0 = Negative

1 = Positive

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

$$\text{Error} = \frac{FN + FP}{(TN + FN + FP + TP)}$$

Actual

Pred	0	1
0	TN	FN
1	FP	TP
	N	P

$$P = FN + TP$$

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

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

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

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

Q:- FN or FP :- which is more not good?

Ans :- it's a domain specific Question.

example - Clinical sector data

		Act 0	Act 1
Pred	0	TN	FN
	1	FP	TP

Covid-19

Actually covid = True
Pred ⇒ False
FN

Act covid = False
Pred = True

Precision Recall/Sensitivity F1-Score

AUC & ROC

Case Study - Object - Encoding