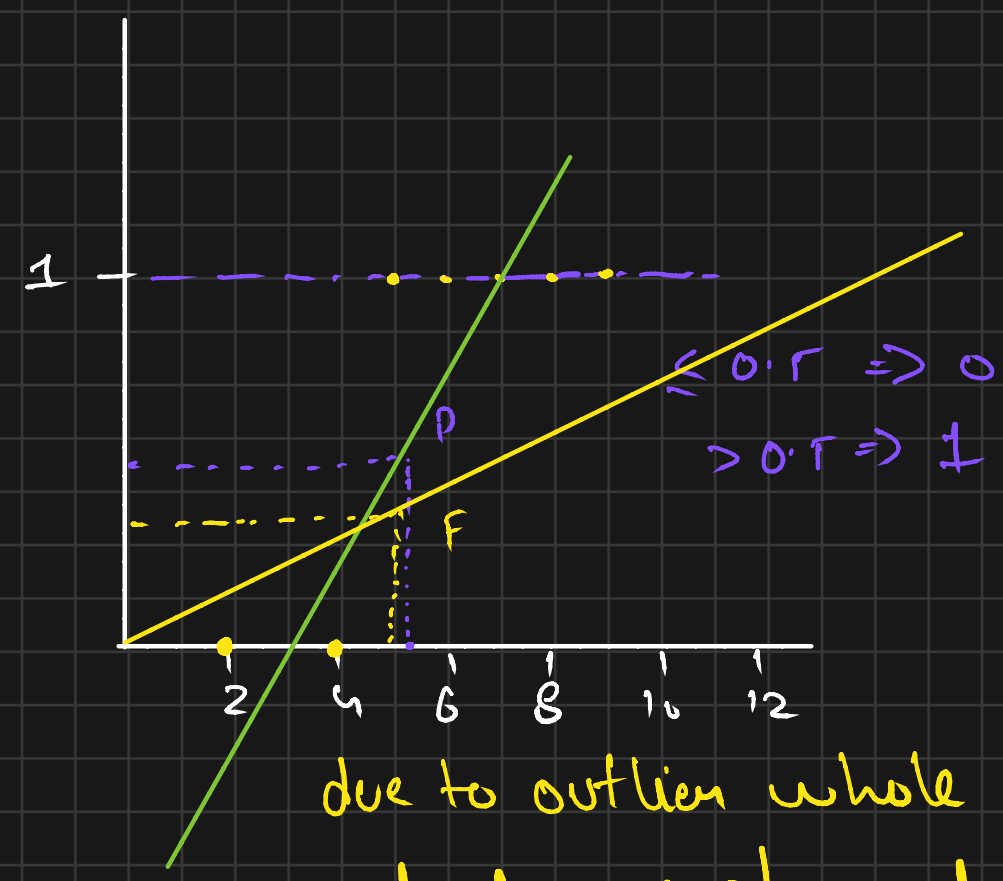
Logistic Regression

Stdy hrs	Pass/ fail
2	F
3	F
4	P
5	P
6	P
7	P
12	P



due to outlier whole output was changed

→ we can't use linear Regression

1) Outliers
2) output > 1 { squashing line }

⇒ Sigmoid fxn

$$= \frac{1}{1 + e^{-z}}$$

$$z$$

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

$$\sigma(z)$$

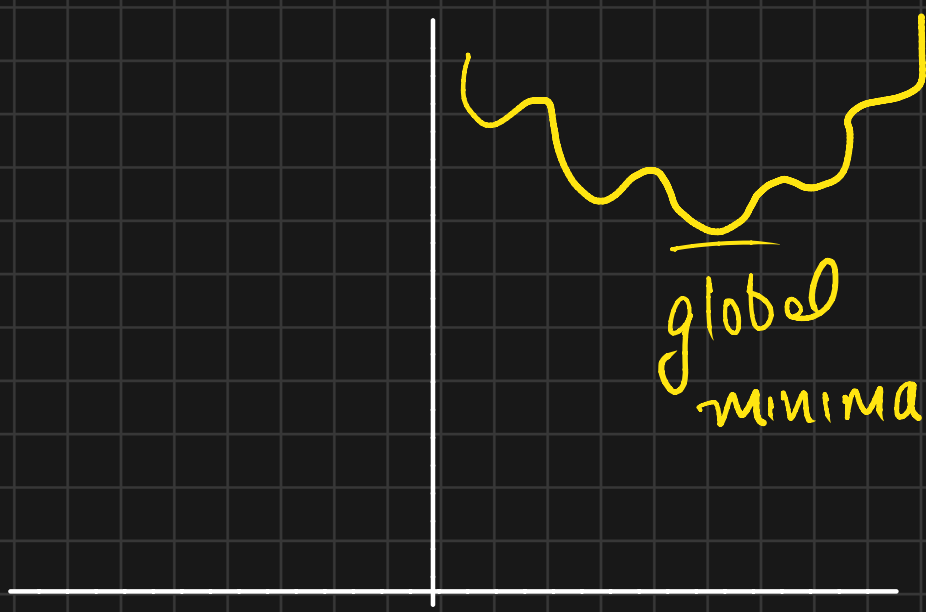
$$h_\theta(x) = \frac{1}{1 + e^{-z}} \rightarrow \theta_0 + \theta_1 x$$

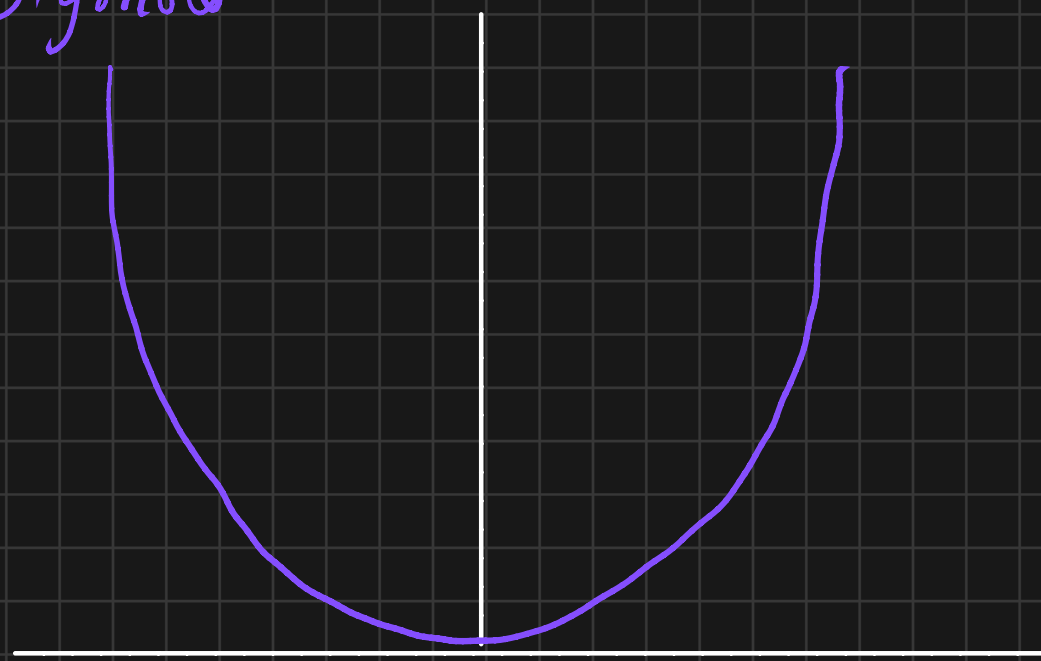
LR cost fxn

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

$$h_\theta(x) = \theta_0 + \theta_1 x \quad \text{— LR}$$

$$h_\theta(x) = \frac{1}{1+e^{-z}} \quad \text{— Sigmoid}$$





$$J(\theta_0, \theta_1) = \frac{1}{2m} (h_\theta(x) - y)^2 \qquad h_\theta(x)^1 = \frac{1}{1+e^{-z}}$$

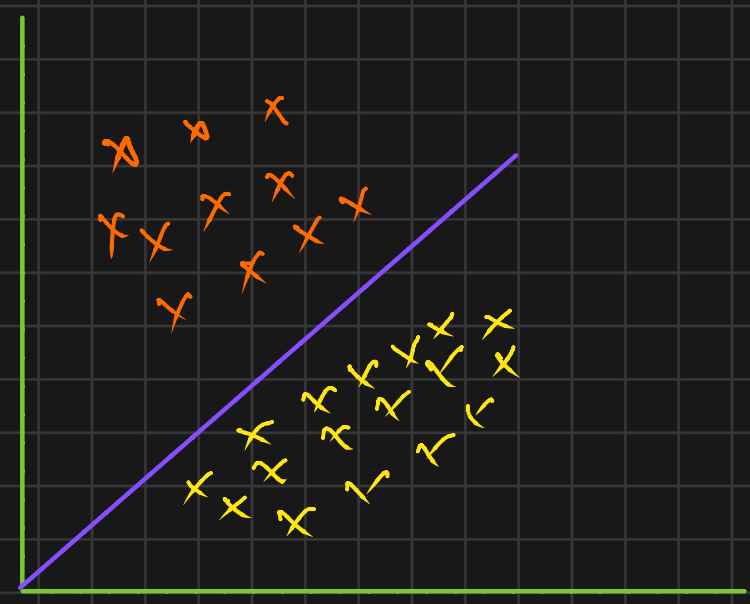
$$\text{cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

$$\text{cost}(h_\theta(x), y) = -y \log(h_\theta(x) - (1-y)(1 - h_\theta(x)))$$

minimize cost fxn $J(\theta_0, \theta_1)$ by changing θ_0, θ_1

Performance Matrices

1. Confusion matrix
2. Accuracy
3. Precision
4. F-Beta Score



Dataset

x_1	x_2	o/p y	$\hat{y}$ Pred by model
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0

1. Confusion Matrix

Pred \ actual	1	0
1	3	1+1
0	1	1

Pred \ actual	1	0
1	TP	FP
0	FN	TN

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

$$\frac{3+1}{3+2+1+1} = \frac{4}{7}$$

Dataset

1000 datapoints

Imbalance

900 → 1

100 → 0

① Precission = $\frac{TP}{TP+FP}$ } out of all the actual values how many are correctly predicted

Predic \ Actual	1	0
1	TP	FP
0	FN	TN

② Recall = $\frac{TP}{TP+FN}$ } out of all the predicted value how many are correctly predicted

Spam Classification : Mail → Spam, model → Spam } good

Mail → Not Spam, model → Spam } Blunder

	Spam (1)	!Spam (0) (Actual)
Spam	TP	FP
!Spam	FN	TN

(pred)

here we should decrease FP

here we use precision

Precision = $\frac{TP}{TP+FP}$ if FP decrease

TP will go up

Case 2 wether a person have Diabetes

	have	!have (Actual)
have	TP	FP
!have	FN	TN

(pred)

Truth → diabetes, model → No } Blunder

Truth → ! Diabetes, model → have Diabetes } 2nd option

here FN is a problem

because person is having diabetes but model predicts he doesn't have diabetes so it's a mistake

here we use Recall = $\frac{TP}{TP+FN}$

⇒ Tommorow Stock market will crash

	Crash (actual)	! crash (actual)
Crash (predict)	TP	FN
! Crash (predict)	FP	TN

Truth → Crash
model → Crash

here we should decrease Precision

⇒ F-Beta Score = $(1+\beta^2)\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

If FP & FN are both important

$\beta = 1$

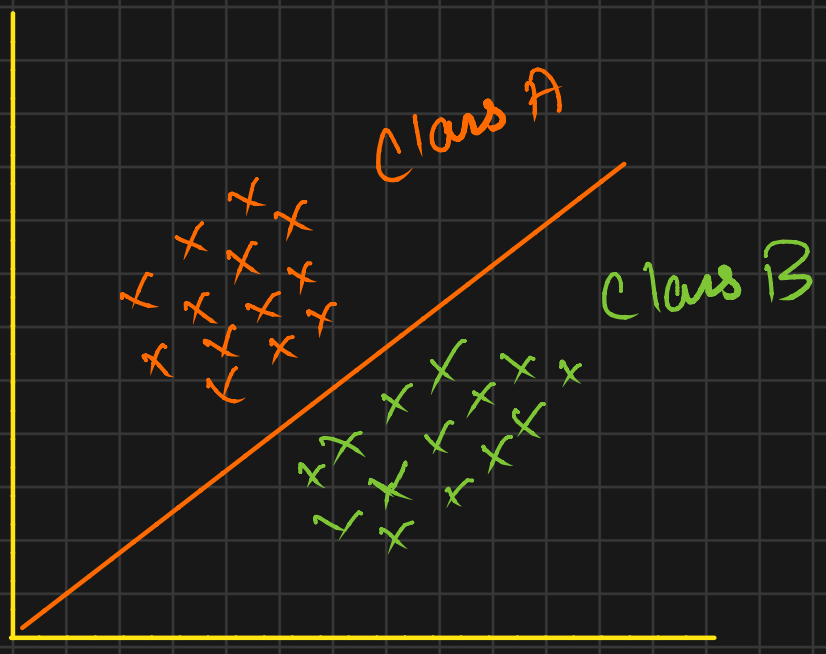
F1 Score = $(1+1^2)\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

Harmonic Mean

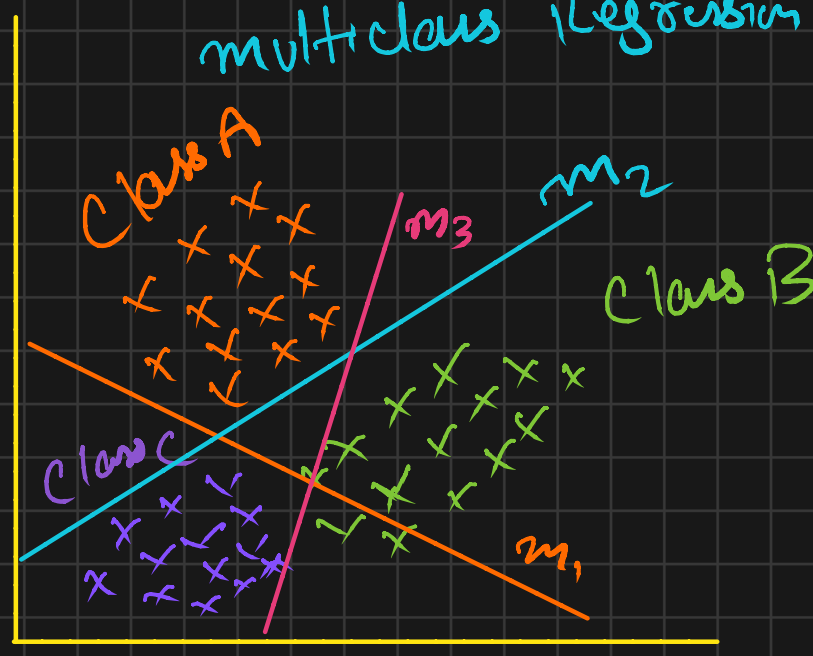
⇒ If FP is more important FN

$\beta = 0.5$

⇒ Logistic Regression one vs Rest



multiclass Regression



m_1 differentiate Class A, B with C

model m_1 → Binary Class-
model m_2 → Binary Class-
model m_3 → Binary Classfic-

f_1	f_2	f_3	op
-	-	-	O_1
-	-	-	O_2
-	-	-	O_3
-	-	-	O_1
-	-	-	O_2
-	-	-	O_3

$m_1 \rightarrow f_1, f_2, f_3 \rightarrow \{O_1\}$
$m_2 \rightarrow f_1, f_2, f_3 \rightarrow \{O_2\}$
$m_3 \rightarrow f_1, f_2, f_3 \rightarrow \{O_3\}$

f_1	f_2	f_3	op	O_1	O_2	O_3
-	-	-	O_1	1	0	0
-	-	-	O_2	0	1	0
-	-	-	O_3	0	0	1
-	-	-	O_1	1	0	0
-	-	-	O_2	0	1	0
-	-	-	O_3	0	0	1