

• ROC Curve.

↳ Used in Threshold Selection

→ Supervised ML

Regression
o/p → numerical

Classification
o/p → categorized.

→ ROC curve is specifically used in binary classification

→ Suppose -

ig | cgpa | plant

Now we run test data -

ig	cgpa	plant	Predict ⁿ	Pred-Prob
-	-	0	1	0.45
-	-	1	0.1	0.38
-	-	0	0	0.61

↳ Our model doesn't give 0 or 1 so how

if gives us predictⁿ prob. Now we select threshold (let 0.5). Now if Prob > 0.5 then 1 Prob < 0.5 → 0

↳ Using 0.5 threshold always is not good.

→ email classifier → 2 mistakes

Actual → Spam	A = Not Spam
Pred → Not Spam	P = Spam
1	2

Here mistake 2 is heavier than mistake 1
 So we set threshold = 0.15

$\text{Prob} > 0.15 \rightarrow \text{spam}$
 else $\rightarrow \text{Not spam}$

- Confusion matrix -

Predicted

	1	0
1	TP	FN
0	FP	TN

Actual

- True Positive Rate (TPR) (Benefit)

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

} out of all spam emails
 how much we have marked spam
 we want to maximize it.

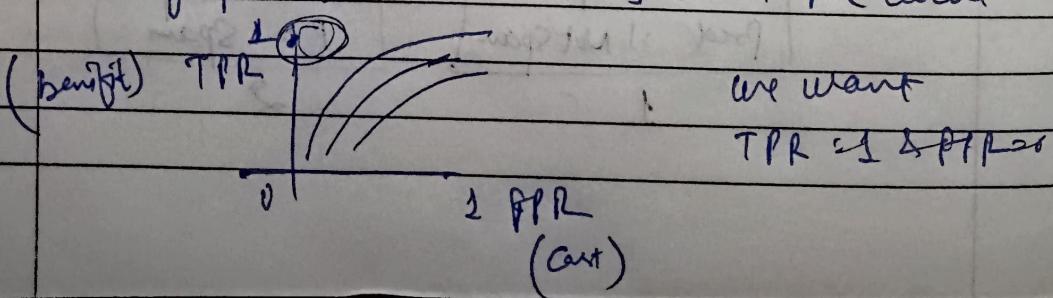
- False Positive Rate (FPR) (Loss) (Cost)

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

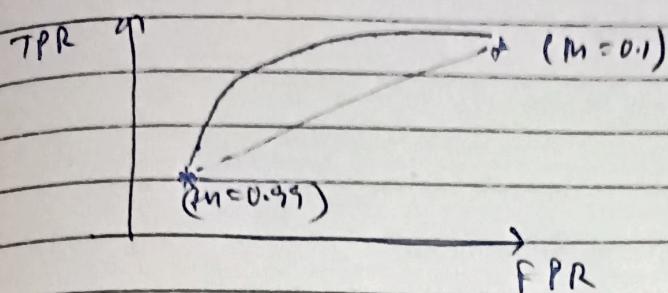
} out of all Not spams
 how many emails you have said spam

} we want to minimize it.

→ Graph b/w TPR & FPR is called ROC curve



Now seeing Different cases -



→ threshold = 0.1 (very low)

$$TPR = \frac{TP}{TP + FN} \quad \uparrow \text{ increases}$$

$$FPR = \frac{FP}{FP + TN} \quad \uparrow \text{ increases}$$

→ threshold = 0.99

TPR → TP ↓ TPR ↓ decreases

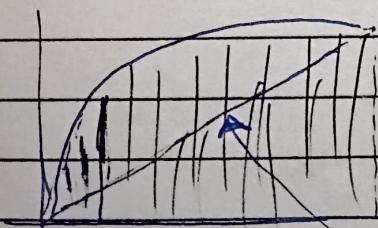
FPR → FP ↓ FPR ↓ "

→ threshold → 0.8

TPR ↑ FPR ↑

but rate of increase of TPR is more than
the rate of increase of FPR

• AUC ROC (Area Under Curve ROC)



• ~~If AUC = 1~~ If AUC = 1
(discriminate b/w +ve & -ve class
perfectly)

• If AUC = 0.5

• If AUC = 0
↳ Perfectly wrong