

## ROC curve

↳ used in "Threshold select"

→ Supervised ML

Regression  
o/p → numerical

classification  
o/p → categorical

→ ROC curve is specifically used in binary classifier

→ Suppose -

iq | cgpa | placed

Now we run test data -

iq	cgpa	placed	Predict <sup>n</sup>	Pred. Prob
-	-	0	1	0.45
-	-	1	0	0.39
-	-	0	0	0.61

Our model doesn't give 0 or 1 rather

it gives us predict<sup>n</sup> 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  $\approx 0.75$

prob  $> 0.75 \rightarrow$  spam  
else  $\rightarrow$  Not spam

• Confusion matrix -  
Predicted

Actual

	1	0
1	TP	FN
0	FP	TN

• True Positive Rate (TPR) (Benefit)

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

out of all spam emails  
how much we have  
marked spam

we want to maximize it.

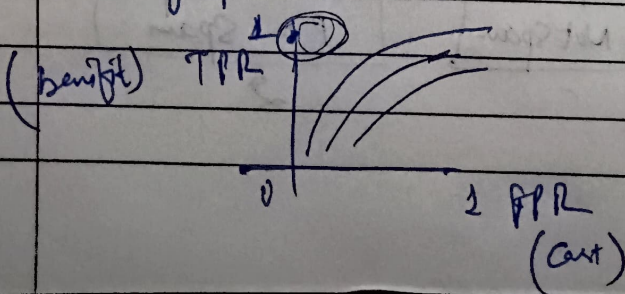
• False Positive Rate (FPR) (Loss) (Cost)

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

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

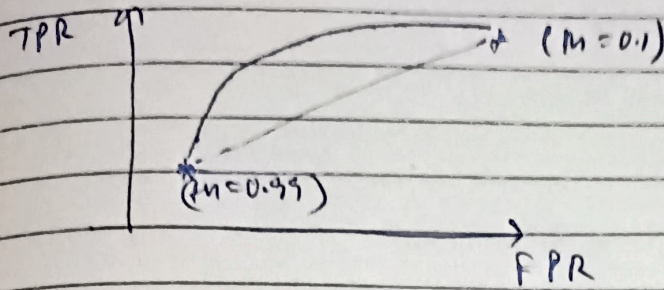
we want to minimize it.

$\rightarrow$  Graph b/w TPR & FPR is called ROC curve



we want  
TPR  $\approx 1$  & FPR  $\approx 0$

• 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

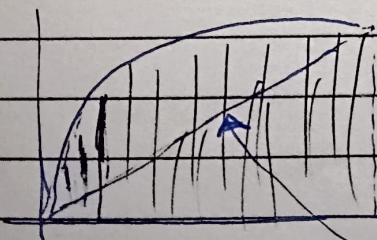
FPR → FP ↓      FPR ↓ "

→ threshold ⇒ 0.5

TPR ↑ & FPR ↑

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

• AUC ROC (Area Under Curve ROC)



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

• If AUC = 0.5

• If AUC = 0  
↳ perfectly wrong