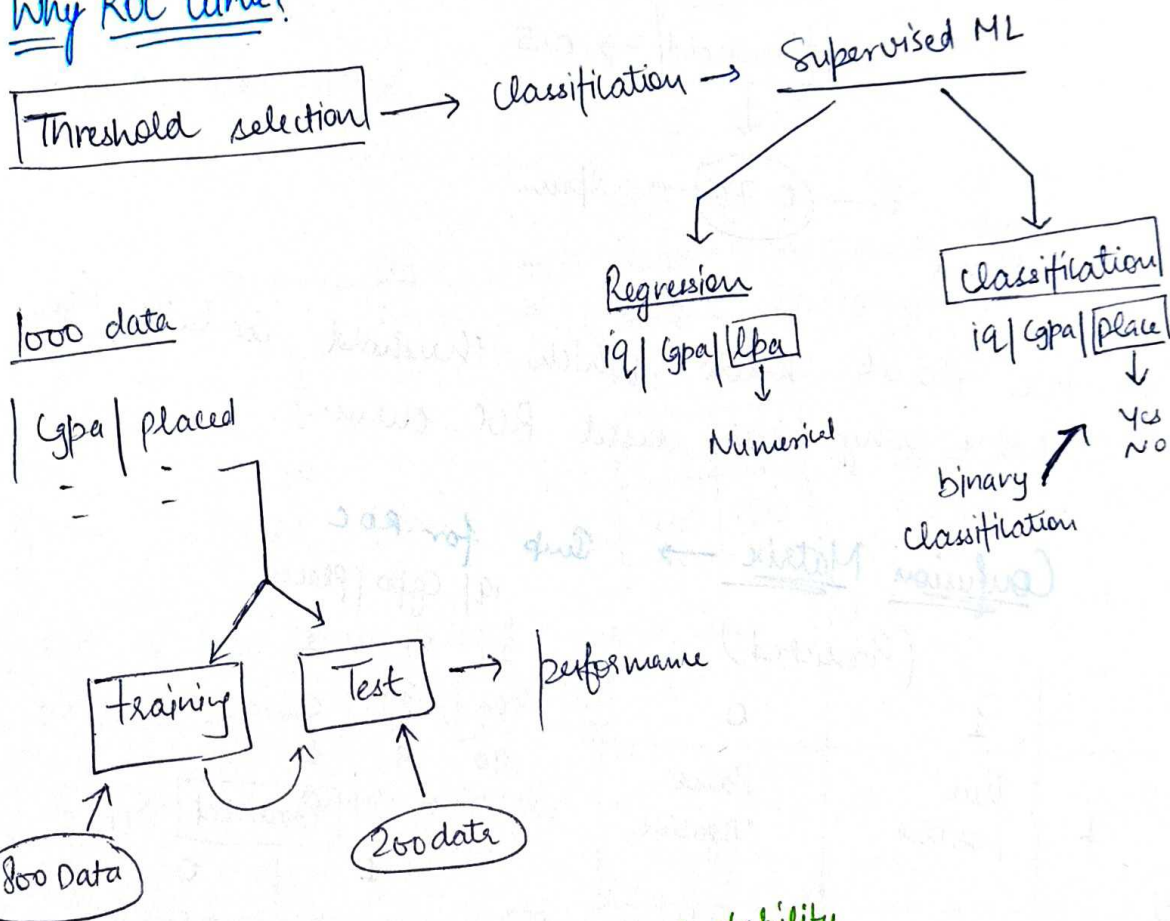


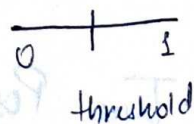
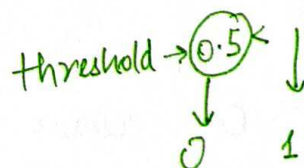
ROC - AUC Curve in Machine Learning

①

Why ROC curve?



iq	Gpa	placed	Prediction	Pred probability
7	70	1	0	0.45
8	80	0	0	0.39
9	90	1	1	0.61



* Not everytime 0.5 → Threshold can work. for ex → email spam classification

email → new email

↓
spam or not spam

binary classification → 2 mistakes

①

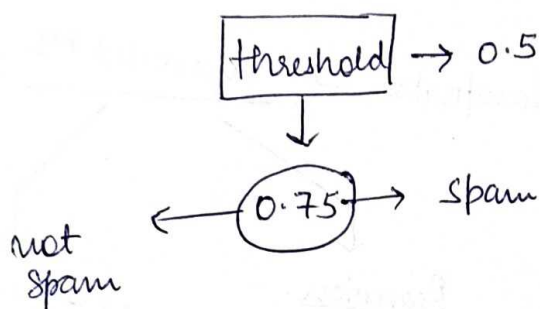
Actual data → spam
Predict → Not spam

②

Actual data → Not spam
Predict → spam

② mistake is more dangerous than mistake ①

threshold from 0.5 to 0.75



* we don't know which threshold is better so, that's why we used ROC curve.

Confusion Matrix \rightarrow Bmp for ROC

		(Predicted)	
		1	0
(Actual)	1	True positive	False Negative
	0	False Positive	True Negative

iq / cgpa / placed

70	7	1
80	8	0
90	9	1

		Predicted	
		1	0
Actual	1	100	5
	0	15	80

True Positive Rate (TPR) \rightarrow Benifit

		Predicted	
		1	0
Actual	1	TP	FN
	0	FP	TN

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

Intuition:

(2)

email spam classifier

1 → spam

0 → not spam

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

out of all the spam emails
Hum kitne emails ko spam
label kar paye

test → 200 emails
↓
100 spam 100 not spam

	1	0
1	80	20
0	20	80

$$TPR = \frac{80}{80 + 20} = \frac{80}{100}$$

* out of 100 spam email, we can label 80 email
as a spam. \Rightarrow $TPR = 80\%$

TPR → Max ↑↑ → good

False Positive Rate (FPR) → Cost

Actual	Prediction	
	1	0
1	TP	FN
0	FP	TN

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

out of ^{all} Not spam emails
kitne email ko spam
label kiya.

Intuition:

Email spam classifier

1 → spam

0 → not spam

$\left[\begin{array}{l} \text{TPR} \rightarrow \text{Benefit} \uparrow \uparrow \\ \text{FPR} \rightarrow \text{Cost} \downarrow \downarrow \end{array} \right]$

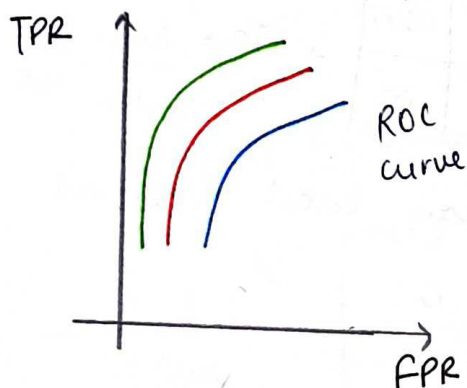
Best case \rightarrow

$$\text{TPR} = \frac{100\% \cdot 0 \cdot 1}{\downarrow \text{FN} = 0}$$

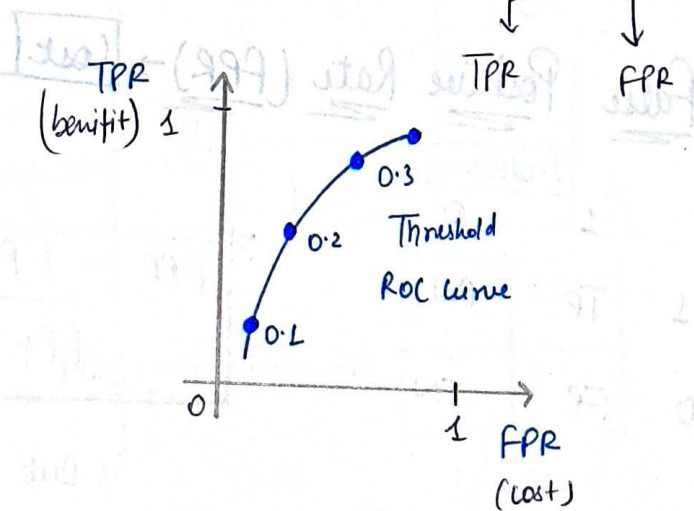
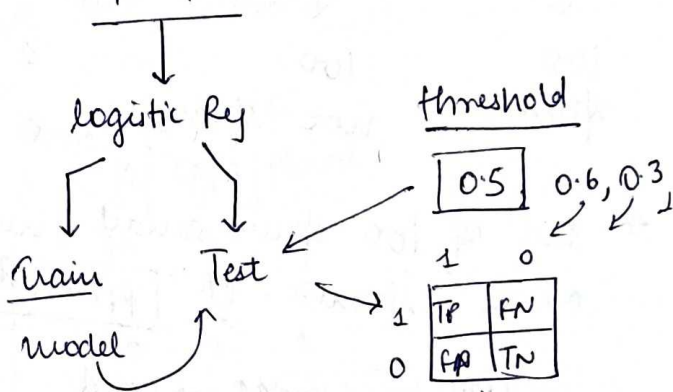
$$\text{FPR} = \frac{0 \cdot 0}{\downarrow \text{FP} = 0}$$

ROC Curve

TPR \leftrightarrow FPR



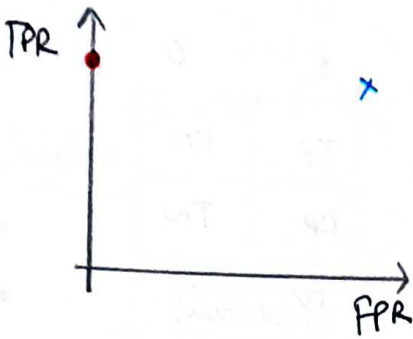
cgpa | iq | placement



Different Cases

3

Case-I



	1	0	
1	TP ↑	FN ↓	Actual
0	FP ↑	TN ↓	
	Predicted		

email spam classifier

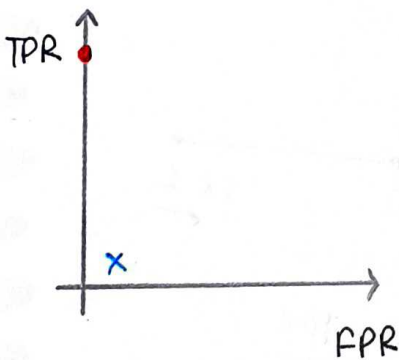
small → $0.1 \approx 0$
↑
Threshold

decrease threshold → FPR ↑
TPR ↑

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

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

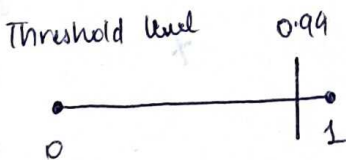
Case-II



email spam classifier

large → $0.99 \approx 1$

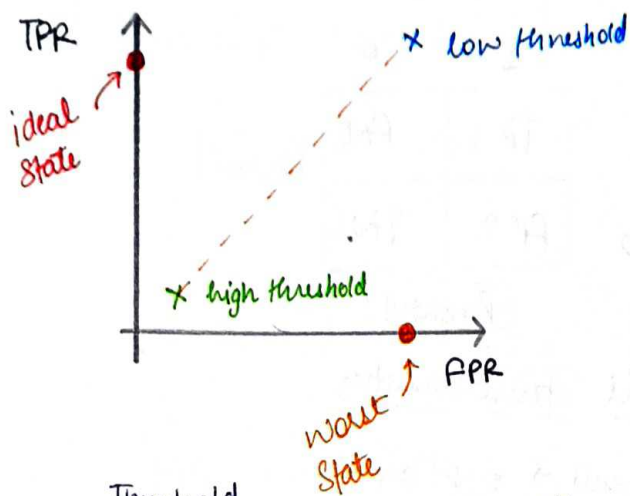
	1	0	
1	TP ↓	FN ↑	Actual
0	FP ↓	TN ↑	
	Predicted		



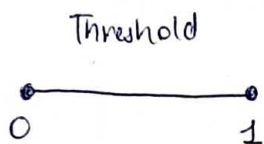
increase threshold → FPR ↓
TPR ↓

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

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

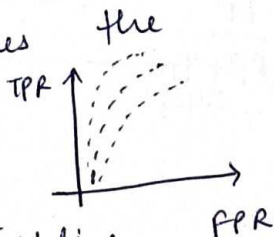


	1	0	
1	TP	FN	Actual
0	FP	TN	
	Prediction		



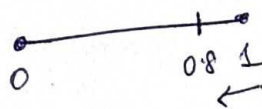
* Before 2 case, $TPR \propto FPR$ but this is not same for everytime. \times (high threshold) and \times (low threshold) line not same everytime.

* Sometimes or many times the line is not straight line



Change in Threshold

Threshold \rightarrow $[0.8]$ \rightarrow relax condition



TPR \rightarrow \uparrow increasing

FPR \rightarrow also increasing but not same rate as TPR.

