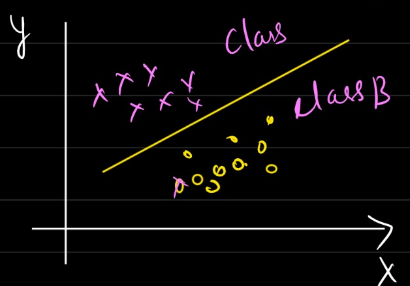
Evaluation metrics for classification problems

1. Confusion matrix
2. Accuracy | misclassification rate
3. Precision
4. Recall
5. F-beta score
6. True Positive Rate (Sensitivity)
7. False Positive Rate
8. True Negative rate (specificity)
9. ROC AUC
10. Precision - Recall | sensitivity - specificity tradeoff.



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

$\downarrow$ 0-1

$\begin{cases} \geq 0.5 & \text{— class 1} \\ < 0.5 & \text{— class 0} \end{cases}$

1. Confusion matrix

		Actual value: 1	Actual value: 0
Predicted value	1	TP	FP
	0	FN	TN

TP — True Positive
FP — False positive
FN — False Negative
TN — True Negative

X_1	X_2	y_{act}	y_{pred}	
—	—	0	1	→ Wrong Classification
—	—	1	1	→ Correct Classification
—	—	0	0	→ Correct classification
—	—	1	0	→ Wrong "
—	—	0	1	→ Wrong "
—	—	1	0	→ "
		1	1	→ "

		1	0	y_{act}
	1	1	2	
y_{pred}	0	2	1	

2. Accuracy → How many are correctly predicted from all the data points

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

Pred \ actual	1	0
1	TP	FP
0	FN	TN

→ Correctly classified DP's – TP & TN.

180

y_{pr} \ y_{act}	1	0
1	100 TP	20 FP
0	10 FN	50 TN

$$Acc = \frac{100+50}{100+20+10+50} = \frac{150}{180} = 0.83$$

* Misclassification rate → opposite of accuracy

$$\frac{FP+FN}{TP+FP+FN+TN} \Rightarrow 1 - \text{Accuracy} = 1-0.83 = 0.17$$

Why Precision?

1000 datapoints → 900 class 1, → 100 class 0 } imbalanced dataset

$$\text{Accuracy} = \frac{900+0}{1000} = 0.90 = 90\%$$

Pred \ Act	1	0
1	900	100
0	0	0

③ Precision $\Rightarrow \frac{TP}{TP+FP}$ } → Out of all actual values how many are correctly predicted.

y_{pred} \ y_{act}	1	0
1	✓TP	✓FP
0	FN	TN

⇒ FP is important here

* for class 1, out of all predicted ones, how many are actually ones

$$\text{Precision} = \frac{TP}{TP+FP}$$

④ Recall – $\frac{TP}{TP+FN}$ { out of all predicted value, how many are correctly predicted with Actual value.

		1	0	actual
Pred	1	$\boxed{TP}$	FP	
	0	FN	TN	

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

Usecase 1: Spam Classification

Text ⇒ mode ⇒ Spam | Not spam.

		1	0	Actual
Pred	1	TP	FP	
	0	FN	TN	

TP ⇒ { Mail – Spam (1), model – Spam (1) } Accurate case

TN ⇒ { Mail – Not spam (0), model – Not spam (0) } Accurate case

FP ⇒ { Mail → Not a spam, model – Spam } (Blunder) wrong Prediction

FN { Mail – Spam, Model – Not a spam } wrong Prediction

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

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

Here FP becomes important.

⇓

Precision

↳ we want a model with high Precision.

Usecase-2 FN is important.

		1	0	Act
Prd	1	TP	FP	
	0	FN	TN	

TP ⇐ Actual – Diabetes, mode – Diabetes } Correct

TN ⇐ Actual – Not diabetic, Model – " " } Correct

FP ⇐ Actual – No Diabetic, Model – Diabetic } wrong

FN

FN ⇐ Actual – Diabetes
model – No diabetes } wrong
Blunder

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

* internal assignment

both the cases FN will be important } → Conviction of a person in the court trial
→ Stock market will crash or not.

(5) F-Beta score :

$$(1+\beta^2)\ \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(1) if FP and FN both are important

$\beta = 1$

$$F_1 \text{ score} = \boxed{2 \times \frac{P \times R}{P + R}} \rightarrow \text{Harmonic mean of P \& R}$$

(2) if FP is more important than FN

$\beta = 0.5$

$$F_{0.5} \text{ Score} = (1 + 0.25)\ \frac{P \times R}{P + R}$$

(3) if FN is more important than FP

$\beta = 2$

$$F2 \text{ score} = (1 + 4)\ \frac{P \times R}{P + R}$$

⑥ True Positive Rate (TPR)

— Out of all actual 1, it is actually Predicted One

		1	0	Act
pred	1	TP	FP	
	0	FN	TN	

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

→ Also Called as Sensitity, Recall.

⑦ False Positive Rate — when its actual 0, how often it is predicted 1.

		1	0
pred	1		FP
	0		TN

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

⑧ True Negative rate ⇒ Out of all actual 0, it is actually Predicted No

		1	0 actual
pred	1	TP	FP
	→ 0	FN	TN

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

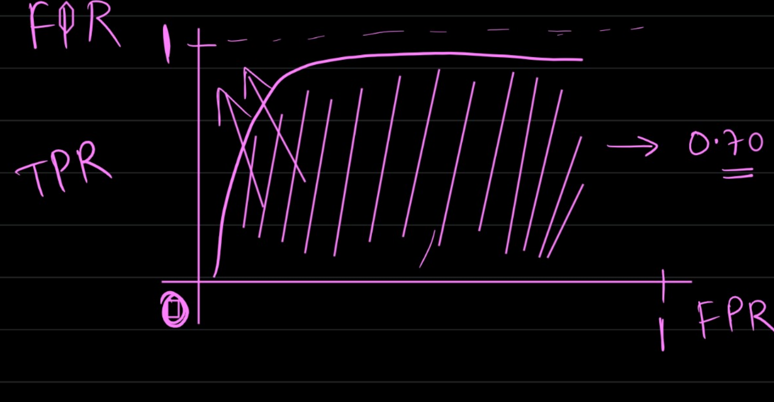
$$TNR = 1 - FPR$$

Specificity

→ It is also known as Specificity

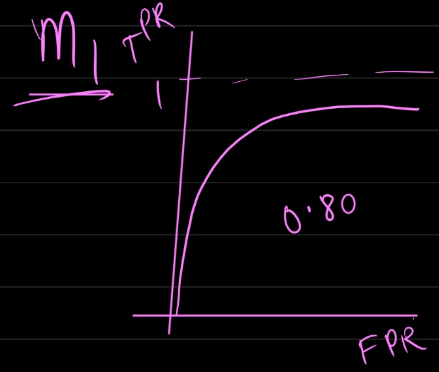
⑨ ROC-AUC — Reciever operating Characteristics Area under Curve

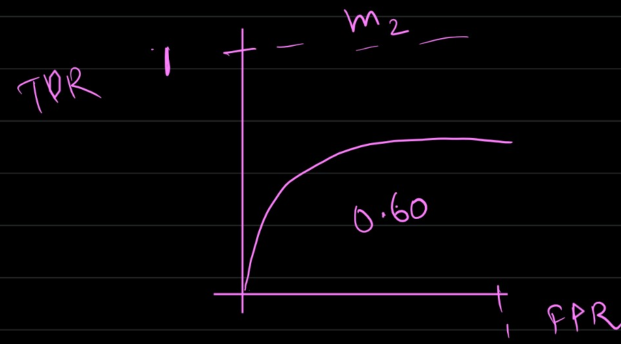
TPR vs FPR



→ 0.70 → model will correctly predict 70% of the time

Higher the AUC better the model will be





m_1 is better than m_2

AUC↑ for m_1

(10) Precision – Recall – Accuracy tradeoff

or

Sensitivity – Specificity – Accuracy tradeoff.

x_1	x_2	y_{act}	$y_{pred\ prob}$
–	–	0	0.3
–	–	1	0.4
–	–	1	–
		0	–
		1	–

≥ 0.5 — Class 1

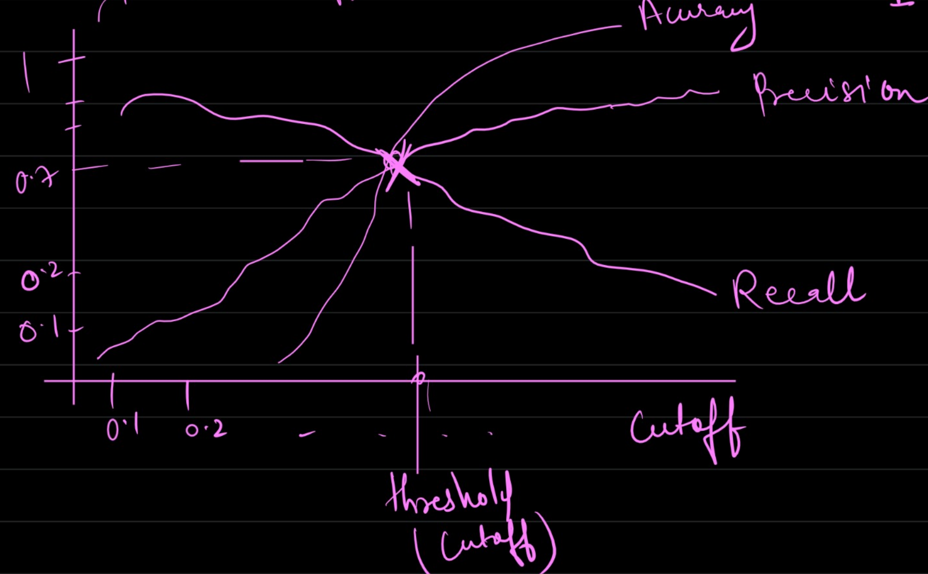
< 0.5 — Class 0

* Diabetic | not diabetic

↓

1 → cutoff very very hight → 0.9

P–R tradeoff



X	Y prob	Y pred cutoff 0.1	Y pred cutoff 0.2
	0.1	0	0
	0.2	1	0
	0.3	1	1
	0.1	1	1
	0.67	1	1
	0.25	1	1
	0.02	0	0

$0.1 \geq$ – class 1

$<$ – class 0

Precision, Recall, accuracy.