Lecture - Classification Metrics - 2

Revision

Problem with accuracy - Doesn't work imbalance data.

CM

		$\hat{y}$ = 0	$\hat{y}$ = 1
y	$-$ 0	TN	FP
	$+$ 1	FN	TP

$$\text{Precision} = \frac{\text{Correctly pred. positive}}{\text{Predicted Positive}} = \frac{TP}{TP + FP}$$

↓

How precisely your models predicts a positive class.

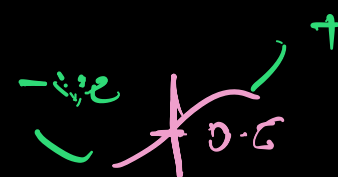
$$\text{Recall}^{+} = \frac{\text{Correct Positive Pred}}{\text{Actual Positives}} = \frac{TP}{TP + FN}$$

How many positives is the model able to recall / remember.

My personal Hack to remember Precision and Recall

$$\text{Precision}^{+} = \frac{TP}{\text{Pred +ve}}$$

$$\text{Recall}^{+} = \frac{TP}{\text{Real +ve}}$$



Agenda :

F1-score — P and R

Optimise Threshold.

ROC-AUC Curve

Sensitivity, Specificity — Healthcare/ Medical Interview.

F1-score

Doesn't just relies on ML + Rules.

Detect fraudulent transaction

Fraud(1, +ve) , Non-Fraud (-ve, 0)

(+ve)

1. Predicted as fraud but are legitmate

Wrong classification, Pred: +ve

FP ↑

- Bad user experience
- Fin loss to business.

2. Predicted as legitmate but are fraud

FN ↑

- Loss to banks in transaction pipeline

FP ⟺ FN

Disclaimer: Digital Notes Fraud (-ve), N.Fraud (+ve)

$\Downarrow$

Both FP and FN are required to be reduced.

P↑ R↑

	Precision	Recall
M1	0.3	0.8
M2	0.2	0.9
M3	0.7	0.4

Which one is best? Not sure.

Equal weightage to both P and R

~~Arith.~~ Mean ?? Harmonic

Arithm. Mean AM	$\frac{x_1+x_2+x_3}{3}$	Same units
Geometric Mean GM	$\sqrt[3]{x_1 \cdot x_2 \cdot x_3}$	Different units
Harmonic Mean HM	$\frac{3}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}}$	Rates/ Proportion.

P and R proportions $P = \frac{TP}{PP}$, $R = \frac{TP}{RP}$ Proportion

☆☆☆

$$\text{f1-score} = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \boxed{\frac{2P.R}{P+R}}$$

Equal weightage to P and R

	Precision	Recall	f1-score
M1	0.3	0.8	0.44
M2	0.2	0.9	0.33
M3	0.7	0.4	0.51

Best model

Depending on Business case

Assume

P → 0.6

R → 0.4

Weighted F1 score $= \cancel{\dfrac{2}{w_1 \frac{1}{P} + w_2 \frac{1}{R}}}$

Out of syllabus.

f-beta score $= \dfrac{2P.R}{\beta^2 P + R}$ ← No need remember

Regulates how much weight we give to P vs R.

Interview Question

How will you decide the value of β? — Hyper param

~~- Cross Validation~~ ↑ Performance.

- Business expectation PART OF METRIC

should not be optimized

Sensitivity (Imp Medical/Healthcare)

COVID Days - Screening Tests

Rapid antigen - less trusted

RT-PCR - more trusted.

Rapid Antigen.

+ve → Def COVID

-ve → May or not have COVID
check RTPCR.

Sensitivity and Specificity of the test

Sensitivity - Ability to identify the disease if it exists. (unhealthy)

No Disease (-ve) Disease (+ve)

Recall $= \dfrac{TP}{\text{Real Positives}} = \dfrac{TP}{TP+FN}$

=

Sensitivity

How sensitive is the test to detect the disease in a patient who actually suffers from the disease.

True +ve Rate

$\dfrac{TP}{RP}$ = Recall = Sensitivity = TPR

Rates

rates = denominator is real.

Sensitive Test - TP ↑, FN ↓

TPR ↑, FNR ↓

$$FNR = 1 - TPR$$

$$= 1 - \left(\frac{TP}{TP+FN}\right) = \frac{FN}{TP+FN} = \frac{FN}{RP}$$

$$\boxed{\begin{aligned} TPR &= TP/RP \\ FNR &= FN/RP \end{aligned}}$$

False Negative Rate

Miss Rate ☆☆☆

How many patients have we missed to detect the disease.

ACTUALLY DISEASED

Why sensitivity important?

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

TP↓ FN↑ Sensitivity/ Recall / TPR.

Higher chances of misses.

Progressive Diseases and deadly diseases

CANCER

Treatment doesn't have major side effects

Specificity

$$\text{Sensitivity} = \frac{TP}{\text{Real +ve}}$$

Specificity is actually same as sensitivity BUT for negative class.

$$\text{Specificity} = \frac{TN}{RNeg} = \frac{TN}{TN+FP}$$

Real.

$$= TNR$$

True Negative Rate ✓

lets say your model is not specific

TN↓, FP↑ - a lot healthy patients are being falsely diagnosed with disease.

- social stigma, anxiety - HIV+
- Higher treatment costs.
- Medication is higher side effects

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

$$= \text{True Negative Rate}$$

$$FPR = 1 - \text{Specificity}$$

$$= 1 - \frac{TN}{TN+FP} = \frac{FP}{TN+FP}$$

demo.

RN

+ve

Sensitivity = Recall = TPR

$= \frac{TP}{TP+FN} = \frac{TP}{Real\ P}$

Opp.

Miss Rate = 1 - Sensitivity = FNR.

$= \frac{FN}{TP+FN} = \frac{FN}{Real\ P}$

Defined for +ve class.

-ve

Specificity = TNR

$= \frac{TN}{TN+FP} = \frac{TN}{Real\ Neg}$

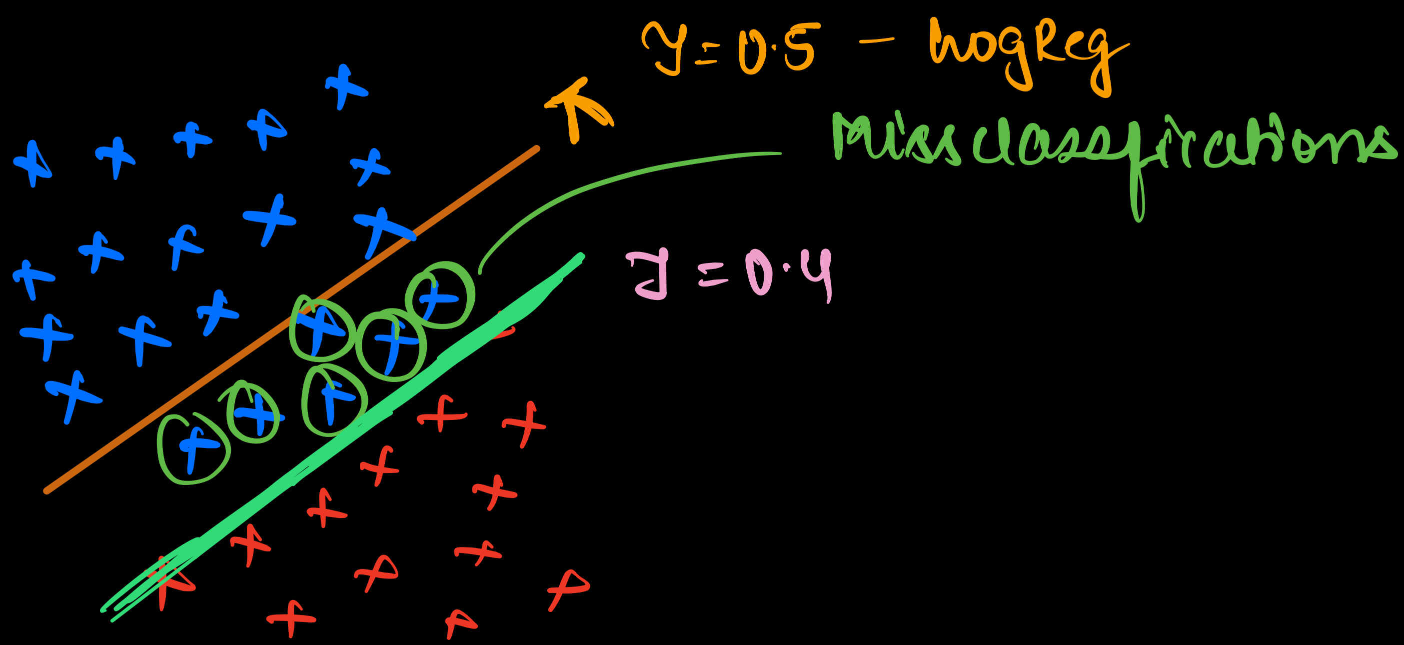
FPR = 1 - Specificity

$= \frac{FP}{Real\ Neg}$

Defined for -ve class.

ROC Curve

Cancer / No Cancer



Hyperparameter Tuning C based on VAL data.

↓

?? What values of C will you use?

Step1. For val dataset

$P(y^i = 1 | x^i)$

X	Y	$\sigma(w^T x)$
$x^{(1)}$	$y^{(1)}$	$p^{(1)}$
$x^{(2)}$	$y^{(2)}$	$p^{(2)}$
$x^{(3)}$	$y^{(3)}$	$p^{(3)}$
$\vdots$	$\vdots$	$\vdots$
$x^{(m)}$	$y^{(m)}$	$p^{(m)}$

sort the e.g.'s based on descending order of prob.

X	Y	P
x^1	1	.68
x^2	1	.94
x^3	0	.30
x^4	1	.92
x^5	0	.70
x^6	0	.20

$\longrightarrow$

X	Y	P
x^2	1	0.94
x^4	1	0.92
x^5	0	0.72
x^1	1	0.68
x^3	0	0.30
x^6	0	0.20

Step 2. Use all diff P as diff-threshold

x	y	p	[0.94 $\hat{y}$	[0.92 $\hat{y}$	[0.72 $\hat{y}$	[0.20 $\hat{y}$
x^2	1	.94	1	1	1	1
x^4	1	.92	0	1	1	1
x^5	0	.72	0	0	1	1
x^1	1	.68	0	0	0	1
x^3	0	.30	0	0	0	1
x^6	0	.20	0	0	0	1
			M1	M2	M3	M_m

Step3. TPR and FPR for all diff thresholds

x	y	P	$\hat{y}$ [0.94	$\hat{y}$ [0.92	$\hat{y}$ [0.72	- - -	$\hat{y}$ [0.20
x_2	1	.94	1	1	1		1
x_4	1	.92	0	1	1		1
x_5	0	.72	0	0	1		1
x_1	1	.68	0	0	0		1
x_3	0	.30	0	0	0		1
x_6	0	.20	0	0	0		1

Recall

$$TPR_{M1} = \frac{1}{1+2} = \frac{1}{3}$$

1-specificity

$$FPR_{M1} = \frac{0}{0+5} = 0$$

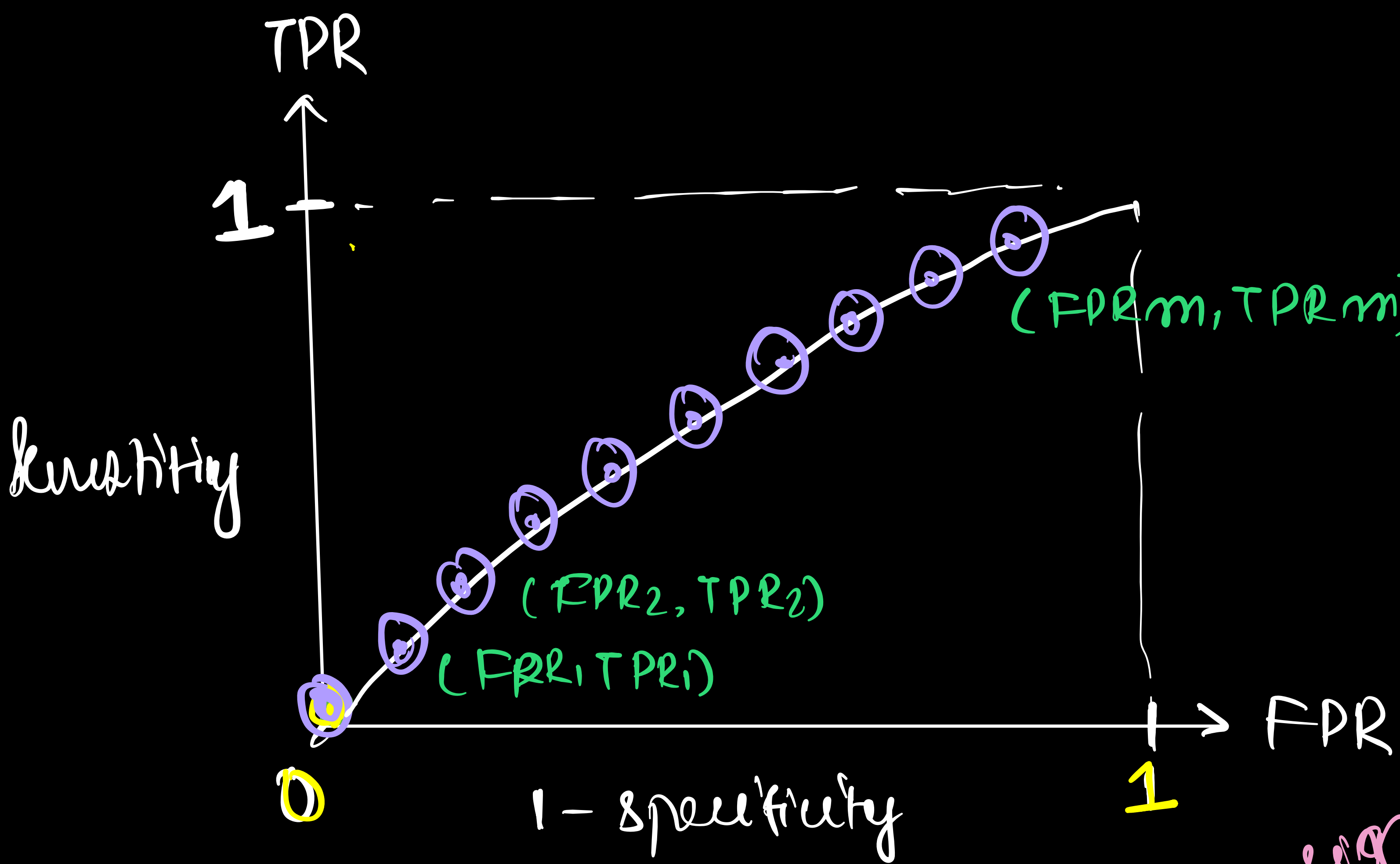
$(1/3, 0)\ (x_2, y_2)\ (x_2, y_2)$.

TPR M2
FPR M2 (x_2, y_2)

TPR M3
FPR M3 (x_3, y_3)

TPR Mn
FPR Mn

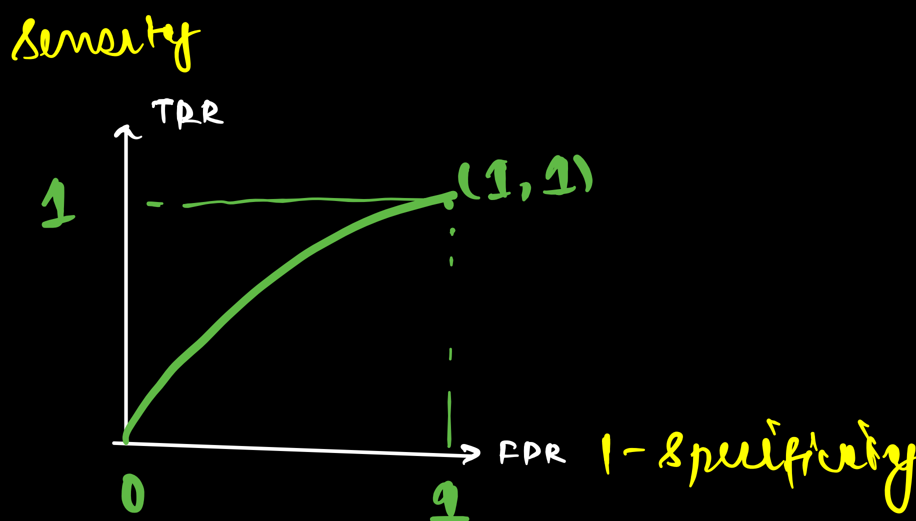
Step 4. Plot a TPR v/s FPR curve



ROC - Receiver Operating Characteristics - signal processing

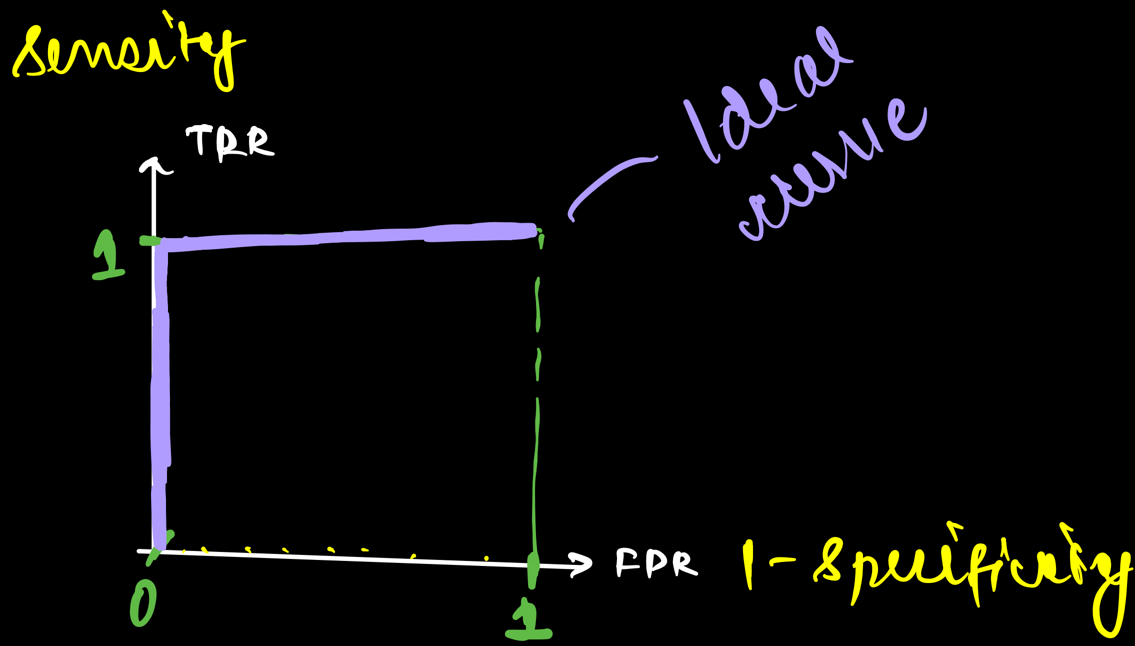
Revision (ROC curve) - AUC (Area under curve)

step1) sorting all the rows basis on P.
step2) calculating $\hat{y}$ based on every $\Gamma = p_i$
step3) calculating TPR and FPR for every $\Gamma = p_i$



ROC curve
(Receiver Operating Characteristics)

Case 1) Ideal Model / Setting

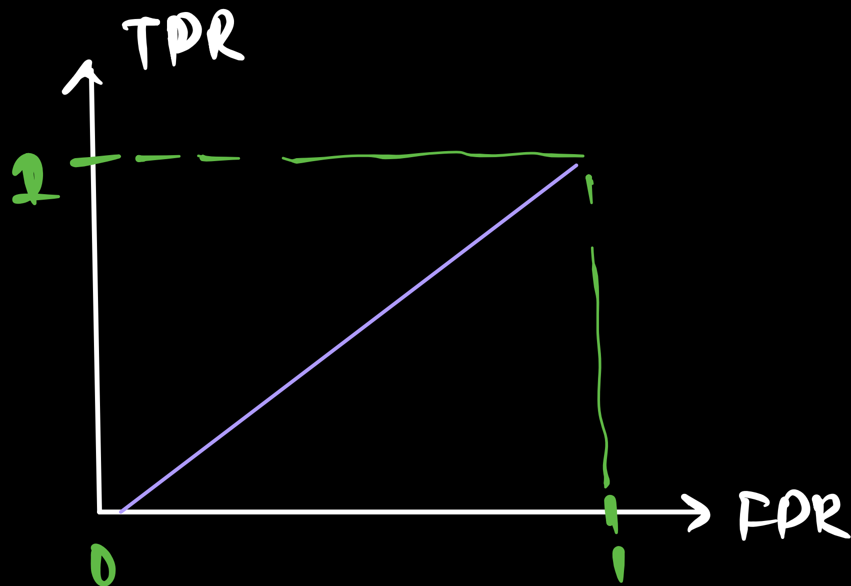


Case 2) Random Model

Assum…

Balanced Data

3 : +ve | 3 : −ve

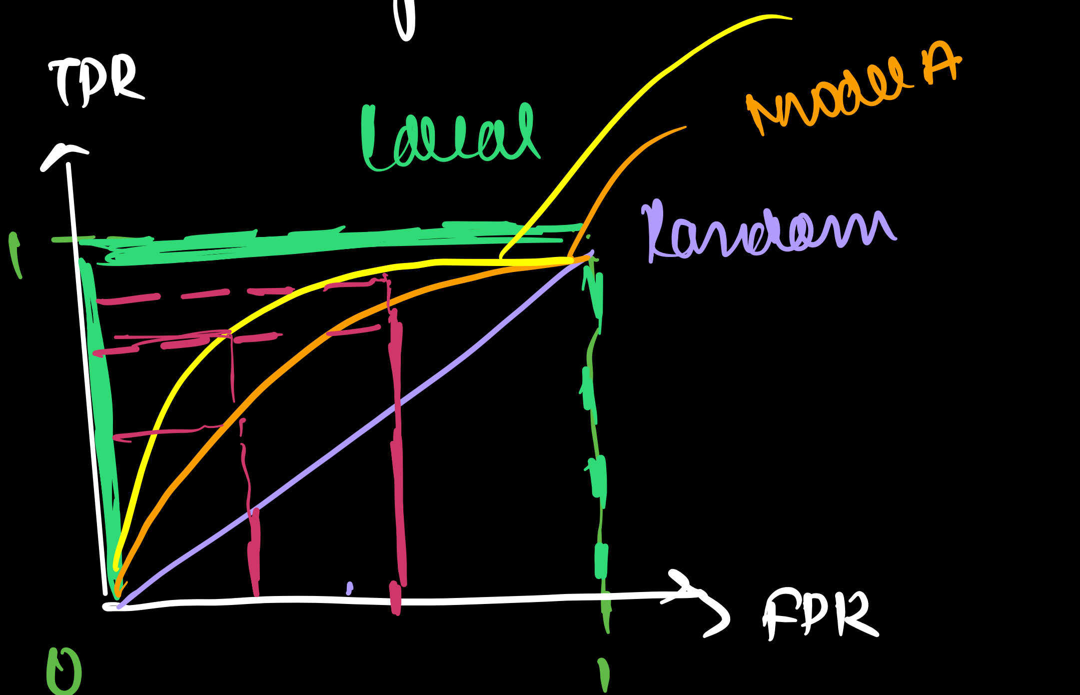


Randomly Assign any class to a label.

50% → 0 K   TPR = x

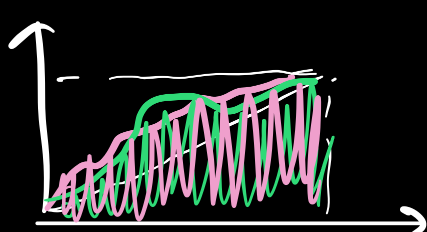
50% → 1 K   FPR = x

Case 3: Typical case



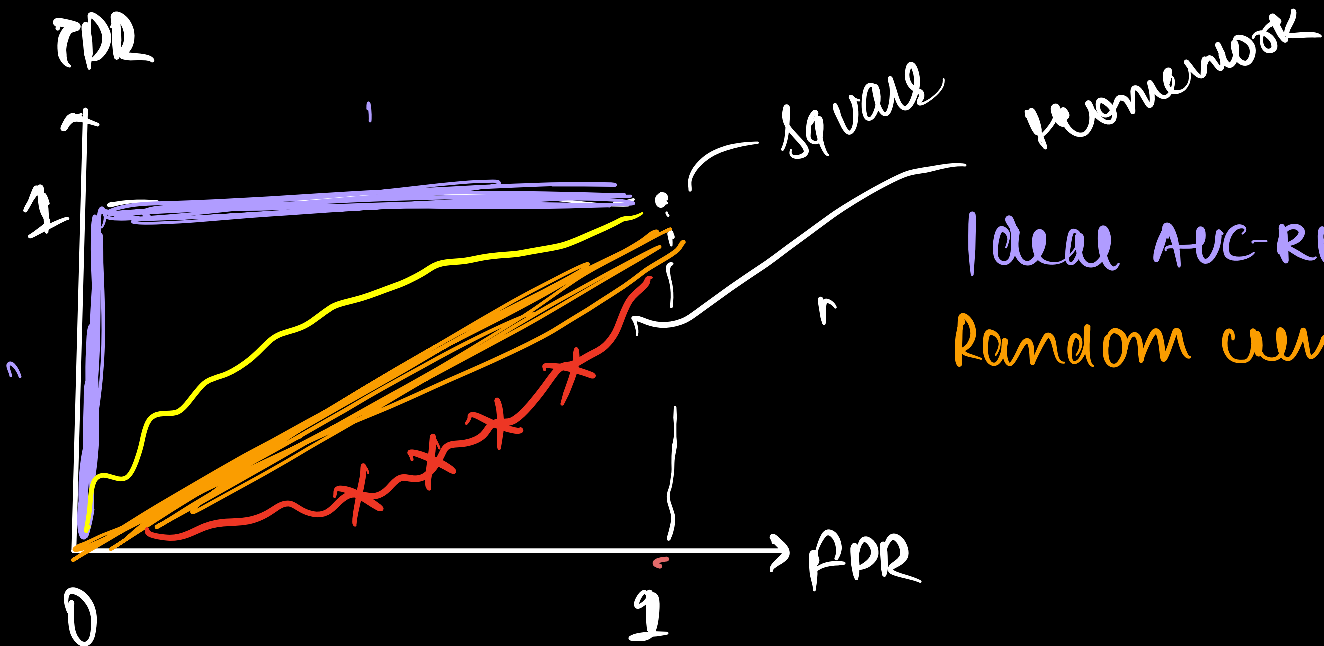
Which one is better?

Model which more above is better



Area under the curve

AUC-ROC curve



Ideal AUC-ROC = 1 unit^2

Random curve = 0.5 unit^2

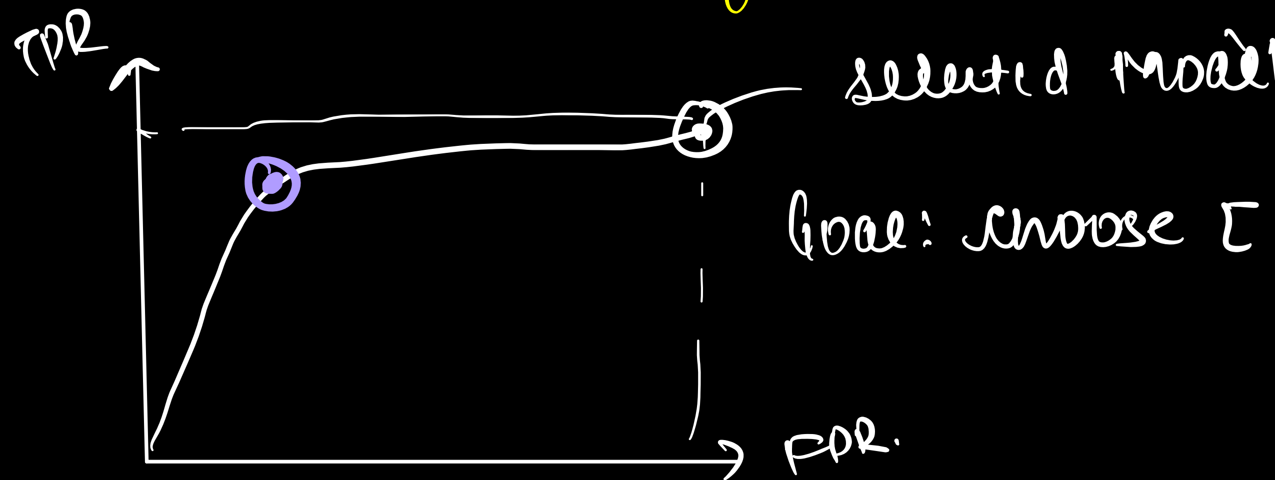
Your model $0.5 < \text{AUC-ROC} < 1$

Why a model can't be worse than a random model?

Homework.

Q How to choose b/w two models ✓

How to choose τ for the selected model?



Goal: choose τ

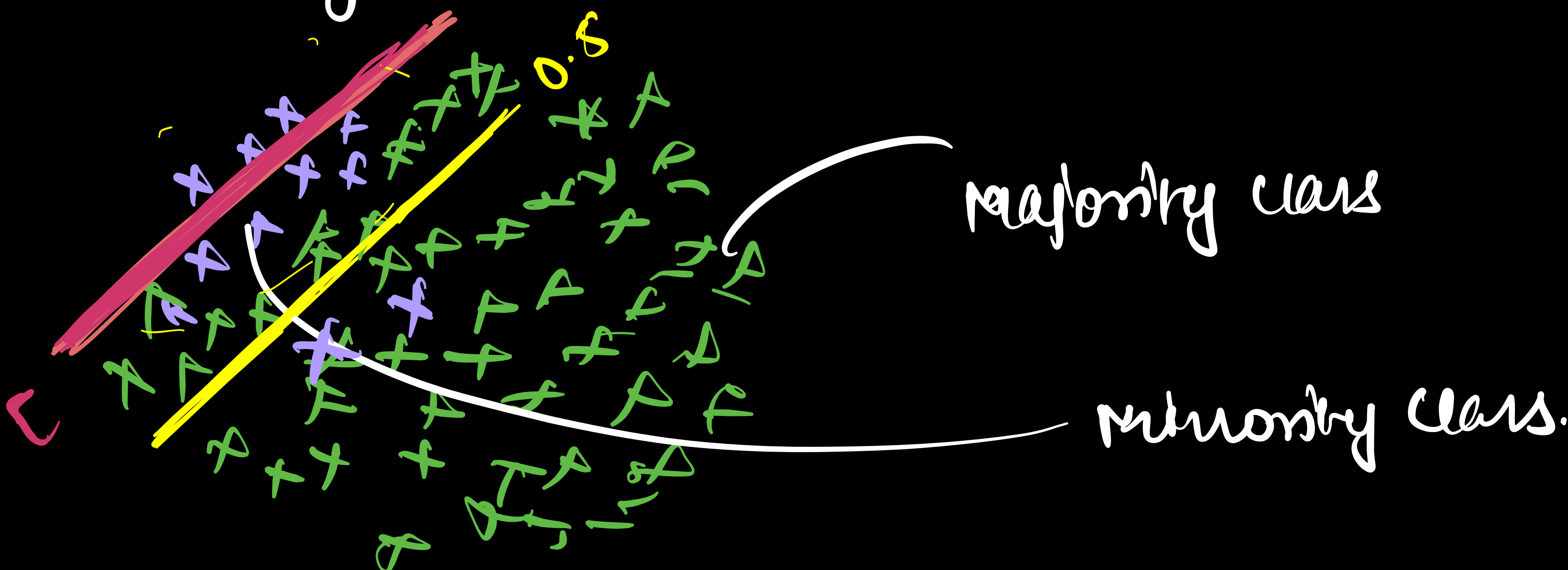
Tipping point of this curve.

Choose a point which actually give ↑↑↑ TPR for ↓↓↓ FPR.

Problem ROC-AUC metric.

Assume Imbalance

Tuning the threshold



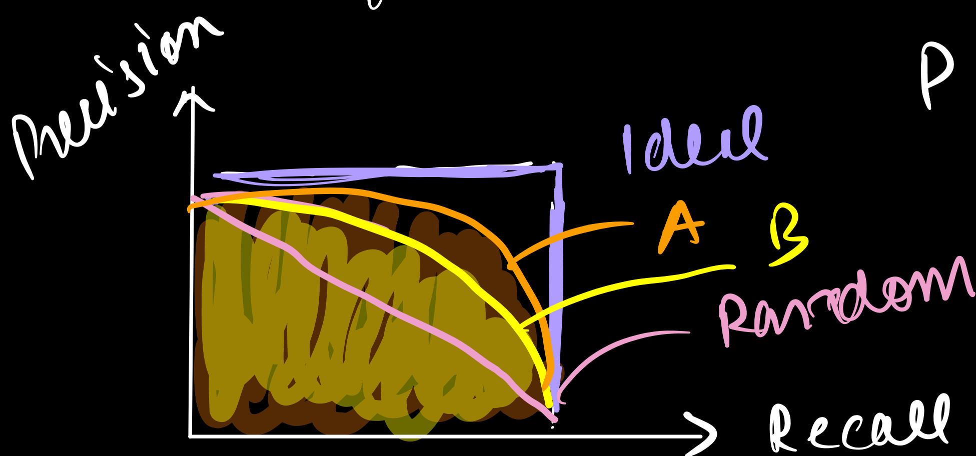
By selecting — we have reduced the misclassification for majority class but increased the misclassification for minority

Major class influence metric more.

⇓

AUC-ROC curve shouldn't be used for highly imbalanced data.

Your choice ⊏ might be influenced highly by performance of majority class.



PR curve. for imbalance data.

$$\text{Mean} = \frac{x_1 + x_2}{\cancel{2}\ 3}$$

Rlar

Va

$$\beta\text{-score} = \frac{2PR}{\cancel{\beta}\,P + R}$$

Near

$$= \boxed{\frac{2P \cdot R}{1.5\,P + R}}$$

Bamber

FP

FN

ML → P R F1

Classification

Optimise

Revenue

Profit

Good . Revenue

↓

Expense

Best m

P

R