* To decide which one should we look for between Sensitivity & Specificity

* If identifying +ve is more important to us, then we will select Algo that has high Sensitivity

* If correctly identifying -ve is more important then we will select algo that has high Specificity

Precision & Recall :->

-> Used for Information Retrieval.

-> Google Search Engine ->

-> Query fired

-> Have millions of Related Records

-> From these top 10-100 records are returned.

$$\text{Precision} = \frac{\text{Correctly Predicted +ve}}{\text{Total +ve Predicted}} = \frac{TP}{TP + FP}$$

Range (0 — 1)

$$\text{Recall} = \frac{\text{Correctly Identified +ve}}{\text{Total Actual +ve}} = \frac{TP}{TP + FN}$$

(TPR)
(range 0 - 1)

* Ideally we want precision to be high (i.e 1) for a good classifier

(range 0 - 1)

* Ideally we want precision to be high (i.e 1) for a good classifier

$$\text{Precision} = 1 = \frac{TP}{TP+FP} = 1 \Rightarrow \text{when } \boxed{FP = 0}$$

* Ideally we want Recall to be very high (i.e 1) for a good classifier

$$\therefore \text{Recall} = 1 = \frac{TP}{TP+FN} \Rightarrow \text{when } \boxed{FN = 0}$$

So Ideally a good classifier has High Precision & Recall

But in Reality there is trade-off

* When we tweak (update) our model to increase one, then the other decreases.

Q) Explain

F1- Score

* In reality we need a metric that takes into account both precision and recall.

* F1 score is a metric that takes into account both precision & recall.

* F1 Score is harmonic mean of Precision & Recall.

$$\text{F1-Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Harmonic mean of two variables a & b

$$H = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2\, a * b}{a + b}.$$

for n variables

$$H = \frac{n}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots \frac{1}{n_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{i}}$$

If F1-Score = 1 ⇒ when Precision = 1, Recall = 1

* When Precision and Recall both are high then F1 Score is High

When to Use F1 Score →

→ Accuracy is not a good metric to use when we have class Imbalance.

Ex → Let say 99% of people visiting site are onlookers and not purchasing anything.

→ Suppose we have a model that predicts that 100% people visiting site are onlookers.

" 1% Error is Acceptable

people visiting site are onlookers.

→ The model is 1% wrong, Generally 1% Error is Acceptable

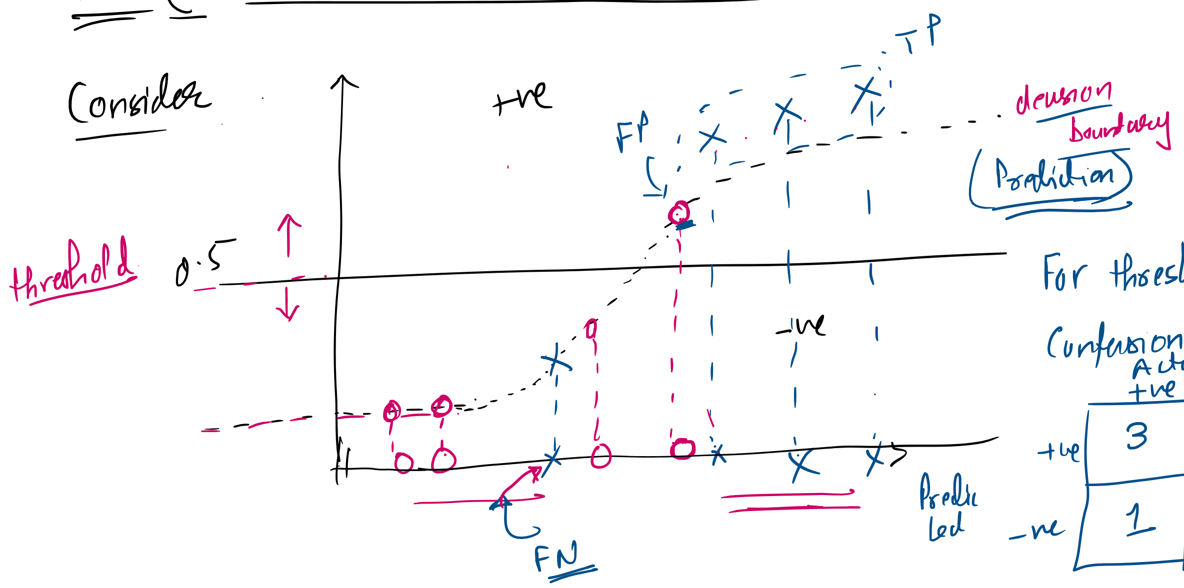
→ But such model in this case is useless

→ In such case instead of accuracy, we will prefer F1-Score.

ROC [Reciever Operator characteristic]

x = Actual +ve
0 => Actual -ve

Consider



For threshold 0.5
Confusion matrix

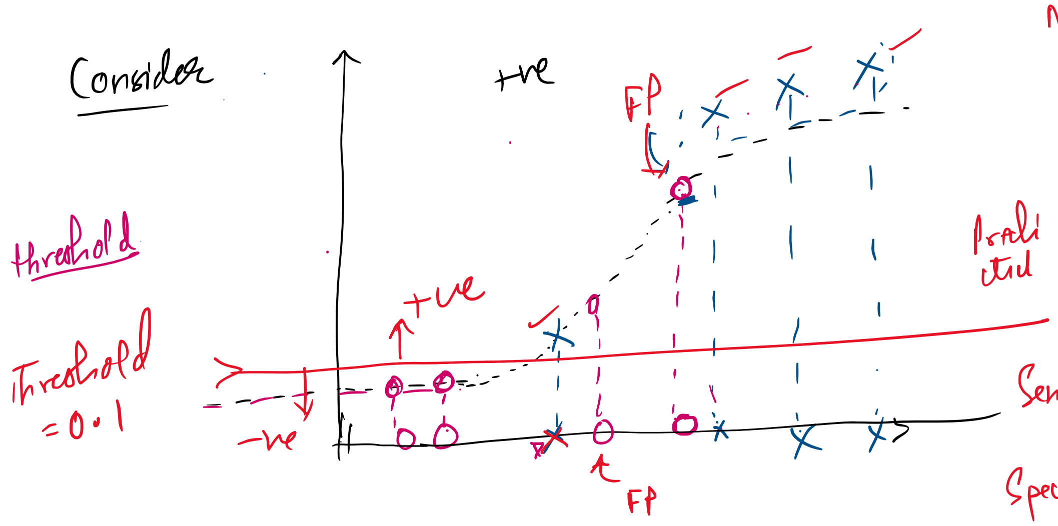
		Actual +ve	Actual -ve
Predicted	+ve	3	1
	-ve	1	3

$$\text{Sensitivity} = \frac{3}{4} \quad \text{TPR}$$

$$\text{Specificity} = \frac{3}{4} \quad \text{TNR}$$

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

Consider



Now threshold = 0.1
Confusion matrix

		Actual +ve	Actual -ve
Predicted	+ve	4	2
	-ve	0	2

$$\text{Sensitivity} = \frac{4}{4} = 1$$

$$\text{Specificity} = \frac{2}{4} = 1/2$$

* Consider for logistic Regression, where we identify a threshold point and prepare confusion matrix and calculate Sensitivity & Specificity.

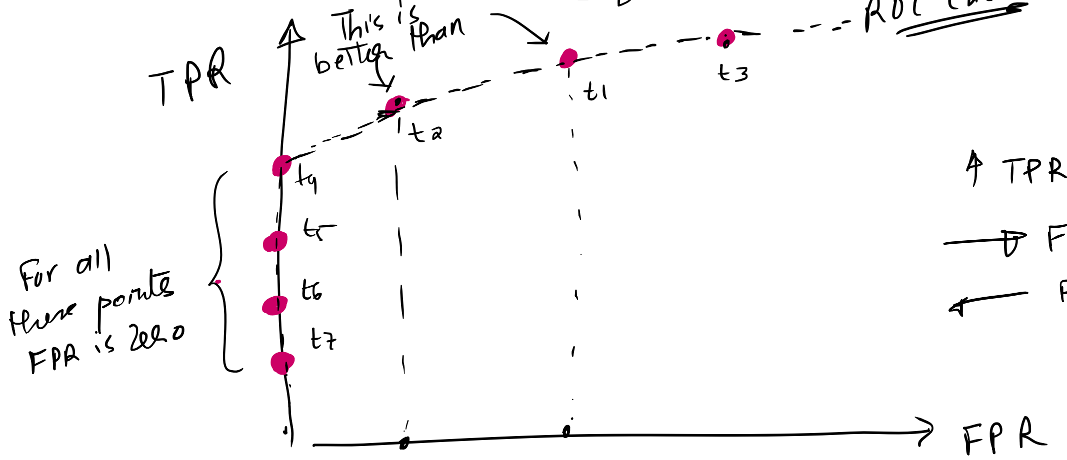
and Calculate Sensitivity & Specificity

* If threshold changes then the Confusion matrix and accordingly the Sensitivity and Specificity changes.

* We can have many such thresholds betⁿ 0 → 1

* We want to analyze the performance at diff threshold. and want to identify the best of it.

* For this we will plot TPR and FPR for diff threshold. → ROC Curve for (Random Forest)



→ From the above ROC Curve for Random Forest model, it will help us to find the best threshold

Q) Explain

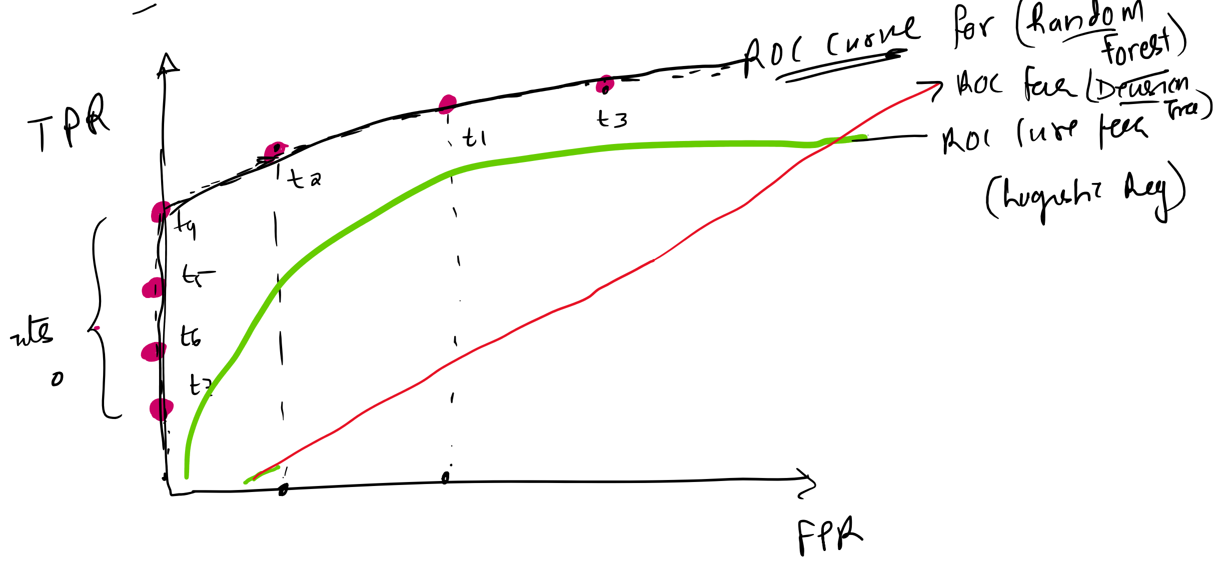
AUC [Area Under Curve] → It is a method to compare ROC for more than one method and will help to judge which one is better.

→ Here will find Area Under ROC for each method.

→ ROC that has maximum Area, the corresponding method will be best.

ROC Curve for (Random Forest)

→ ROC that has maximum Area, the correspon...



From above ROC Curves, we find that there is maximum area under ROC Curve for Random forest, hence the method Random forest will be the best method.

Q) Explain Kappa Statistic.?

* Kappa Statistic or Cohen's Kappa is statistical measure of inter-rater reliability for categorical variable.

* It is used when two/more raters apply a criteria based on a tool to assess whether or not some condition occur.

Ex Let say Two doctors rates whether or not each of 20 patients has diabetes based on symptoms.

* If two raters uses same criteria on same target to evaluate and then their agreement is very high then we will have evidence of reliable rating

* If their agreement is not very high then ⇒
  - → Either criterion tool is not useful
  - → or raters are not trained enough.

* Kappa Statistics correct for $\boxed{\text{chance agreement}}$ and not percent agreement

Ex

Evalor B \ Evalor A	Yes	NO
Yes	35	20
NO	15	40

⇒ 35 times both agreed - said Yes
⇒ 40 times both agreed ⇒ said NO
⇒ 20 time A said NO but B said Yes
→ 15 time A said Yes but B said NO.

$\Rightarrow$ 20 time A said NO but B said Yes

$\Rightarrow$ 15 time A said Yes but B said NO.

Cohen suggested following statistics. $\Rightarrow$

value ≤ 0 $\Rightarrow$ No agreement

0.01 $\rightarrow$ 0.20 $\Rightarrow$ as none to slight

0.21 $\rightarrow$ 0.40 $\Rightarrow$ as fair

0.41 $\rightarrow$ 0.60 $\Rightarrow$ as moderate

0.60 — 0.80 $\Rightarrow$ as substantial

0.81 — 1.00 $\Rightarrow$ perfect agreement.

* Rather than calculating the percentage of items, the raters agreed on Cohen's Kappa attempts to account the fact that raters may happen to agree on some items purely by chance

Ex Two Curators asked to rate 70 paintings.

		Curator C_2 Yes	Curator C_2 NO
Curator C_1	Yes	25	10
	NO	15	20

step ① Calculate Relative Agreement betⁿ Curators.

$$P_o = \frac{\text{Both said Yes} + \text{Both said NO}}{\text{Total}} = \frac{25+20}{70} = 0.6429$$

step 2 > Calculate hypothetical probabilities of chance Agreement betⁿ Curators.

$$P(\text{Yes}) = C_1(\text{Yes}) * C_2(\text{Yes}) = \frac{(25+10)}{70} * \frac{(25+15)}{70} = 0.2857$$

$$P(Yes) = \frac{C_1(Yes)}{Total\ Res} * \frac{C_2(Yes)}{Total\ Res} = \frac{(25+10)}{70} * \frac{(25+15)}{70} = \underline{0.2857}$$

$$P(No) = \frac{C_1(No)}{Total\ Res} * \frac{C_2(No)}{Total\ Res} = \frac{(15+20)}{70} * \frac{(10+20)}{70} = \underline{0.214285}$$

$$\underline{P_e} = P(Yes) + P(No) = 0.2857 + 0.214285 \approx \underline{0.5}$$

Calculate Cohen's Kappa $= K = \dfrac{P_0 - P_e}{1 - P_e} = \dfrac{(0.6429 - 0.5)}{1 - 0.5} = \underline{0.2857}$

It is in range $0.21 \rightarrow 0.40$ so the agreement bet^n two Curator is <u>fair</u>