

- For classification problem with a class output, the confusion matrix gives the counts of correct and erroneous predictions:

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

		Actual	
		1	0
Predicted	1	TP	FP
	0	FN	TN

**I** (points to FP)

**II** (points to FN)

- Classification Error Rate: sum of Type 1 (FP) and Type 2 (FN) Errors (in percentage). Accuracy is 1-(error rate)
- Sensitivity (also called Recall or True Positive Rate): proportion of Total Positives that were correctly identified
- Specificity (also called True Negative Rate): proportion of Total Negatives that were correctly identified

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

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

		Truth	
		P	N
Predicted	P	TP	FP (Type 1)
	N	FN (Type 2)	TN

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

		Truth	
		P	N
Predicted	P	0	0
	N	10	$10^9 - 10$

Out of 1 Billion People there are 10 terrorists

$$\begin{aligned}
 \text{Accuracy} &= \frac{10^9 - 10}{10^9} \\
 &= 1 - 10^{-8} \\
 &= 0.99999 \\
 &\text{or } 99.9999\%
 \end{aligned}$$

## Recall (Sensitivity or TPR)

- $\text{Recall} = \text{TP} / \text{TP} + \text{FN}$
- Recall: Out of all terrorist what fraction did you identify

## Precision:

- $\text{Precision} = \text{TP} / \text{TP} + \text{FP}$
- Out of all the predicted terrorists what fraction were really terrorists.

Label all as not a terrorist

	P	N
P	0	0
N	10	$10^9 - 10$

$$ACC = (10^9 - 10)/(10^9)$$

Label all as terrorist

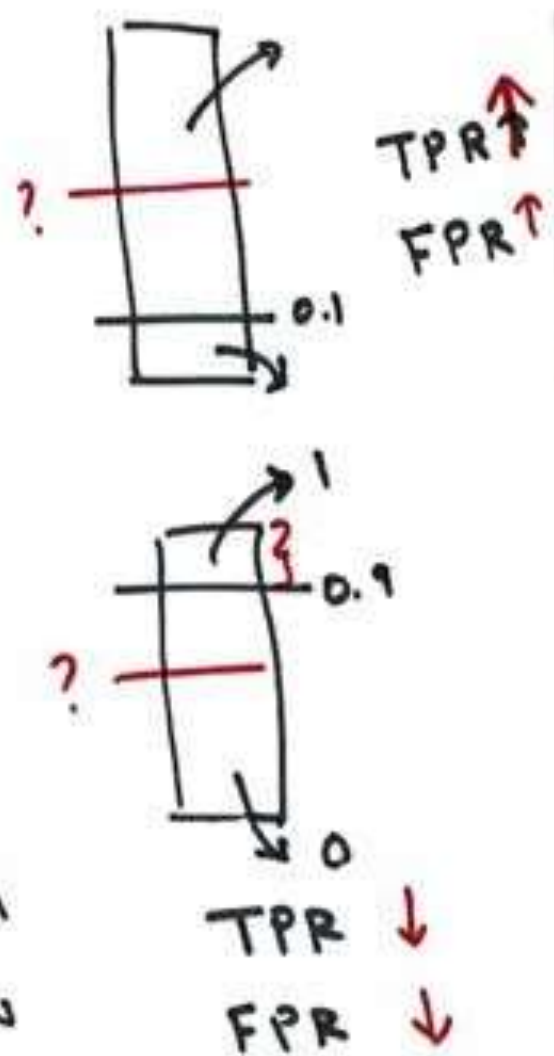
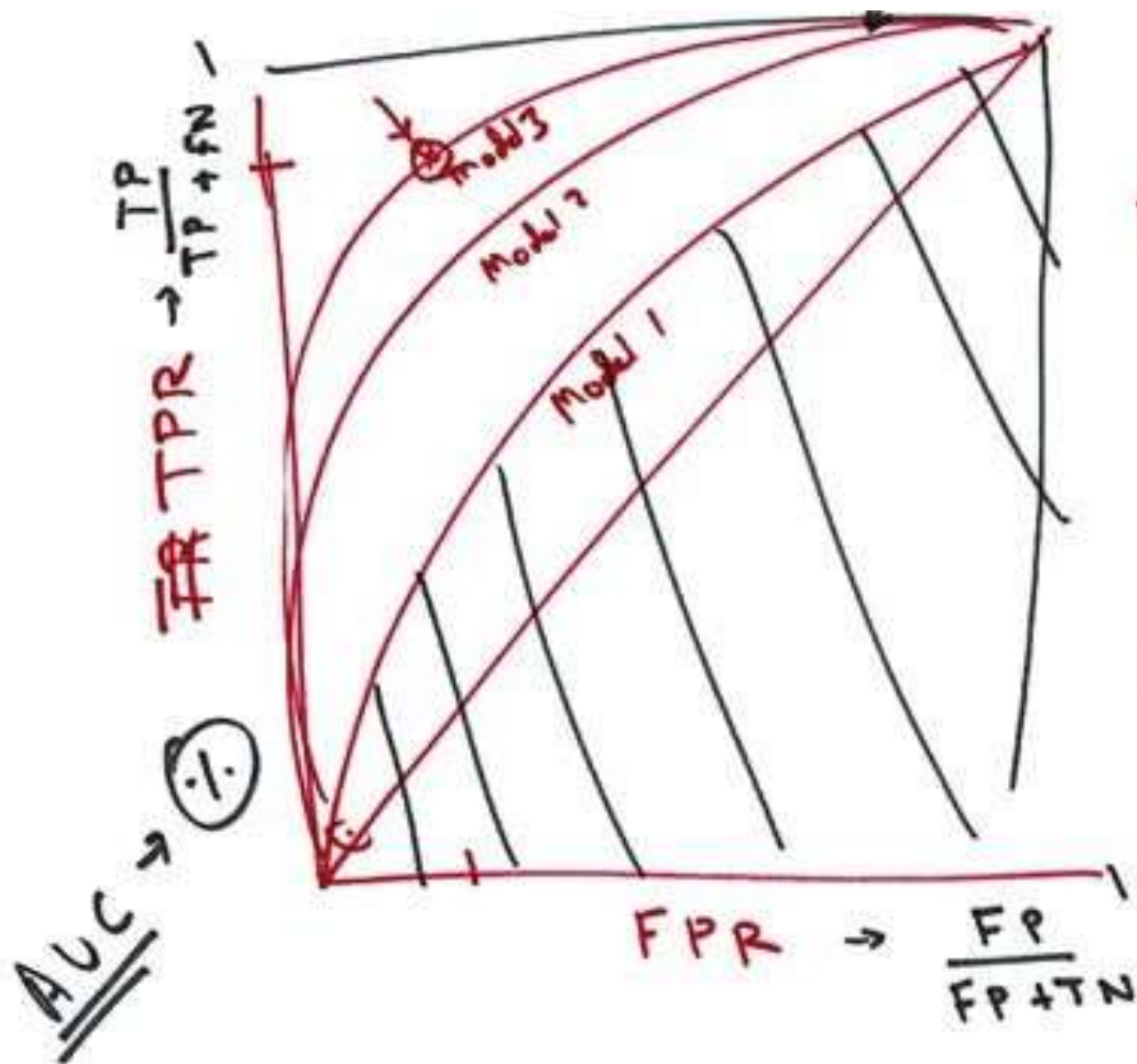
$$= 1 - 10^{-8}$$

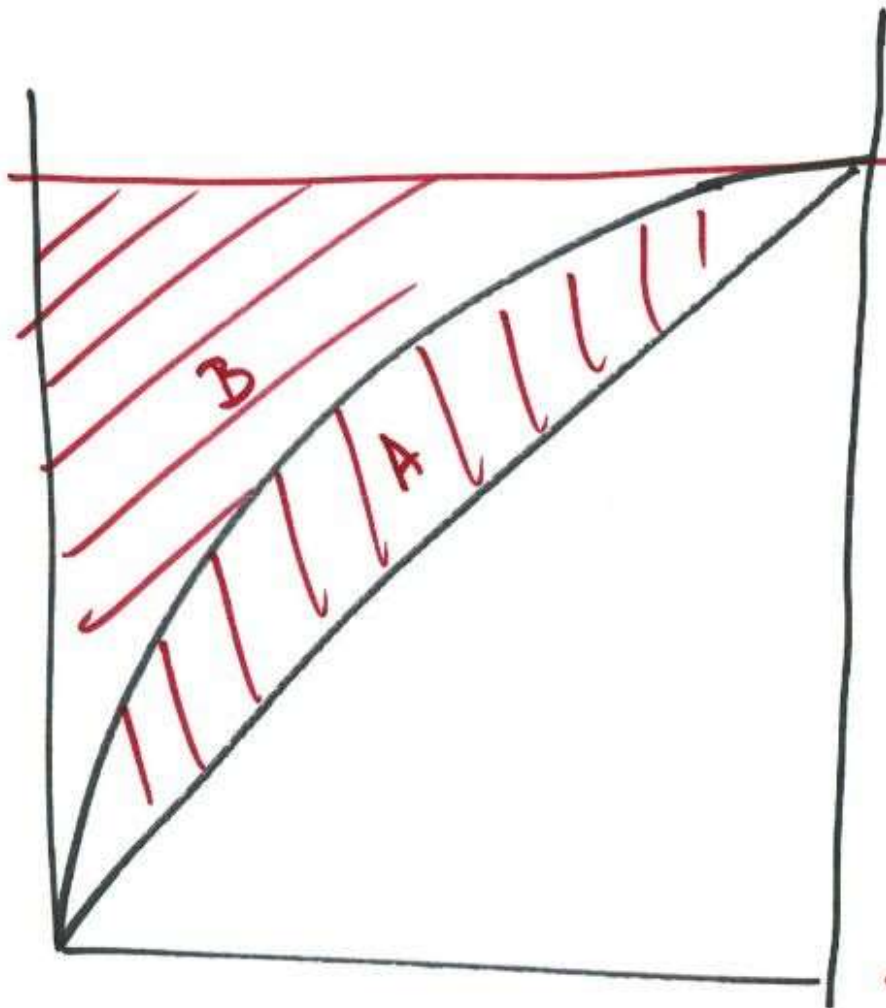
$$= 0.999$$

	P	N
P	10	$10^9 - 10$
N	0	0

	Acc	Recall	Precision
All as not terrorist	1	0	0
All as terrorist		1	low
Predicts the top terrible only		0	1

$$F_1 = (2 * P * R) / (P+R)$$





$$\text{Gini Coefficient} = A / A + B$$

$$A = \text{AUC} - 0.5$$

$$A + B = 0.5$$

$$\text{Gini Coefficient} = \text{AUC} - 0.5 / 0.5$$

$$\text{Gini Coefficient} = 2 * \text{AUC} - 1$$