Confusion Matrix

* Binary classification

Email are Spam or Not-spam

100 i/p email → ML → Spam / Not spam → Predicted

Actual data

ML ↳ Probability

0.8

0.9

* How to evaluate this model

* confusion Matrix

* Threshold

0.67

0.75

0.9

100 email

80 Not spam (−ve) | 20 spam (+ve) → Actual data

60 Not spam | 20 spam | 15 spam | 5 Not spam

Predicted o/p by ML

75

$\frac{75}{100}$

Piyush Wairale

A **confusion matrix** is a table used in machine learning to evaluate the performance of a classification model, particularly in supervised learning.
It provides a detailed breakdown of the model's correct and incorrect predictions, allowing you to understand not just how often the model is right or wrong, but the types of errors it makes.

error matrix

For a binary classification problem, the confusion matrix is a 2x2 table that compares the actual labels with the predicted labels:

Predicted

Actual

	Predicted Positive	Predicted Negative	
Actual Positive 20	True Positive (TP) 15	False Negative (FN) 5	→ 20
Actual Negative 80	False Positive (FP) 20	True Negative (TN) 60	→ 80
100			

TP = 15
FN = 5

$$\frac{20}{N_a}$$

Definitions:

- **True Positives (TP)**: Cases where the model correctly predicts the positive class.
- **True Negatives (TN)**: Cases where the model correctly predicts the negative class.
- **False Positives (FP)**: Cases where the model incorrectly predicts the positive class (Type I error).
- **False Negatives (FN)**: Cases where the model incorrectly predicts the negative class (Type II error).

Confusion Matrix

Imp

- **Precision:** The proportion of positive identifications that were actually correct.

$$\boxed{\text{Precision} = \frac{TP}{TP + FP}} = \frac{15}{15+20} = 15/35$$

- **Recall (Sensitivity):** The proportion of actual positives that were correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{15}{15+5} = 15/20$$

- **F1 Score:** The harmonic mean of precision and recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{\frac{15}{35} \times \frac{15}{20}}{\frac{15}{35} + \frac{15}{20}}$$

- **Specificity:** The proportion of actual negatives that were correctly identified.

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

- **Accuracy:** The proportion of total correct predictions.

60

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad 100 \quad = \frac{75}{100}$$

Receiver Operating characteristic curve

AUC → Area under ROC

The ROC curve is a graphical representation of a classifier's performance across all classification thresholds. It plots two metrics:

- **True Positive Rate (TPR)**: Also known as Sensitivity or Recall, it measures the proportion of actual positives that are correctly identified.

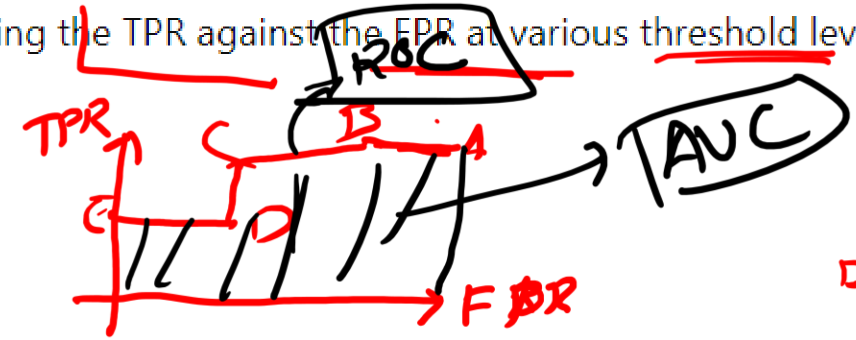
$$TPR = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{TP}{P_a} = \frac{15}{20}$$

- **False Positive Rate (FPR)**: It measures the proportion of actual negatives that are incorrectly identified as positives.

⇒ for each threshold value

$$FPR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} = \frac{20}{80} = \frac{1}{4} = \frac{FP}{N_a}$$

The ROC curve is created by plotting the TPR against the FPR at various threshold levels.



A $TH_1 = 0.6$, CM_1

B $TH_2 = 0.7$, CM_2

C $TH_3 = 0.75$, CM_3

D $TH_4 = 0.85$, CM_4

E $TH_5 = 0.9$

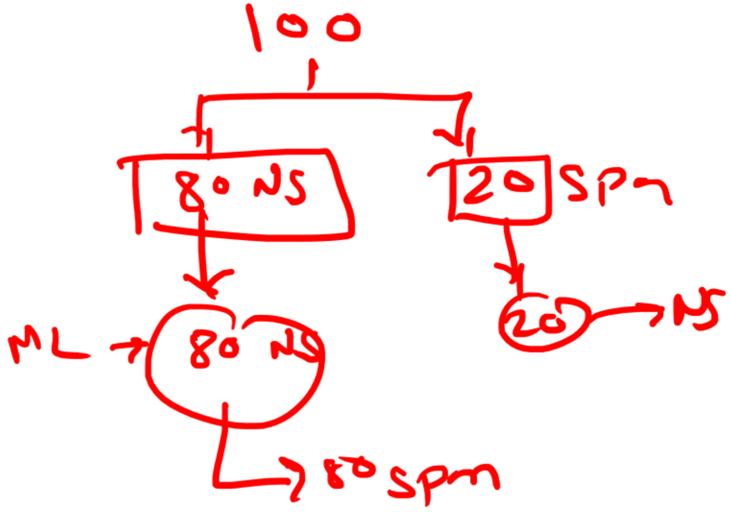
AUC (Area Under the ROC Curve)

AUC quantifies the overall performance of a model by calculating the area under the ROC curve. It provides a single value that summarizes the model's ability to discriminate between the positive and negative classes.

- **AUC = 1**: Perfect model (no false positives or false negatives).
- **AUC = 0.5**: Model performs no better than random chance.
- **AUC < 0.5**: Indicates a model that is performing worse than random guessing.

$\boxed{AUC = 0}$

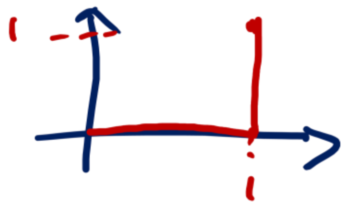
$\boxed{0.5 < AUC < 1}$ ↑

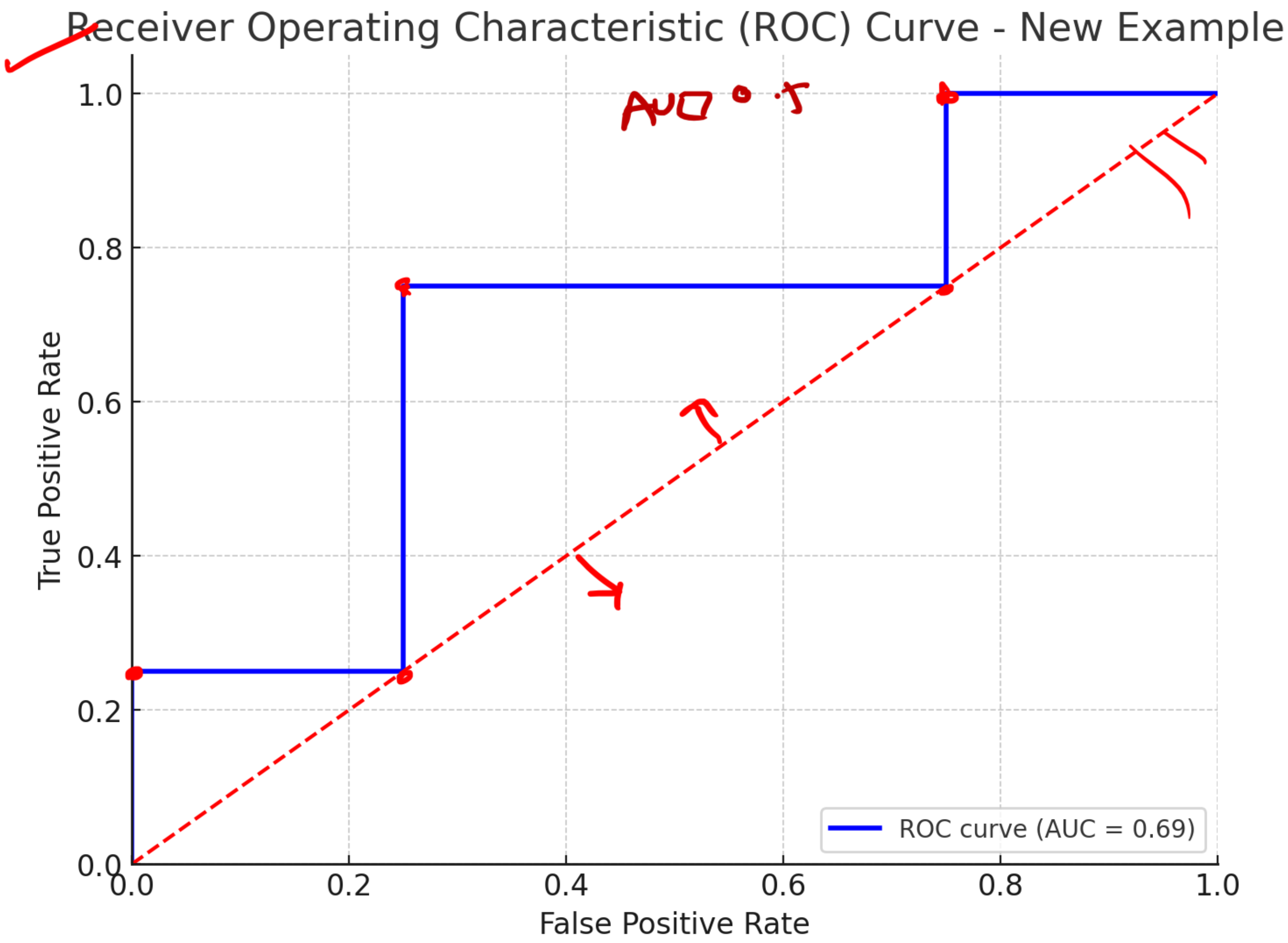


Piyush Wairale

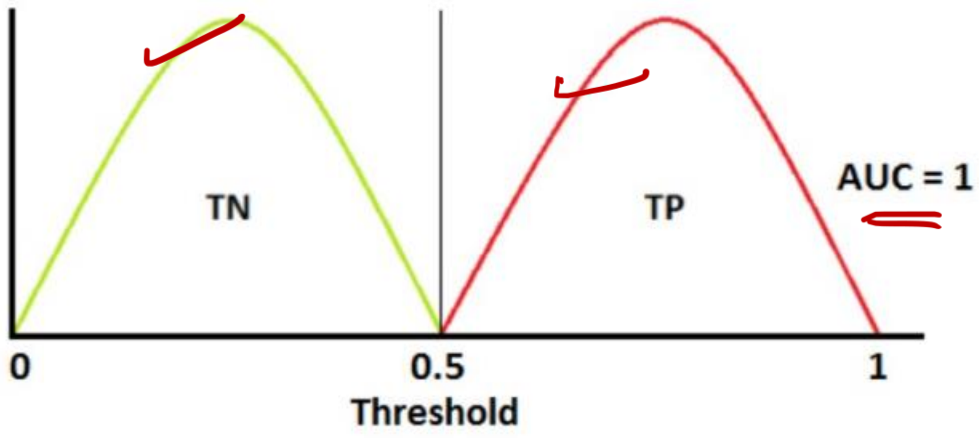
① AUC = 1

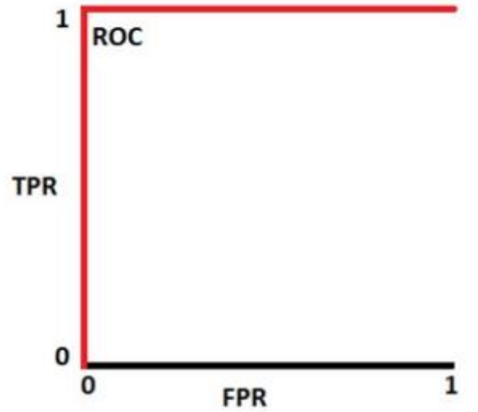
② AUC = 0

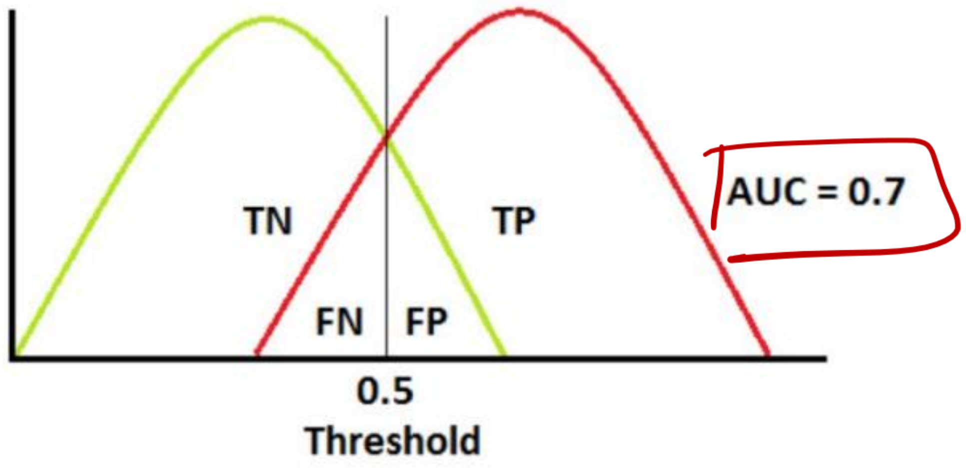


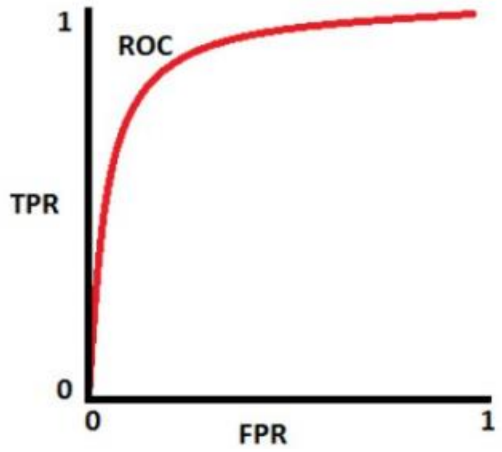


0.5

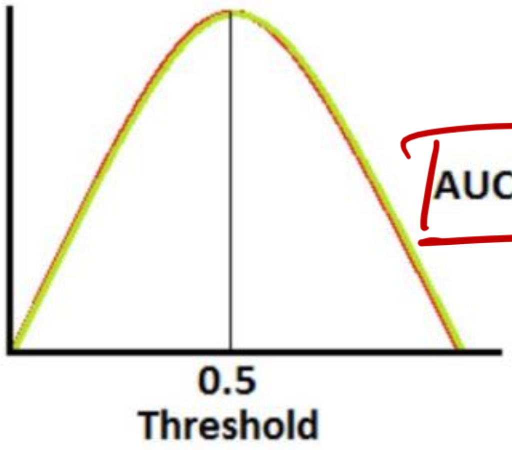

TN
TP
AUC = 1
0
0.5
1
Threshold


1
ROC
TPR
0
0
FPR
1

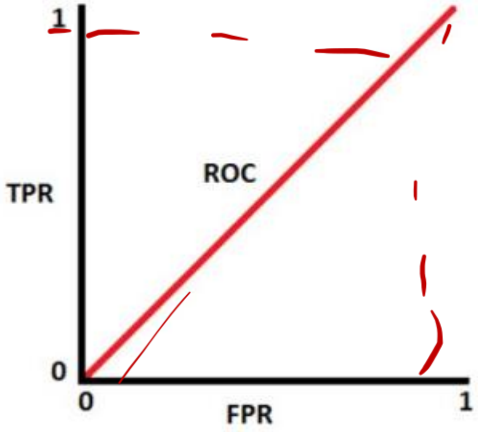

TN
TP
AUC = 0.7
FN
FP
0.5
Threshold

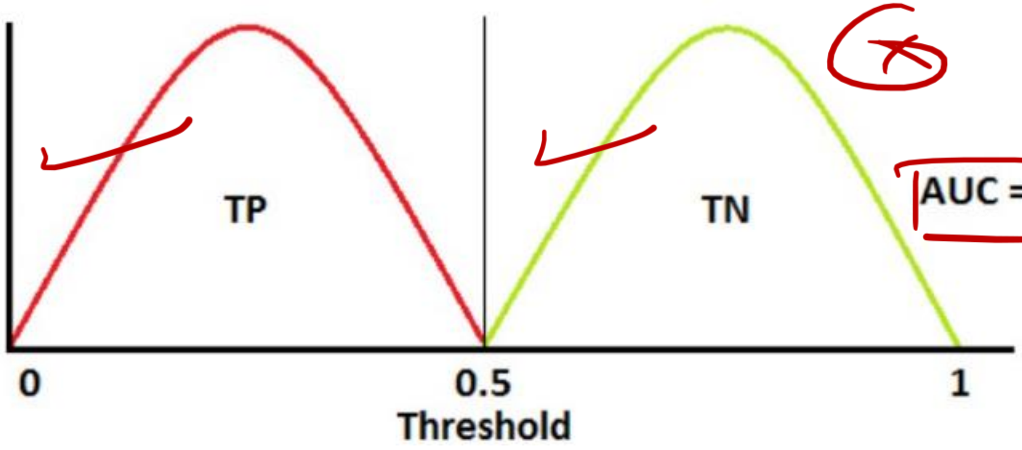

1
ROC
TPR
0
0
FPR
1

ROC - AUC

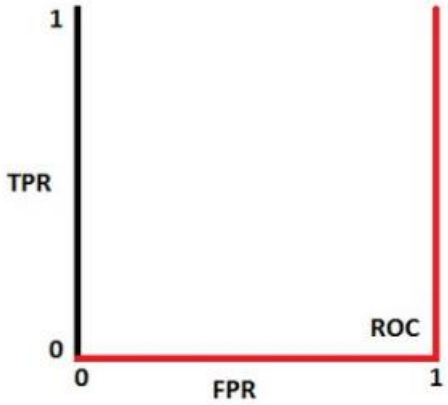


AUC = 0.5





AUC = 0



Piyush Wairale

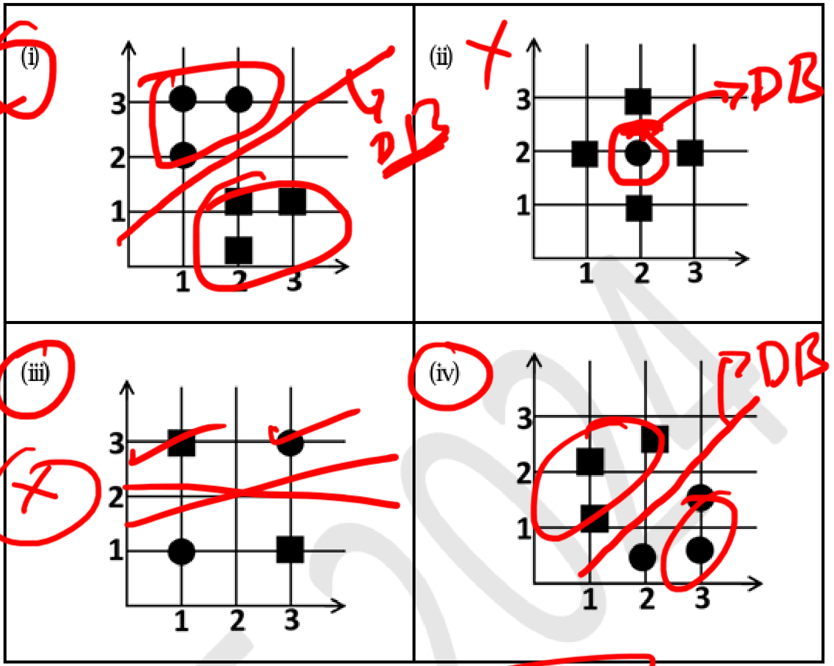
Non-Linearly Separable

Definition: A dataset is considered **non-linearly separable** if there is no single straight line or hyperplane that can separate the classes. Instead, the classes are intermixed in such a way that a more complex boundary is required to achieve perfect classification.

Characteristics:

- **Complex Decision Boundaries:** The classification boundary may need to be curved or take on more complex shapes to separate the classes.
- **Advanced Algorithms Required:** Algorithms such as Support Vector Machines (with non-linear kernels), Decision Trees, and Neural Networks are better suited for non-linearly separable data.

Q.53 Consider the following figures representing datasets consisting of two-dimensional features with two classes denoted by circles and squares. 2M



Which of the following is/are TRUE? MSQ

(A)	(i) is linearly separable.
(B)	(ii) is linearly separable.
(C)	(iii) is linearly separable.
(D)	(iv) is linearly separable.

1M

LS

Yes → LDA

NO → PCA