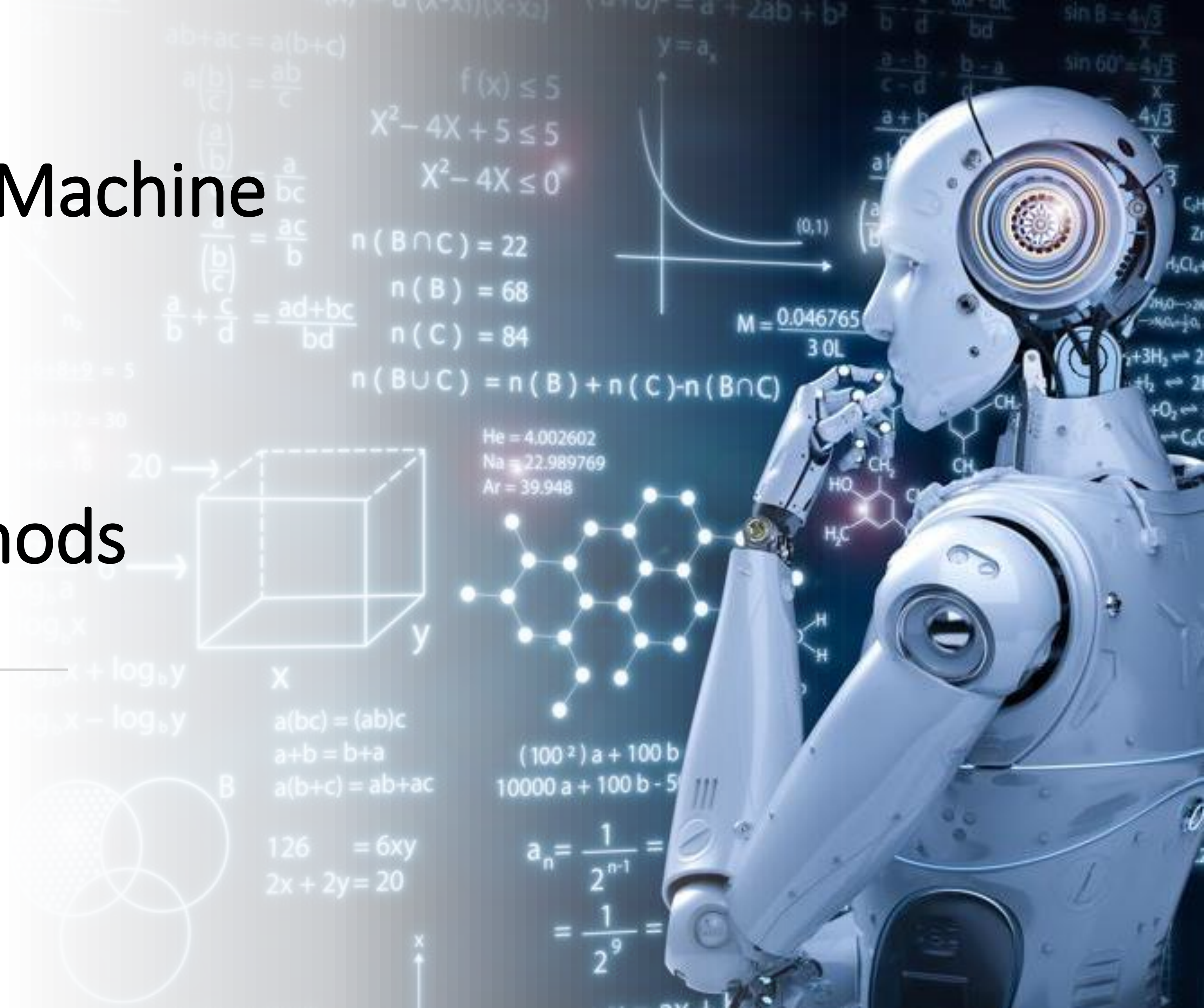
Introduction to Machine Learning

Lecture 4
Evaluation Methods



A. S. M. Sanwar Hosen

Email: sanwar@jbnu.ac.kr

Date: 19 Oct., 2023

Evaluation Methods

❑ Evaluation Metrics in ML

Evaluation metrics are used to measure the quality of the statistical or ML model. The metrics commonly used in ML are as follows:

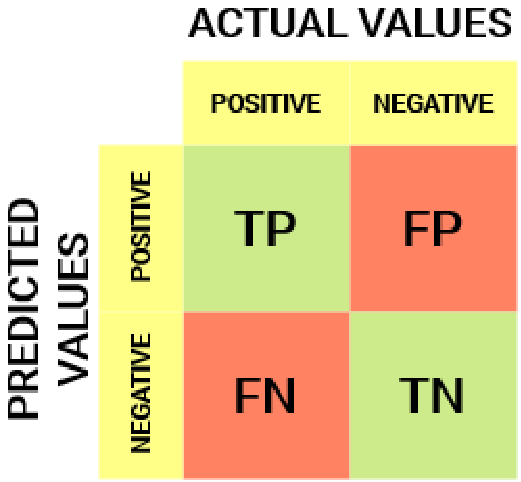
- ✓ Confusion Matrix
- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ F1-Score
- ✓ Receiver Operating Characteristics (ROC) Curve
- ✓ Area Under the ROC Curve (AUC-ROC)
- ✓ Training Loss
- ✓ Validation Loss

Evaluation Methods

❑ Confusion Matrix

A confusion matrix is an $(N \times N)$ matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compare the actual target values with the predicted values by the ML model. This gives us a holistic view of how well out classification model performing and what kinds of errors it is making.

For a binary classification problem, we would have (2×2) matrix as shown below with 4 values:



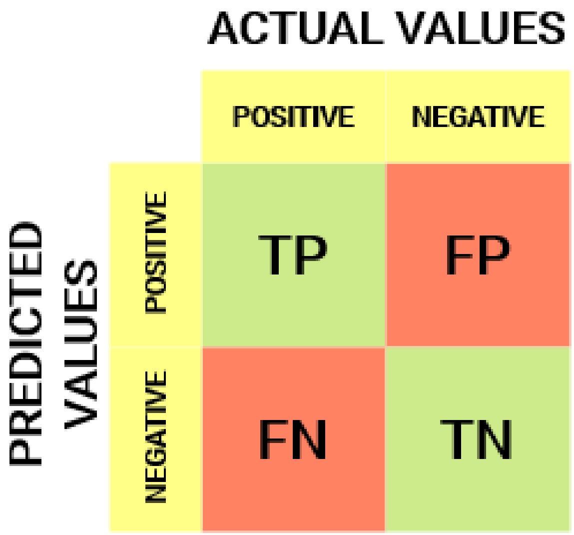
- ✓ The target variable has two values: Positive (1) or Negative (0)
- ✓ The columns represent the actual values of the target variable
- ✓ The rows represent the predicted values of the target variable

Evaluation Methods

❑ Confusion Matrix (cont..)

There are 4 important terms in a confusion matrix:

- ✓ True Positives (TP)
- ✓ True Negative (TN)
- ✓ False Positive (FP)
- ✓ False Negative (FN)



❑ Confusion Matrix (cont..)

✓ True Positive (TP): The case in which the ML model predicted Positive (Yes or 1) and the actual output was also Positive (Yes or 1).

Example: Predicting apple class on the following dataset.

Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP	FP
	NEGATIVE	FN	TN

What is the TP of the ML model?

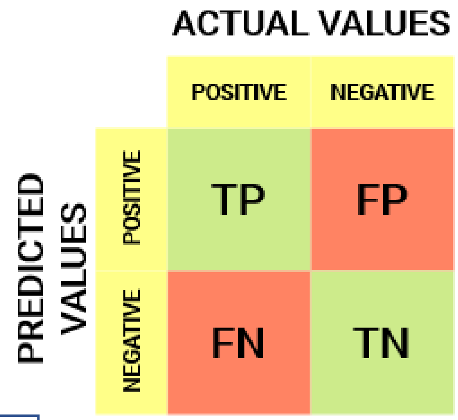
$TP = 2$

Evaluation Methods

❑ Confusion Matrix (cont..)

✓ True Negative (TN): The case in which the ML model predicted Negative (No or 0) and the actual output was also Negative (No or 0).

Example: Predicting apple class on the following dataset.



Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

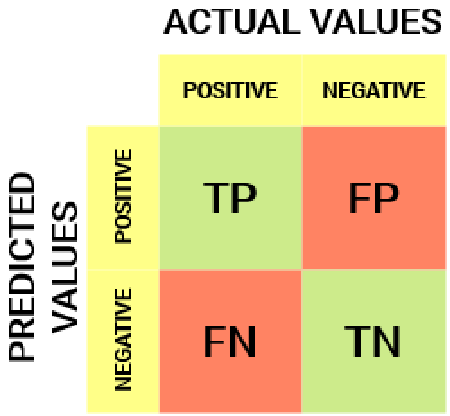
What is the TN of the ML model?

$TN = 1$

Confusion Matrix (cont..)

✓ False Positive (FP): The case in which the ML model predicted Positive (Yes or 1) and the actual output was Negative (No or 0).

Example: Predicting apple class on the following dataset.



Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

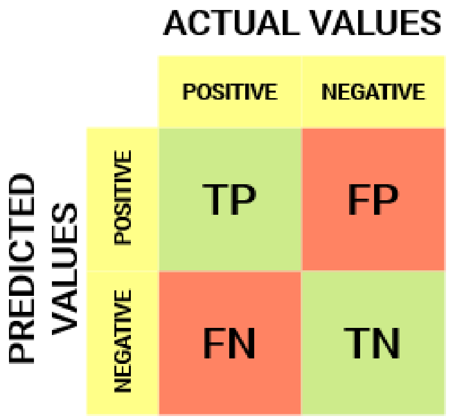
What is the FP of the ML model?

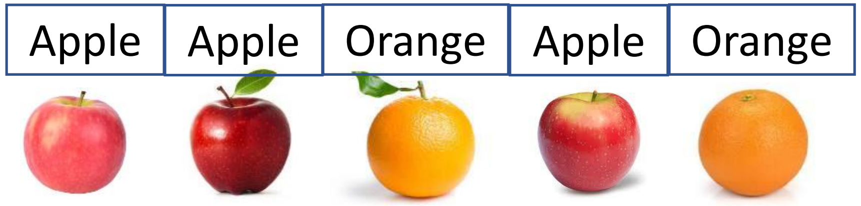
$FP = 1$

Confusion Matrix (cont..)

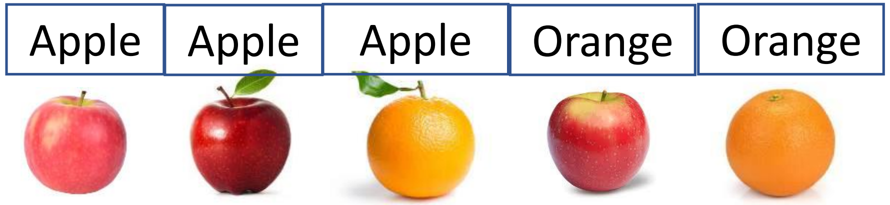
✓ False Negative (FN): The case in which the ML model predicted Negative (No or 0) and the actual output was Positive (Yes or 1).

Example: Predicting apple class on the following dataset.





Test data and actual output



Prediction

What is the FN of the ML model?

$FN = 1$

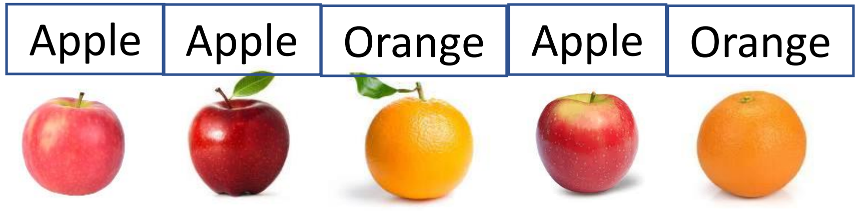
Evaluation Methods

❑ Evaluation Metrics in ML

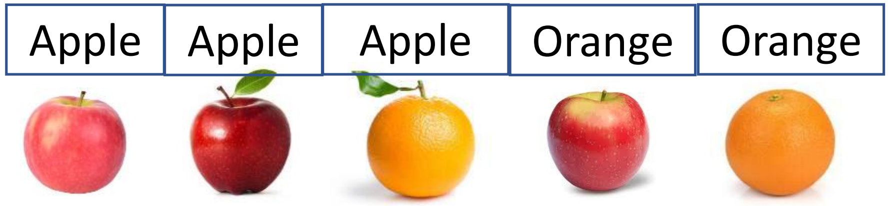
✓ Accuracy: It is the number of correct predictions to the total number of input samples.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ Sample} = \frac{True\ Positive\ (TP) + True\ Negetive (TN)}{Total\ Sample}$$

Example: Predicting apple class on the following dataset.



Test data and actual output



Prediction

What is the Accuracy of the ML model?

$$Accuracy = \frac{2+1}{5} = 0.6 = 0.6 \times 100 = 60\%$$

Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

✓ Precision: It measures the quality of a positive prediction made by the ML model.

$$Precision = \frac{True\ Positive\ (TP)}{Total\ Positive\ in\ prediction} = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive(FP)}$$

Example: Predicting apple class on the following dataset.

Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

What is the Precision of the ML model?

$$Precision = \frac{2}{2+1} = 0.6667 = 0.6667 \times 100 = 66.67\%$$

Evaluation Metrics in ML (cont..)

✓ Recall (True Positive Rate or Sensitivity): It measures the actual Positive values correctly identified by the ML model.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative (FN)}$$

Example: Predicting apple class on the following dataset.

Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

What is the Recall of the ML model?

$$Recall = \frac{2}{2+1} = 0.6667 = 0.6667 \times 100 = 66.67\%$$

Evaluation Methods

Evaluation Metrics in ML (cont..)

✓ False Positive Rate (FPR): It measures the proportion of actual negative cases that are incorrectly classified as positive by the model. It is calculated as:

$$FPR = \frac{False\ Positive\ (FP)}{False\ Positive\ (FP) + True\ Negative\ (TN)}$$

Example: Predicting apple class on the following dataset.

Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

What is the FPR of the ML model?

$$FPR = \frac{1}{1+1} = 0.50 = 0.50 \times 100 = 50\%$$

Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

✓ True Negative Rate (TNR) or Specificity: It measures the proportion of actual negative cases that are correctly identified as negative by the model. It is calculated as:

$$TNR = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN) + False\ Positive\ (FP)}$$

Example: Predicting apple class on the following dataset.

Apple	Apple	Orange	Apple	Orange

Test data and actual output

Apple	Apple	Apple	Orange	Orange

Prediction

What is the TNR of the ML model?

$$TNR = \frac{1}{1+1} = 0.50 = 0.50 \times 100 = 50\%$$

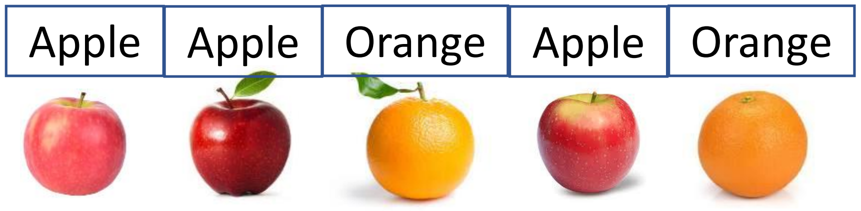
Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

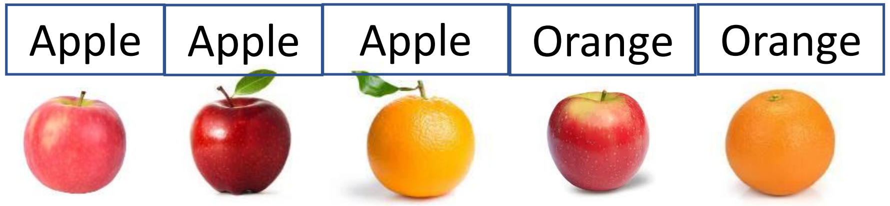
✓ False Negative Rate (FNR): It measures the proportion of actual positive cases that are incorrectly classified as negative by the model. It is calculated as:

$$FNR = \frac{False\ Negative\ (FN)}{True\ Positive\ (TP) + False\ Negative\ (FN)}$$

Example: Predicting apple class on the following dataset.



Test data and actual output



Prediction

What is the FNR of the ML model?

$$FNR = \frac{1}{2+1} = 0.33 = 0.33 \times 100 = 33.33\%$$

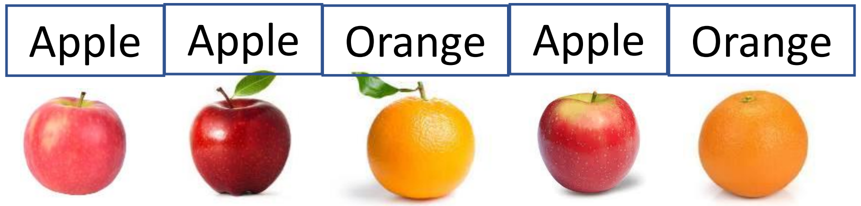
Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

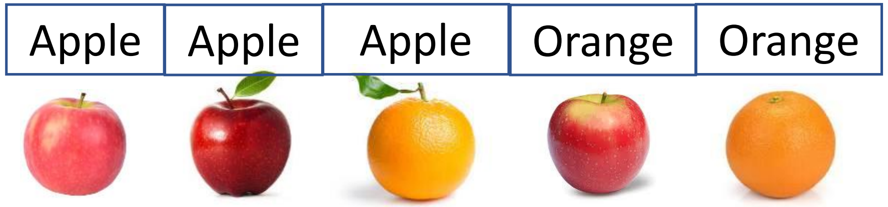
✓ F1-Score: It is the Harmonic Mean between Precision and Recall and it tells how precise the ML model is.

$$F1 - Score \ = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2\times(Precision\times Recall)}{Precision+Recall}$$

Example: Predicting apple class on the following dataset.



Test data and actual output



Prediction

What is the F1-Score of the ML model?

$$F1 - Score = \frac{2\times(0.6667\times0.6667)}{0.6667+0.6667} = \frac{0.8890}{1.3334} = 0.6667 = 0.6667 \times 100 = 66.67\%$$

Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

✓ Receiver Operating Characteristics (ROC) Curve: The ROC curve is a graphical representation used in ML and statistics to evaluate the performance of binary classification models. It illustrate the trade-off between the true positive rate TPR (Sensitivity) and false positive rate FPR (1-Specificity) defined bellow.

1) $TPR = \frac{TP}{TP + FN}$ and 2) $FPR = \frac{FP}{FP + TN}$

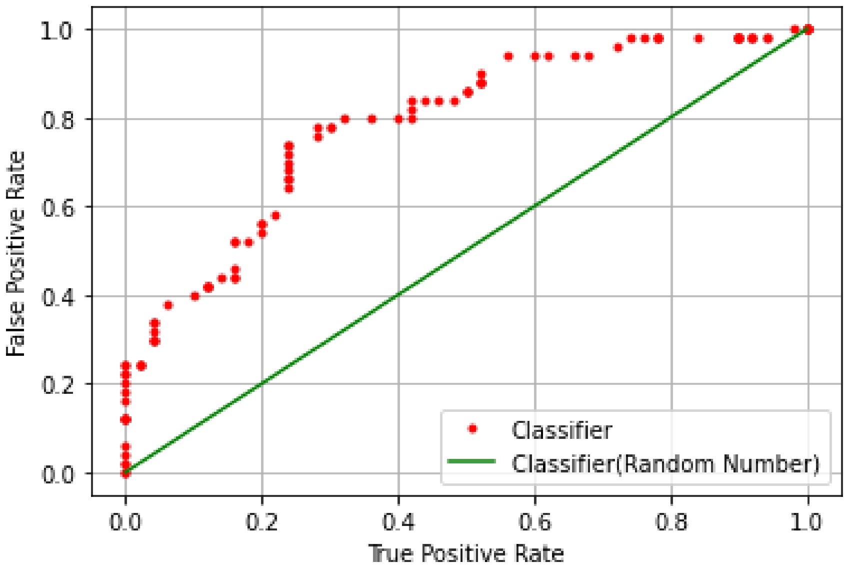


Figure: ROC curve.

Table: Predictions probability

ID	Actual Output	Prediction Probability	Threshold > 0.6	Threshold > 0.7	Threshold > 0.8	Performance Metric
1	0	0.97	1	1	1	
2	1	0.67	1	0	0	
3	1	0.56	0	0	0	
4	0	0.75	1	1	0	
5	1	0.81	1	1	1	
6	0	0.85	1	1	1	
7	0	0.71	1	1	0	
8	0	0.85	1	1	1	
9	1	0.80	1	1	1	
10	0	0.81	1	1	1	
			0.75	0.5	0.5	**TPR**
			1	1	0.66	**FPR**
			0	0	0.33	**TNR**
			0.25	0.5	0.5	**FNR**

Evaluation Methods

❑ Evaluation Metrics in ML (cont..)

✓ Area Under ROC Curve (AUC-ROC): It is a metric used to evaluate the performance of binary classification models. It is commonly used in ML and statistics to assess the ability of a model to distinguish between two classes (usually a positive class and a negative class).

The AUC value ranges from 0 to 1, where:

- AUC = 0.5: This indicates that the model's performance is no better than random guessing.
- AUC < 0.5: The model's performance is worse than random guessing, indicating that it is making incorrect predictions.
- AUC > 0.5: The model's performance is better than random guessing, and the higher the AUC value, the better the model's ability to discriminate between the two classes.

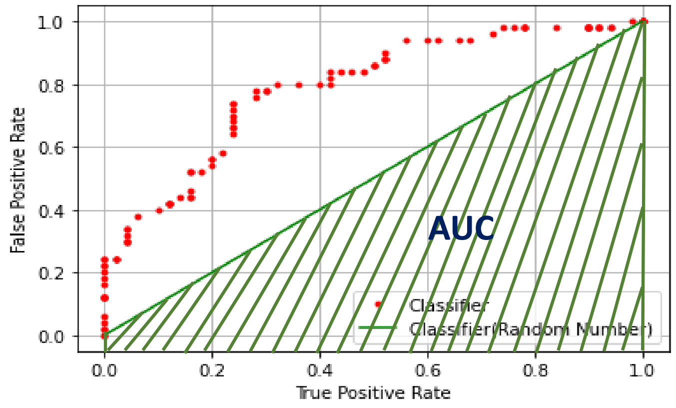


Figure: AUC-ROC curve (random guessing).

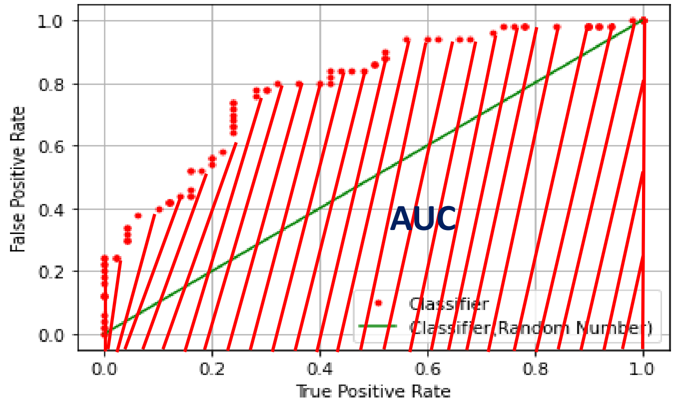


Figure: AUC-ROC curve (classifier).

Evaluation Methods

❑ **Lecture Overview**

✓ **Evaluation Metrics**

- Confusion Matrix

(True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (TN))

- Accuracy
- Precision
- Recall
- F1-Score
- Receiver Operating Characteristics (ROC)
- Area Under the ROC Curve (AUC-ROC)