END-TO-END AI

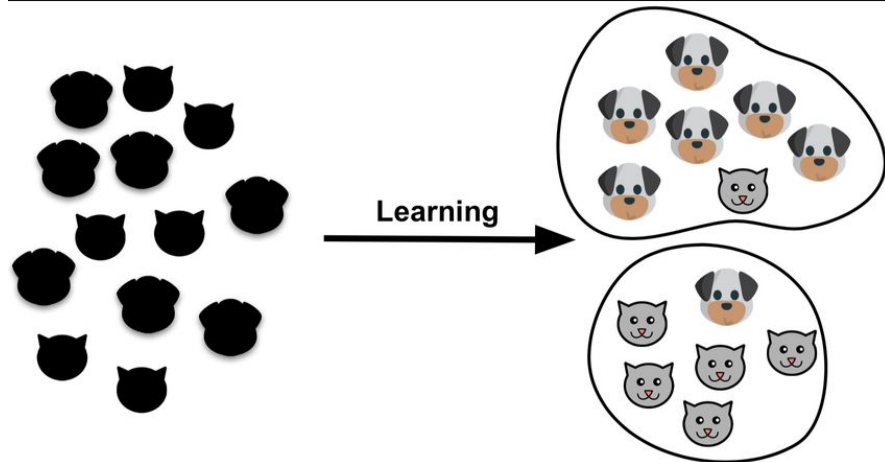## MODEL TRAINING AND EXPERIMENT TRACKING

Q1 COHORT(WEEK-3), 2022

What is Classification

Classification is the **division** of dataset into their respective **distinct classes** or categories.

Classification is applicable to both **structured**(e.g. Tabular data) and **unstructured** data(texts, images data).

Learning

Type of Classification

There are different type of classifications, these are:

➔ Binary Classification: Classification problem with two class labels.

➔ Multi-class classification: This involves those classification problem that has two more classes.

➔ Multi Label classification: This involves predicting one or more classes.

➔ Imbalanced classification: This refers to a classification problem where the one class is more represented over the other.

Binary Classification

This involves classification problem that possess only two classes or categories. It is used in solving problems such as:

- Email spam detection

- Customer churn

- Customer subscription

- Conversion prediction
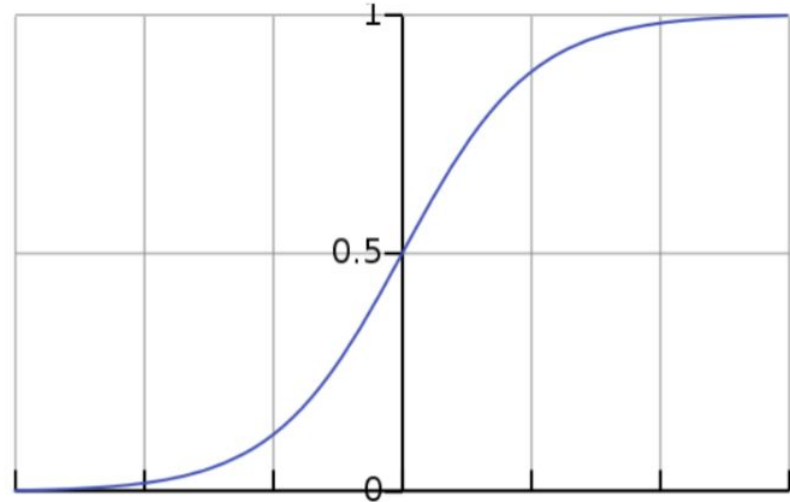
https://en.wikipedia.org/wiki/Sigmoid_function#/media/File:Logistic-curve.svg

Testing Methods for Binary Classification

➔ **Confusion Matrix**

- True positives (TP): This is the number of instances that are pos and are classified as positive.
- False positives (FP): This is the number of instances that are negative and are classified as positive.
- False negatives (FN): This is the number of instances that are positive and are classified as negative.
- True negatives (TN): This is the number of instances that are negative and are classified as negative.

➔ Binary Classification Test
➔ ROC(Receiver Operating Characteristic) Curve

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Testing Methods for Binary Classification

➔ Confusion Matrix:

➔ **Binary Classification Test:** This is includes the following: classification accuracy, error rate, sensitivity, specificity.

➔ ROC(Receiver Operating Characteristic) Curve

➔ Classification Accuracy = (True Positive + True Negative)/ Total Instance

➔ Error rate = (False Positive + False Negative)/ Total Instance

➔ Sensitivity = True Positive / Positive Instances

➔ Specificity = True Negative / Negative Instances

Testing Methods for Binary Classification

➜ Confusion Matrix:

➜ Binary Classification Test:

➜ **ROC(Receiver Operating Characteristic) Curve**

Reference: https://en.wikipedia.org/wiki/Receiver_operating_characteristic
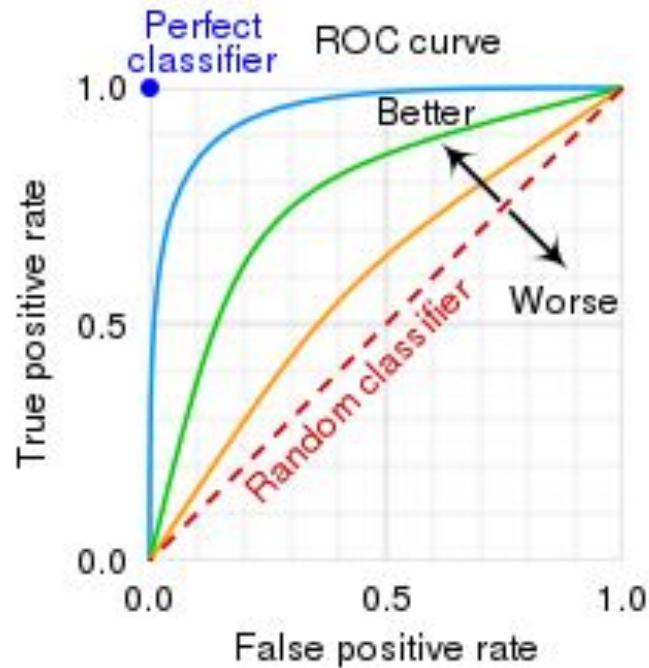
Algorithms for Binary Classification

Below are the list of algorithms for binary classification:

➔ Logistics regression

➔ Support Vector Machine

➔ Bayesian Classification

➔ Decision Trees

HANDS ON SESSION