ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 11 **7 July 2025**

Confusion Matrix in Machine Learning

Confusion matrix is a simple table used to measure how well a classification model is performing. It compares the predictions made by the model with the actual results and shows where the model was right or wrong. This helps you understand where the model is making mistakes so you can improve it. It breaks down the predictions into four categories:

- **True Positive (TP):** The model correctly predicted a positive outcome i.e the actual outcome was positive.
- **True Negative (TN):** The model correctly predicted a negative outcome i.e the actual outcome was negative.
- **False Positive (FP):** The model incorrectly predicted a positive outcome i.e the actual outcome was negative. It is also known as a Type I error.
- **False Negative (FN):** The model incorrectly predicted a negative outcome i.e the actual outcome was positive. It is also known as a Type II error.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Metrics based on Confusion Matrix Data

- **Accuracy** measures the overall correctness of the model by showing how many predictions were correct out of all predictions. It can be misleading if one class dominates.
  Formula: Accuracy = (TP + TN) / (TP + TN + FP + FN)
- **Precision** indicates how many of the positive predictions were actually correct. It is important when false positives should be minimized, like in spam detection.
  Formula: Precision = TP / (TP + FP)

- **Recall** (or sensitivity) measures how well the model captures actual positive cases. It is crucial in situations like medical testing, where missing a positive case is dangerous.
  Formula: Recall = TP / (TP + FN)
- **F1-Score** is the harmonic mean of precision and recall, providing a balanced metric when both false positives and false negatives matter.
  Formula: F1-Score = 2 × (Precision × Recall) / (Precision + Recall)
- **Specificity** measures the ability of the model to correctly identify negative cases. It is especially useful in medical screening to avoid false alarms.
  Formula: Specificity = TN / (TN + FP)
- **Type I Error (False Positive)** occurs when a negative instance is incorrectly predicted as positive. This affects the model's precision.
  Formula: Type I Error Rate = FP / (FP + TN)
- **Type II Error (False Negative)** happens when a positive instance is incorrectly predicted as negative. This impacts recall.
  Formula: Type II Error Rate = FN / (TP + FN)

Large Language Models(LLMs)

A **large language model** is a type of artificial intelligence algorithm that applies neural network techniques with lots of parameters to process and understand human languages or text using self-supervised learning techniques. Tasks like text generation, machine translation, summary writing, image generation from texts, machine coding, chat-bots, or Conversational AI are applications of the Large Language Model.

How do Large Language Models work?

Large Language Models (LLMs) operate on the principles of deep learning, leveraging neural network architectures to process and understand human languages.

These models, are trained on vast datasets using self-supervised learning techniques. The core of their functionality lies in the intricate patterns and relationships they learn from diverse language data during training. LLMs consist of multiple layers, including feedforward layers, embedding layers, and attention layers. They employ attention mechanisms, like self-attention, to weigh the importance of different tokens in a sequence, allowing the model to capture dependencies and relationships.

Artificial Intelligence V/s Machine Learning

Aspect	Artificial Intelligence (AI)	Machine Learning (ML)
Definition	A field of computer science focused on building smart machines	A subset of AI that enables systems to learn from data
Goal	Mimic human intelligence and behavior	Learn from data to make predictions or decisions
Approach	Can be rule-based or learning-based	Always data-driven and learning-based
Scope	Broad (includes ML, NLP, robotics, expert systems, etc.)	Narrow (focused on learning from data)
Examples	Chatbots, self-driving cars, game AI	Spam filters, recommendation systems, fraud detection
Dependency on Data	May or may not require large amounts of data	Requires large datasets to train models

By : Aditi Tangri URN : 2302460 CRN : 2315004