ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 11

Date: Jul 07, 2025

Understanding the 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)

It also helps calculate key measures like accuracy, precision and recall which give a better idea of performance especially when the data is imbalanced.

Metrics based on Confusion Matrix Data

1. Accuracy

Accuracy shows how many predictions the model got right out of all the predictions. It gives idea of overall performance but it can be misleading when one class is more dominant over the other. For example a model that predicts the majority class correctly most of the time might have high accuracy but still fail to capture important details about other classes. It can be calculated using the below formula:

Accuracy=TP+TN/(TP+TN+FP+FN)

2. Precision

Precision focus on the quality of the model's positive predictions. It tells us how many of the "positive" predictions were actually correct. It is important in situations where false positives need to be minimized such as detecting spam emails or fraud. The formula of precision is:

$$\text{Precision}=TP/(TP+FP)$$

3. Recall

Recall measures how how good the model is at predicting positives. It shows the proportion of true positives detected out of all the actual positive instances. High recall is essential when missing positive cases has significant consequences like in medical tests.

$$\text{Recall}=TP/(TP+FN)$$

4. F1-Score

F1-score combines precision and recall into a single metric to balance their trade-off. It provides a better sense of a model's overall performance particularly for imbalanced datasets. It is helpful when both false positives and false negatives are important though it assumes precision and recall are equally important but in some situations one might matter more than the other.

$$\text{F1-Score}=2\cdot\text{Precision}\cdot\text{Recall}/(\text{Precision}+\text{Recall})$$

5. Specificity

Specificity is another important metric in the evaluation of classification models particularly in binary classification. It measures the ability of a model to correctly identify negative instances. Specificity is also known as the True Negative Rate Formula is given by:

$$\text{Specificity}=TN/(TN+FP)$$

6. Type 1 and Type 2 error

Type 1 and Type 2 error are:

Type 1 error: It occurs when the model incorrectly predicts a positive instance but the actual instance is negative. This is also known as a false positive. Type 1 Errors affect the precision of a model which measures the accuracy of positive predictions.
$$\text{Type 1 Error}=FP/(FP+TN)$$

Type 2 error: This occurs when the model fails to predict a positive instance even though it is actually positive. This is also known as a false negative. Type 2 Errors impact the recall of a model which measures how well the model identifies all actual positive cases.
$$\text{Type 2 Error}=FN/(TP+FN)$$

What are 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.

Examples of such LLM models are Chat GPT by open AI, BERT (Bidirectional Encoder Representations from Transformers) by Google, etc.

There are many techniques that were tried to perform natural language-related tasks but the LLM is purely based on the deep learning methodologies. LLM (Large language model) models are highly efficient in capturing the complex entity relationships in the text at hand and can generate the text using the semantic and syntactic of that particular language in which we wish to do so.

Architecture of LLM

Large Language Model's (LLM) architecture is determined by a number of factors, like the objective of the specific model design, the available computational resources, and the kind of language processing tasks that are to be carried out by the LLM. The general architecture of LLM consists of many layers such as the feed forward layers, embedding layers, attention layers. A text which is embedded inside is collaborated together to generate predictions.

Important components to influence Large Language Model architecture:

- Model Size and Parameter Count
- input representations
- Self-Attention Mechanisms
- Training Objectives
- Computational Efficiency
- Decoding and Output Generation

Popular Large Language Models

- **GPT-3:** GPT 3 is developed by OpenAI, stands for Generative Pre-trained Transformer 3. This model powers ChatGPT and is widely recognized for its ability to generate human-like text across a variety of applications.
- **BERT:** It is created by Google, is commonly used for natural language processing tasks and generating text embeddings, which can also be utilized for training other models.
- **RoBERTa:** RoBERTa is an advanced version of BERT, stands for Robustly Optimized BERT Pretraining Approach. Developed by Facebook AI Research, it enhances the performance of the transformer architecture.

Applications of Large Language Models

LLMs, such as GPT-3, have a wide range of applications across various domains. Few of them are:

- **Natural Language Understanding (NLU):**Large language models power advanced chatbots capable of engaging in natural conversations.They can be used to create intelligent virtual assistants for tasks like scheduling, reminders, and information retrieval.
- **Content Generation:**Creating human-like text for various purposes, including content creation, creative writing, and storytelling.Writing code snippets based on natural language descriptions or commands.
- **Language Translation:** Large language models can aid in translating text between different languages with improved accuracy and fluency.
- **Text Summarization:** Generating concise summaries of longer texts or articles.
- **Sentiment Analysis:** Analyzing and understanding sentiments expressed in social media posts, reviews, and comments.

Advantages of Large Language Models

- Large Language Models (LLMs) come with several advantages that contribute to their widespread adoption and success in various applications:
- LLMs can perform zero-shot learning, meaning they can generalize to tasks for which they were not explicitly trained. This capability allows for adaptability to new applications and scenarios without additional training.
- LLMs efficiently handle vast amounts of data, making them suitable for tasks that require a deep understanding of extensive text corpora, such as language translation and document summarization.
- LLMs enable the automation of various language-related tasks, from code generation to content creation, freeing up human resources for more strategic and complex aspects of a project.

Artificial Intelligence	Machine Learning

AI is a broader field focused on creating systems that mimic human intelligence, including reasoning, decision-making, and problem-solving.	ML is a subset of AI that focuses on teaching machines to learn patterns from data and improve over time without explicit programming
The main goal of AI is to develop machines that can perform complex tasks intelligently, similar to how humans think and act.	ML focuses on finding patterns in data and using them to make predictions or decisions. It aims to help systems improve automatically with experience.
AI systems aim to simulate human intelligence and can perform tasks across multiple domains.	ML focuses on training systems for specific tasks, such as prediction or classification.
AI aims to create systems that can think, learn, and make decisions autonomously.	ML aims to create systems that learn from data and improve their performance for a particular task.
AI can operate with minimal human intervention, depending on its complexity and design.	ML requires human involvement for data preparation, model training, and optimization
AI involves broader goals, including natural language processing, vision, and reasonin	ML focuses specifically on building models that identify patterns and relationships in data
Examples: Robotics, virtual assistants like Siri, autonomous vehicles, and intelligent chatbots.	Examples: Recommender systems, fraud detection, stock price forecasting, and social media friend suggestions.