Chapter 29: Model Evaluation Terminology

Introduction

In Artificial Intelligence and Machine Learning, simply building a model is not enough. Once a model is trained, we need to **evaluate** how well it performs. This process is called **Model Evaluation**. It helps us understand how accurate and reliable the model's predictions are. For this purpose, certain **terms and metrics** are commonly used.

Understanding model evaluation terminology is crucial because it helps us:

- Judge the effectiveness of a model.
- Compare different models.
- Improve the performance of AI systems.

In this chapter, you will learn about the key terminologies used in evaluating AI models in a simple and understandable way.

29.1 What is Model Evaluation?

Model evaluation refers to **measuring the performance** of an AI model on given data. The goal is to check whether the model is predicting correctly or not.

For example, if an AI model predicts whether an email is spam or not, model evaluation checks how many times it got it **right or wrong**.

29.2 Important Model Evaluation Terminologies

Below are the most commonly used terms in model evaluation:

---

1. True Positive (TP)

- The model **predicted YES**, and the actual answer was **YES**.
- Example: The AI says a person has a disease, and they actually do.

---

2. True Negative (TN)

- The model **predicted NO**, and the actual answer was **NO**.
- Example: The AI says a person does not have a disease, and they truly don’t.

3. False Positive (FP) *(Type I Error)*

- The model **predicted YES**, but the actual answer was **NO**.
- Example: The AI says a person has a disease, but they don't.

---

4. False Negative (FN) *(Type II Error)*

- The model **predicted NO**, but the actual answer was **YES**.
- Example: The AI says a person does not have a disease, but they do.

29.3 Confusion Matrix

A **confusion matrix** is a table used to describe the performance of a classification model. It shows the numbers of:

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

Structure of a Confusion Matrix:

	Predicted: Yes	**Predicted: No**
Actual: Yes	True Positive (TP)	False Negative (FN)
Actual: No	False Positive (FP)	True Negative (TN)

29.4 Accuracy

Accuracy tells how often the model is correct.

Formula:

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

Example:

If out of 100 predictions, the model got 90 right (TP + TN), then accuracy = 90%.

29.5 Precision

Precision tells how many of the predicted "yes" cases were actually "yes".

Formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Use Case: Important when **false positives** are harmful, like spam detection.

29.6 Recall (Sensitivity or True Positive Rate)

Recall tells how many of the actual "yes" cases were correctly predicted.

Formula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Use Case: Important when **false negatives** are dangerous, like in disease detection.

29.7 F1 Score

The **F1 Score** is a balance between Precision and Recall.

Formula:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Use Case: When you need a balance between precision and recall.

29.8 Overfitting and Underfitting

Overfitting:

- The model performs very well on training data but poorly on new data.
- It has **memorized** the data instead of learning patterns.

Underfitting:

- The model performs poorly on both training and testing data.
- It has **not learned** enough from the data.

29.9 Cross-Validation

Cross-validation is a technique to test how well your model performs on **unseen data** by splitting the dataset into multiple parts.

For example:

- Split the data into 5 parts.
- Train on 4 parts, test on 1.
- Repeat 5 times with different test sets.

29.10 Bias and Variance

Bias:

- Error due to wrong assumptions in the model.
- High bias = underfitting.

Variance:

- Error due to too much sensitivity to small variations in the training set.
- High variance = overfitting.

Summary

Model evaluation helps us determine whether our AI model is performing well or not. Key terminologies like **True Positive, False Negative, Precision, Recall, Accuracy**, and others give us insight into the model's strengths and weaknesses.

Here's a quick recap:

Term	Description
TP	Correctly predicted YES
TN	Correctly predicted NO
FP	Incorrectly predicted YES
FN	Incorrectly predicted NO
Accuracy	Overall correctness
Precision	Correct YES predictions among all predicted YES
Recall	Correct YES predictions among all actual YES
F1 Score	Balance of Precision and Recall
Overfitting	Model learns too much from training data
Underfitting	Model learns too little
Cross-validation	Testing model on different parts of the dataset
Bias	Error from wrong assumptions
Variance	Error from too much complexity

By learning and applying these concepts, you can become a better AI developer and understand how to improve your models effectively.
