WEEK 3: CORE MATHEMATICS FOR MACHINE LEARNING

DAY 13 (09/07/2025)

Probability and Statistics for Machine Learning:

Probability and statistics form the backbone of machine learning. They help models **understand uncertainty, detect patterns**, and **make predictions** based on data. Many ML algorithms are deeply based on these mathematical ideas.

1. Probability Basics

Probability measures how likely an event is to happen. It ranges from **0 to 1**, where

- 0 means the event will **never** happen,
- 1 means it will **definitely** happen,
- and values in between show varying degrees of likelihood.

$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$$

Example:
When flipping a fair coin, the probability of getting heads = **1/2 = 0.5**.
This means there's a 50% chance of getting heads in a single toss.

In Machine Learning:
Probability helps models make predictions — for example, predicting whether an email is **spam or not spam** is based on the probability of certain words appearing in spam emails.

2. Random Variables

A **random variable** represents outcomes of a random process in numeric form.
It can take on **different values**, each with an associated probability.

- **Example:** Marks of students in an exam, number of rainy days in a month, or the number of website visitors per day.

In Machine Learning:
Random variables help in defining data distributions and understanding patterns like "How likely is it that a person buys a product if they've clicked on an ad?"

3. Key Statistical Measures

Statistics helps us **summarize** and **analyze** data so that we can understand its behavior.

a. Mean (Average)

It represents the **central value** of data.

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Total number of values}}$$

Example:
For marks [80, 90, 70],
Mean = (80 + 90 + 70) / 3 = 80.

b. Variance

Variance shows **how far the data values are spread** from the mean.
A higher variance means the data points are more spread out.

c. Standard Deviation

It is the **square root of variance** and shows **how much the data deviates** (on average) from the mean.

d. Covariance and Correlation

They measure **how two variables move together**.

- **Covariance** tells whether variables increase or decrease together.

- **Correlation** shows how **strongly** they are related, and it always lies between -1 and +1.

Example:
If "hours studied" and "exam marks" have a **high positive correlation**, it means students who study more tend to score higher.
If the correlation is **close to zero**, it means one has little to do with the other.

In Machine Learning:
Correlation helps in **feature selection** — we often remove features that are highly correlated because they provide duplicate information.

4. Probability Distributions

A **distribution** shows how data or outcomes are spread across possible values. It helps us visualize and model real-world patterns.

Common Distributions in ML:

- **Normal (Gaussian) Distribution:**

    - The most common distribution, where most data points lie near the mean, and fewer are at the extremes.

    - Shape: Bell curve.

    - **Example:** Human heights, exam scores.

- **Bernoulli Distribution:**
    - For binary outcomes — success/failure, yes/no, 0/1.
    - **Example:** Tossing a coin (Head or Tail), predicting if an email is spam (Yes or No).

- **Binomial Distribution:**
    - Represents the number of successes in repeated independent trials.
    - **Example:** Number of heads in 10 coin tosses.

- **Uniform Distribution:**
    - All outcomes are equally likely.
    - **Example:** Rolling a fair dice — each face (1 to 6) has equal probability.

In Machine Learning:
Distributions help models understand data behavior.

Reflection:

Today I learned that **probability and statistics play a major role in Machine Learning.** They allow models to deal with uncertainty, summarize large datasets, and find meaningful relationships between variables. By understanding these concepts, we make our predictions more accurate, confident, and trustworthy.