Training Day-17 Report:

Normalization and Standardization in Machine Learning:-

Normalization and standardization are two techniques used to transform data into a common scale. Normalization is a technique used to scale numerical data in the range of 0 to 1. This technique is useful when the distribution of the data is not known or when the data is not normally distributed. On the other hand, standardization is a technique used to transform data into a standard normal distribution. This technique is useful when the distribution of the data is known and when the data is normally distributed. Both techniques have different applications, and choosing the right technique based on the data and the problem you're trying to solve is important.

What is Normalization?

Normalization in machine learning is a data preprocessing technique used to change the value of the numerical column in the dataset to a common scale without distorting the differences in the range of values or losing information.

In simple terms, Normalization refers to the process of transforming features in a dataset to a specific range. This range can be different depending on the chosen normalization technique.

The two most common normalization techniques are Min-Max Scaling and Z-Score Normalization, which is also called Standardization.

Now, let's discuss Min-Max Scaling.

Min-Max Scaling

This method rescales the features the features to a fixed range, usually 0 to 1. The formula for calculating the scaled value of a feature is:

Normalized Value = Value - Min/ Max - Min

where,

Value: Original Value of the feature

Min: Minimum value of the feature across all the data points.

Max: Maximum value of the feature across all the data points.

What is Standardization?

Standardization is a data preprocessing technique used in statistics and machine learning to transform the features of your dataset so that they have a mean of 0 and a standard deviation of 1. This process involves rescaling the distribution of values so that the mean of observed values is aligned to 0 and the standard deviation to 1.

Standardisation aims to adjust the scale of data without distorting differences in the ranges of values or losing information.

Unlike other scaling techniques, standardization maintains all original data points' information (except for cases of constant columns).

It ensures that no single feature dominates the model's output due to its scale, leading to more balanced and interpretable models.

Formula of Standardization

Z = (x-mean)/standard deviation