

Normalization and standardization are both techniques used to rescale data, but they serve different purposes and are applied differently:

Normalization

- **Purpose:** To rescale the data to a fixed range, typically [0, 1].
- **Method:** Each feature is scaled independently to fit within the specified range.
- **Formula:**

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

- **Use Case:** Useful when you want to ensure that all features have the same scale, especially when the data has different units or ranges. It's often used in algorithms that do not assume any particular distribution of the data, such as K-Nearest Neighbors (KNN) and neural networks.

Standardization (Standard Scaler)

- **Purpose:** To transform the data to have a mean of 0 and a standard deviation of 1.
- **Method:** Each feature is centered around the mean and scaled by the standard deviation.
- **Formula:**

$$X_{\text{standardized}} = \frac{X - \mu}{\sigma}$$

where (μ) is the mean and (σ) is the standard deviation.

- **Use Case:** Useful when the data follows a normal distribution or when you want to ensure that the features have the same variance. It's often used in algorithms that assume normally distributed data, such as linear regression, logistic regression, and support vector machines (SVM).

Key Differences

- **Normalization** scales the data to a specific range, while **standardization** scales the data to have a mean of 0 and a standard deviation of 1.
- **Normalization** is applied row-wise (each sample), whereas **standardization** is applied column-wise (each feature)