



Standardization vs Normalization in ML

Both are **feature scaling techniques** used to **rescale numeric data** so that machine learning models perform better.

1. Standardization

Rescales features so they have a **mean = 0** and **standard deviation = 1**.

Formula:

$$z = \frac{x - \mu}{\sigma}$$

Where:

- x = value
- μ = mean of feature
- σ = standard deviation

Use When:

- Data has **outliers** or **different units**
 - Works well for **algorithms that assume normal distribution** (e.g., Logistic Regression, SVM, Linear Regression)
-

2. Normalization

Rescales features to a **fixed range**, usually **[0, 1]**.

Types of Normalization Techniques

1. Min-Max Scaling

Most common type of normalization

Formula:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

- All values are transformed to **[0, 1]**
 - Sensitive to **outliers**
-

Example: Salary = [30K, 40K, 50K] → scaled to [0.0, 0.5, 1.0]

2. Mean Normalization

Formula:

$$x_{\text{scaled}} = \frac{x - \mu}{x_{\text{max}} - x_{\text{min}}}$$

- Centers data around **0**
 - Range is still **[-1, 1]** (roughly)
-

3. MaxAbs Scaling

Formula:

$$x_{\text{scaled}} = \frac{x}{|x_{\text{max}}|}$$

- Scales data to range **[-1, 1]**
 - Good for **sparse data**
 - Keeps **zero entries** untouched
-

4. Robust Scaling

Formula:

$$x_{\text{scaled}} = \frac{x - \text{median}}{\text{IQR}}$$

Where IQR = Interquartile Range (Q3 - Q1)

- Handles **outliers well**
 - Centers around median instead of mean
-

5. L2 Normalization (also known as just Normalization in sklearn)

- Often used for **text data or deep learning**
 - Scales rows (samples) to unit norm
-

Standardization vs Normalization – Key Differences

Feature	Standardization	Normalization (Min-Max)
Output Range	No fixed range (mean = 0, std = 1)	[0, 1] (or [-1, 1])
Sensitive to Outliers?	Less sensitive (uses mean & std)	Yes, especially Min-Max
Use Cases	Linear Models, SVM, PCA	Neural Networks, Image data
Scaling Method	Centered around 0	Scaled between 0 and 1
Transformation Based On	Mean and Standard Deviation	Min and Max values

Final Thought

- **Use Standardization** when your data has outliers or needs to be centered.
 - **Use Normalization** when you want a bounded scale or are using algorithms like neural networks or kNN.
-