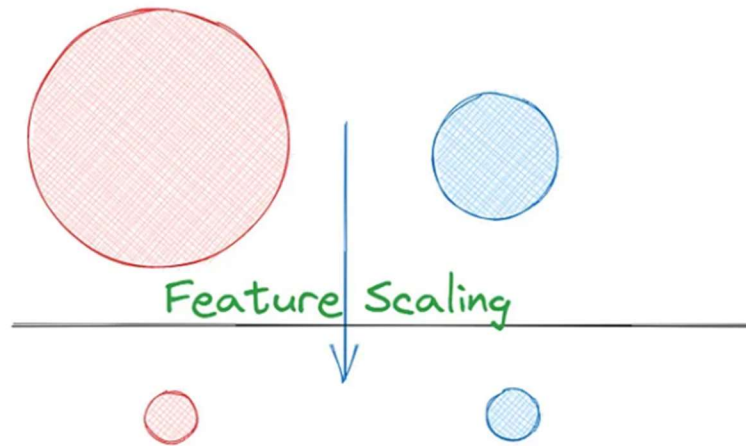


Feature Scaling

Feature scaling is a crucial step in data preprocessing for machine learning. It involves adjusting the range of features in your data so they are consistent.



Types of Feature Scaling:

Standardization:

- Standard Scaler

Normalization:

- Min Max Scaling
- Mean Normalization
- Max Absolute Scaling
- Robust Scaling etc.

Definition:

Standardization (Z-Score Scaling): Adjusts data to have a mean of 0 and standard deviation of centering data and scaling by its standard deviation.

Min-Max Scaling (Min-Max Normalization): Rescales data to a fixed range, typically $[0, 1]$, based on the minimum and maximum values.

Mean Normalization: Scales data to center around 0 with values typically in the range $[-1, 1]$, based on mean and range.

Max Absolute Scaling (MaxAbs Scaling): Divides each feature by its maximum absolute value, preserving the sparsity of the data.

Robust Scaling (Robust Scaler): Scales data using the median and interquartile range (IQR), making it less sensitive to outliers.

Scaling Method	Alternative Name	Description	Formula	Use Cases
Standardization	Z-Score Scaling	Scales features to have mean = 0 and std = 1	$X_{std} = \frac{X - \mu}{\sigma}$	Useful for data with Gaussian distribution
Min-Max Scaling	Min-Max Normalization	Scales features to a fixed range (0, 1)	$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$	When features have different units or ranges
Mean Normalization	-	Centers features around 0 with a fixed range	$X_{mn} = \frac{X - \mu}{X_{max} - X_{min}}$	Similar to Min-Max, but centers data around 0
Max Absolute Scaling	MaxAbs Scaling	Scales features by their maximum absolute value	$(X_{\{maxabs\}} = \frac{X}{\max(X)}$	$X_{\{max\}}$
Robust Scaling	Robust Scaler	Uses median and IQR, robust to outliers	$X_{robust} = \frac{X - \text{median}(X)}{\text{IQR}(X)}$	When data has outliers