Data Cleaning and Normalization

Min-Max Normalization

Definition

- Min-max normalization is a technique to rescale the values of a feature to a fixed range, usually [0, 1].
- This is done by transforming each value in the feature such that it fits within the specified range

Formula

The formula for min-max normalization is:

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

Where:

- X is the original value.
- X_{min} is the minimum value in the feature.
- X_{max} is the maximum value in the feature.
- $X_{\text{normalized}}$ is the normalized value.

Example

Original Value
10
15
20
25
30

Step-by-Step Calculation:

1. **Identify the minimum and maximum values:**
   - $X_{\text{min}} = 10$
   - $X_{\text{max}} = 30$
2. **Apply the min-max normalization formula:**

   For each value in the dataset:

   - For $X = 10$:

$$X_{\text{normalized}} = \frac{10 - 10}{30 - 10} = \frac{0}{20} = 0$$

- For $X = 15$:

$$X_{\text{normalized}} = \frac{15 - 10}{30 - 10} = \frac{5}{20} = 0.25$$

- For $X = 20$:

$$X_{\text{normalized}} = \frac{20 - 10}{30 - 10} = \frac{10}{20} = 0.5$$

- For $X = 25$:

$$X_{\text{normalized}} = \frac{25 - 10}{30 - 10} = \frac{15}{20} = 0.75$$

- For $X = 30$:

$$X_{\text{normalized}} = \frac{30 - 10}{30 - 10} = \frac{20}{20} = 1$$

3. **Normalized Values:**

Original Value	Normalized Value
10	0.0
15	0.25
20	0.5
25	0.75
30	1.0