Feature Scaling

Tuesday 14 May 2024 5:41 AM

1) standardization

$$z = \frac{x - \mu}{\sigma}$$ — mean centering, scaling — $z \sim (0, 1)$

→ Covariance will be maintained

pros – simple and efficient to implement

cons – Doesn't work with algos that assume non-negative values.

StandardScaler [sklearn]

→ Distribution is maintained [individual feature]

→ Outliers will still be outliers

use cases:
- Features are normally distributed
- Dealing with algos that benifit from data centered around 0.

2) MinMax Scaling

$$\frac{x_i - x_{min}}{x_{max} - x_{min}}$$ [0, 1]

MinMaxScaler [sklearn]

→ Distribution & Relative Distances are maintained

→ sensitive to outliers

→ Zero values transformation

→ Reversibility

use cases:
- When there is a bounded scale for features
- When data doesn't exhibit strong normal distribution
- When working with algos which don't work well with -ve values

3) Robust Scaler

→ Reduces the impact of outlier

→ outlier present & important

$$x_i' = \frac{x_i - \text{median}}{IQR}$$

→ Median centering

→ Distribution is maintained

→ No Fixed Range for scaled data

→ Reversibility

4) Max Absolute Scaler

→ sparse data → lots of 0's (has meaning)

$$x_i' = \frac{x_i}{\max(|x|)}$$

[-1, +1] original data +ve & -ve
[0, 1] original data +ve
[-1, 0] original data -ve

→ Preservation of zero

→ sensitive to outliers

→ sign is preserved

→ sensitive to outliers

→ Sign is preserved

→ Relative distances maintained

→ No centering

→ Reversible

→ Distribution maintained

5) L2/L1 Normalization [multivariate scaling technique]

→ Row wise scaling

$$X_i' = \frac{X_i}{\text{L2 Norm of row}}$$

$$X_i' = \frac{X_i}{\text{L1 Norm of row}}$$

Use cases (vectors)
- When Rows are more important than cols
- Text processing → sparse features → L1
- similarity based algos
- Image processing
- Neural Network embeddings

↑ dimension L1 (manhatten distance)

→ similarity / ranking type problems