

→ When you want to find the distance between two points in any of the algo. then you must use standardisation.

⇒ Normalization:- (Day 25):-

"Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without disturbing diff. in the ranges of values or losing information."

→ Normalization etle ek process che jema numerical features na values ne ek common range ma lai apva ma aave che, mostly 0 to 1.

⇒ Types of Normalization:-

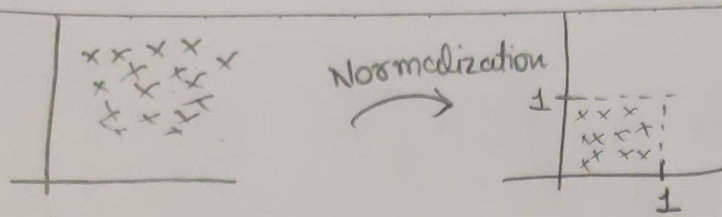
(1) Min-Max Scaling - Intuition (Max. used. default)

→ "Min-Max Scaling is a normalization technique that rescales the values of a feature to a fixed range, usually 0 to 1 (but it can be any range you choose)."

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

→ Min-Max Scaling etle ek technique. che je feature na values ne ek fixed range ma convert kare che, mostly 0 to 1.

Aa method ma min. value ne 0 banavi devay che one max. value ne 1 banavi devay che, one baki na baolha value ne ena vachche proportion ma scale karva ma aave che.



(2) Mean Normalization:-

→ "Mean Normalization is a feature scaling technique where the values of a feature are rescaled such that the mean of the feature becomes 0."

→ Formula:-

$$x'_i = x_i - \bar{x}$$

→ Mean centering

Values:- -1 to 1

$x_{\max} - x_{\min}$: convert in bet

→ Mean Normalization etle ek feature scaling technique che jema feature na values ne rescale kari ne mean 0 banavy che.

Aa method ma data ne shift ane scale karva ma ave che feature na mean ane range (max-min) ni maded thi.

(3) Max. Absolute Scaling:-

→ Max Absolute Scaling etle ek feature scaling technique che jema feature na values ne unscaled max. value thi divide kari ne scale karva ma ave che.

→ Range:- -1 to 1

→ Useful for sparse data. (jyu zero value is imp.)

$$x'_i = \frac{x_i}{|x_{\max}|}$$

→ From sklearn
class: MaxAbsScale

(4) Robust Scaling:- (Use when:- We have outliers in our data)

"Robust Scaling is a feature scaling technique where features are scaled using the median and interquartile range (IQR) instead of the mean and standard deviation."

- This method is less sensitive to outliers compared to Min-Max Scale or standardization.
- It is useful when your data contains extreme values that could distort the scaling.

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

IQR

$$\frac{Q_3 - Q_1}{75\%}$$

→ Robust Scaling etle ek feature scaling technique che jema feature na values ne median and interquartile range (IQR) thi scale kari ne outliers ni effect avoid kariye che.

⇒ For what type of which scaling should be used:-

- Standardization:- Mostly used (commonly)
- Min-Max :- When you know the min. & max. value.
- Robust Scaling :- When you have outliers.
- Max. Absolute Scaling:- When you have sparse data.