

# Feature Scaling in ML

- Feature Scaling का मतलब होता है **डेटा के अलग-अलग features को एक ही scale पर लाना**, ताकि कोई feature ज़्यादा बड़ा या छोटा होने की वजह से model को bias न करे।
- जैसे: Age  $\rightarrow$  0 से 100 and Salary  $\rightarrow$  10,000 से 10,00,000
- Salary का effect ज़्यादा हो जाएगा, इसलिए scaling ज़रूरी है।
- Jaise school mein sabhi bachoo ko ek dress mein padhne jana hota hai. Taki PM/CM/DR./Poor kisi ka beta ho sabko ek saman ek dress mein sako ahmiyat mitle nhi to Bade log chote ho daba denge dominate kr denge chote logo ko ahmiyaat ko khatam kr denge.



# Types of Feature Scaling

- MinMax Scaling(Normalization) → Set the feature range between  $[0, 1]$
- Standardization(z-score Scaling) → Mean =0, Standard deviation =1
- Robust (scaling)

# MinMax Scaling(Normaliztion):

- **MinMax Scaling(Normaliztion):** Min-Max Scaling एक technique है जिसमें हम data को **fixed range** में बदल देते हैं, आमतौर पर **0 से 1** के बीच।
- इसमें सबसे छोटा value **0** बन जाता है
- सबसे बड़ा value **1** बन जाता है
- बाकी values इनके बीच आ जाती हैं
- Jaise school mein sabhi bachoo ko ek dress mein padhne jana hota hai. Taki PM/CM/DR./Poor kisi ka beta ho sabko ek saman ek dress mein sako ahmiyat mitle nhi to Bade log chote ho daba denge dominate kr denge chote logo ko ahmiyaat ko khatam kr denge.

# Formula of Min-Max Scaling

- Min-Max Scaling का Formula

## Min-Max Scaling का Formula

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- जहाँ:  $X$  = original value
- $X_{min}$  = feature का minimum value
- $X_{max}$  = feature का maximum value
- $X_{scaled}$  = scaled value

# Example of Min-Max Scaling



- **Machine Learning** = Data + Learning + Prediction
- मान लो हमारे पास **Age** का data है:

- Age  
10  
15  
20  
25

Min = 10

Max = 25

अब अगर Age = 15 को scale करना है:

$$\frac{15 - 10}{25 - 10} = \frac{5}{15} = 0.33$$

तो scaled value  $\approx$  **0.33**



# Where we use Min-Max Scaling

- KNN (K-Nearest-Neighbors)
- SVM(Support Vector Machines)
- K-means Clustering
- PCA(Principle component analysis)
- Gradient Descent Based Algorithm
- GMM(Gaussian-Mixture Model)

# Standardization(Z-Score Scaling)

- Standardization (Z-Score Scaling) एक feature scaling technique है जिसमें data का mean 0 और standard deviation 1 कर दिया जाता है, ताकि सभी features समान scale पर आ जाएँ और ML models बेहतर perform करें।
- Algorithms हर feature को बराबर importance देते हैं।
- Data का **mean (औसत) = 0**
- Data का **standard deviation = 1**
- यानी data को एक common scale पर ले आते हैं।

# Formula

- Z-Score Scaling का Formula (Math Formula)

-  **Formula:**

$$Z = \frac{X - \mu}{\sigma}$$

- जहाँ:
- **X** = original value
- **μ (mu)** = mean (औसत)
- **σ (sigma)** = standard deviation



# Example on Formula

- मान लो हमारे पास ये data है (Feature: Marks)
- 50, 60, 70, 80, 90
- Step 1: Mean (औसत) निकालो

$$\text{Mean} = \frac{50 + 60 + 70 + 80 + 90}{5} = \frac{350}{5} = 70$$

# Example of SD

- Feature data(Marks )  $\rightarrow$  50, 60, 70, 80, 90
- **Step 2: Standard Deviation निकालो**

✦ Formula:

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{n}}$$

अब हर symbol का मतलब:

- $X \rightarrow$  data की value
- $\mu$  (mu)  $\rightarrow$  mean (औसत)
- $n \rightarrow$  total values
- $\Sigma \rightarrow$  सबको जोड़ना
- $\sqrt{\phantom{x}} \rightarrow$  square root

# Example of SD

- 50 60 70 80 90

**Step 1: Mean (औसत) निकालो**

$$\text{Mean} = \frac{50 + 60 + 70 + 80 + 90}{5}$$
$$= \frac{350}{5} = 70$$

👉 **Mean = 70**

# Example of SD

## ◆ Step 2: Mean से दूरी ( $X - \text{Mean}$ )

X	$X - 70$
50	-20
60	-10
70	0
80	10
90	20

# Example of SD

## ◆ Step 3: Square करो

$X - 70$	$(X - 70)^2$
-20	400
-10	100
0	0
10	100
20	400

# Example of SD

## ◆ Step 4: Average लो (Variance)

$$\begin{aligned}\text{Variance} &= \frac{400 + 100 + 0 + 100 + 400}{5} \\ &= \frac{1000}{5} = 200\end{aligned}$$

## ◆ Step 5: Square Root लो (Standard Deviation)

$$\begin{aligned}\text{Standard Deviation} &= \sqrt{200} \\ &\approx 14.14\end{aligned}$$

✓ Final Answer:

📌 Standard Deviation  $\approx 14.14$

# Now Finally Z-Score



**Data:**

50      60      70      80      90

हम पहले ही निकाल चुके हैं:

- Mean ( $\mu$ ) = 70
- Standard Deviation ( $\sigma$ ) = 14.14

1 2  
3 4

**Z-Score का Formula**

$$Z = \frac{X - \mu}{\sigma}$$

# Now Finally Z-Score

## ◆ Step-by-Step Z-Score Calculation

◆ **1** X = 50

$$Z = \frac{50 - 70}{14.14} = \frac{-20}{14.14} \approx -1.41$$

◆ **2** X = 60

$$Z = \frac{60 - 70}{14.14} = \frac{-10}{14.14} \approx -0.71$$

◆ **3** X = 70

$$Z = \frac{70 - 70}{14.14} = 0$$

◆ **4** X = 80

$$Z = \frac{80 - 70}{14.14} = \frac{10}{14.14} \approx 0.71$$

◆ **5** X = 90

$$Z = \frac{90 - 70}{14.14} \downarrow \frac{20}{14.14} \approx 1.41$$

## ✓ Final Z-Score Table

Original Value	Z-Score
50	-1.41
60	-0.71
70	0
80	0.71
90	1.41





# When we use Standardization

- जब data **normal distribution** के आस-पास हो
- जब outliers हों
- जब ML algorithm distance पर depend करता हो
- **Distance-based Algorithms** के लिए ज़रूरी
- **Example:** Linear Regression, Logistic Regression, KNN, SVM, Neural Networks ये **distance** पर काम करते हैं।

# Where No need Standardization

- कब Standardization नहीं करना चाहिए?
- ✗ Decision Tree
- ✗ Random Forest
- क्योंकि ये **splitting** पर काम करते हैं, distance पर नहीं।

# What



- **Machine Learning** = Data + Learning + Prediction



# What



- **Machine Learning** = Data + Learning + Prediction



# What



- **Machine Learning** = Data + Learning + Prediction



# What



- **Machine Learning** = Data + Learning + Prediction

# What



- **Machine Learning** = Data + Learning + Prediction

# What



- **Machine Learning** = Data + Learning + Prediction