

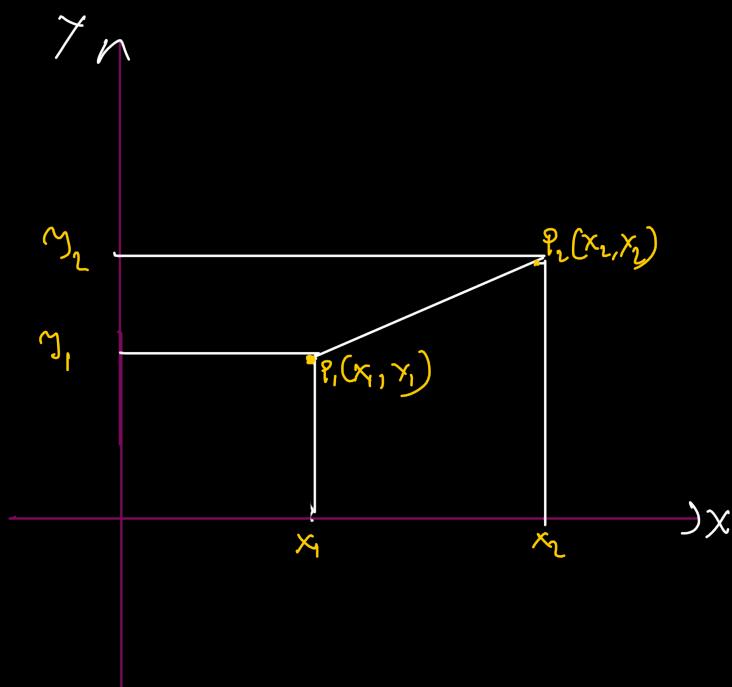
## Feature Scaling (Standardization)

- ↳ Last Step in feature engineering.
- ↳ Just before we give data to machine learning model, we do feature scaling.
- ↳ one of the easiest step in FE.

what is Feature Scaling?

[ Feature Scaling is a technique to Standardize the independent features present in the data in a fixed range. ]

why do we need Feature Scaling?



Euclidean distance between  $P_1$  and  $P_2$  =

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Let  $\pi$  be the no. of Salary and  $y$  be the no. of Age

Example:  $\pi_1$  &  $y_1$  are in row 2,  
 $\pi_2$  &  $y_2$  are in row 9

$$(\pi_2 - \pi_1)^2 = (83000 - 48000)^2 \\ = 1225000000$$

$$(y_2 - y_1)^2 = (50 - 27)^2 = 529$$

→ There are many ML Algo, which works distance b/w two points.

for those Algo Scaling is must for better result.

→ If our IV are not scaled properly then there are many ML Algo which logic will not perform well.

## Types of Feature Scaling

### feature Scaling

#### Normalization

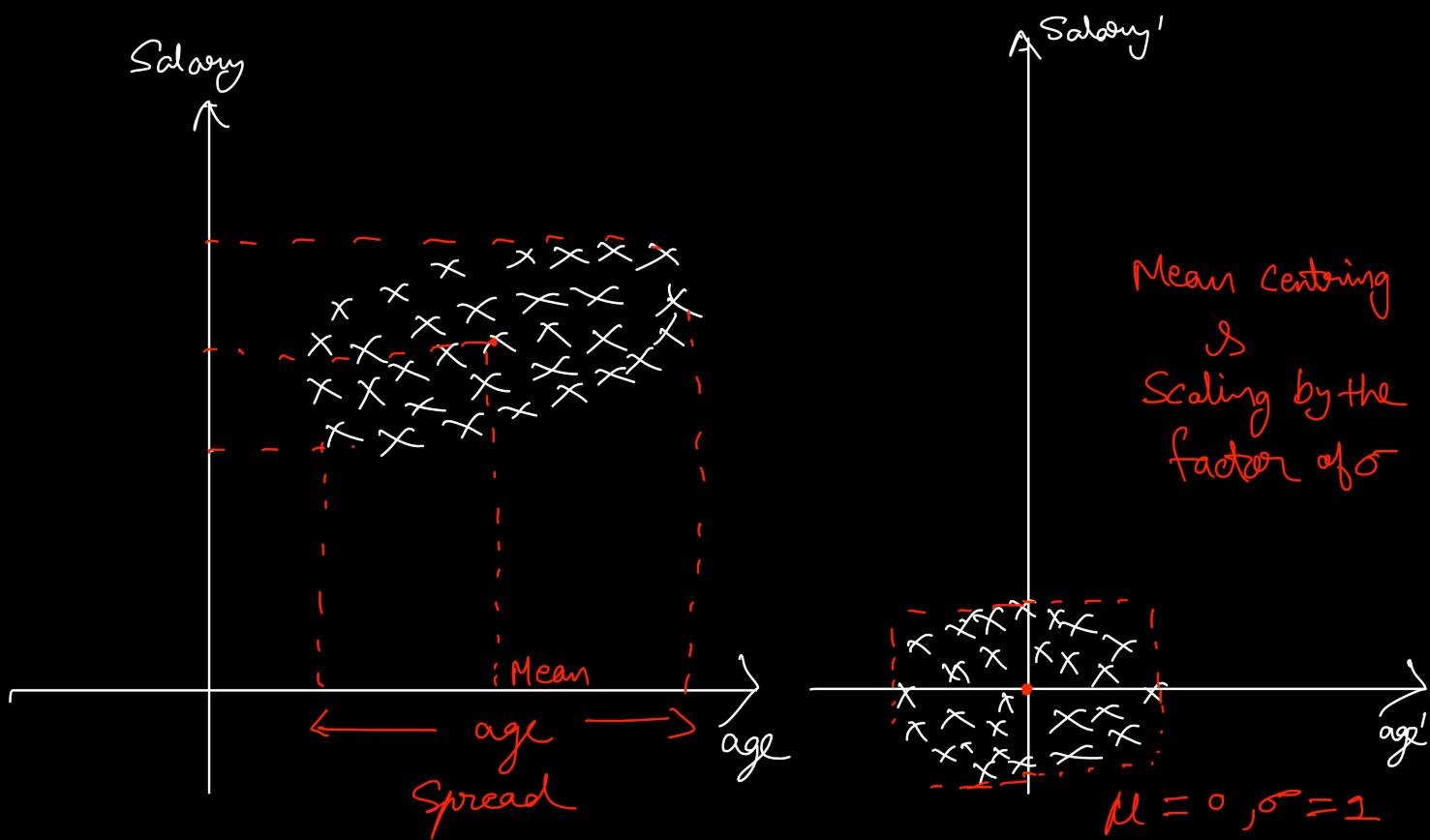
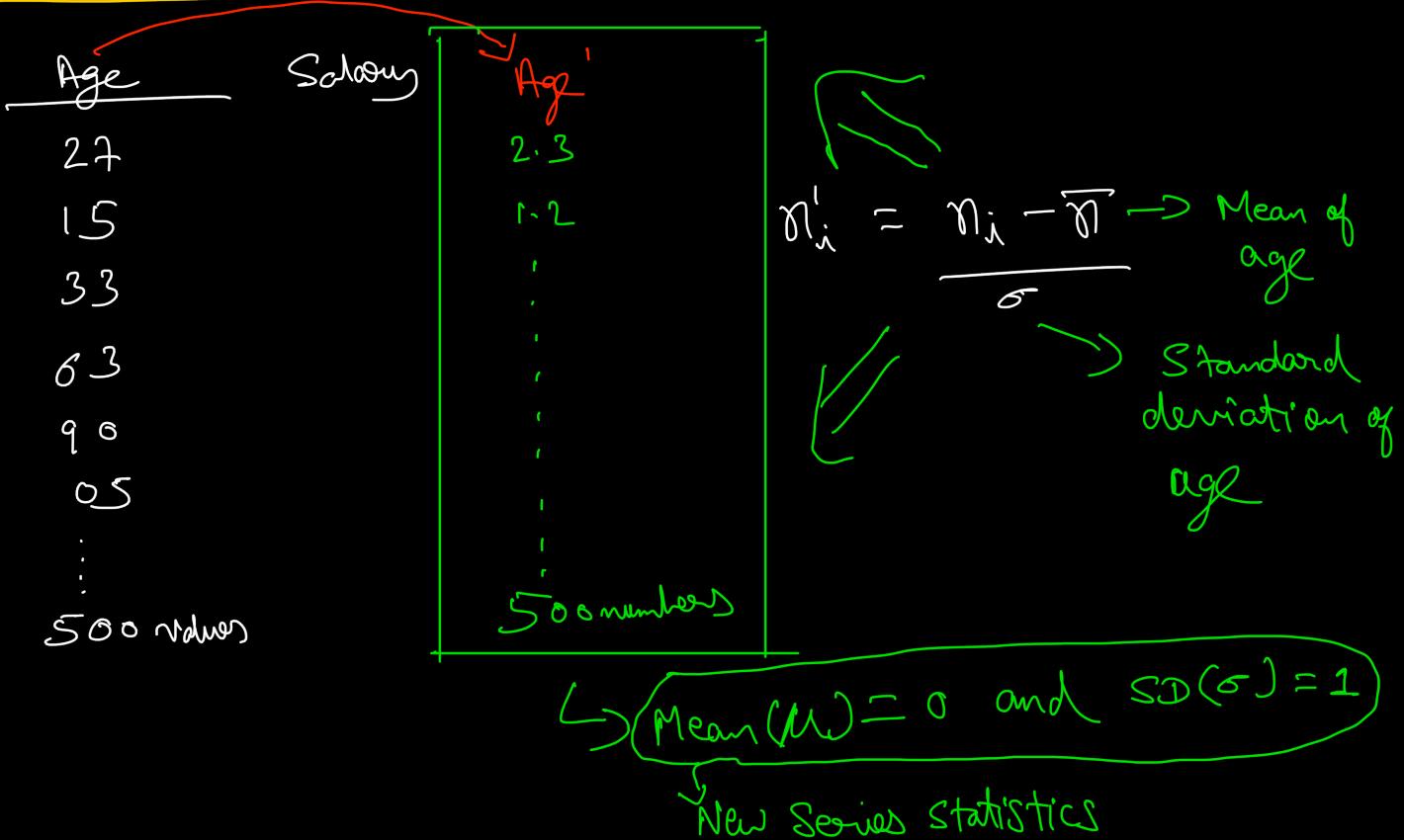
- ↳ Min Max Scaler
- ↳ Robust Scaler

#### Standardization

also known as Z-Score normalization

## Standardization - Intuition

Also called as Z-Score Normalization



## When to use Standardization?

→ First think to note that, whenever we will use Standardization it will not give any harm if it not gives any profit then (कोई फायदा नहीं होता है)

Algorithms → Reason of applying feature scaling

1. K-Means → use of Euclidean distance measure
2. K-Nearest-Neighbours → Measure the distance between pairs of samples and these distances are influenced by the measurement units.
3. Principal Component Analysis → Try to get the feature with maximum variance (PCA)
4. Artificial Neural Network → Apply Gradient Descent
5. Gradient Descent → Theta calculation becomes faster after feature scaling and the learning rate in the update equation of stochastic gradient descent is the same for every parameter.  
without scaling converging becomes hard.

few algorithms in which Standardization does not bring any changes in score or accuracy

- Decision Tree
- Random Forest
- XGBoost, Gradient Boost
- All Tree based Algorithms

## Few Points to note down (Regarding standardization)

- It is recommended to do train test split before performing Scaling (Standardization).
- Scaler learn from training data but transform both train and test data.
- There is no change in shape of distribution of data after scaling (Standardization).
- only scale changes.
- Mean and SD changes.
- There is no change in shape of distribution of any column . only Scale changes.
- Shape of distribution of data never changes when Scaling (Standardization) done.
- Mean will become zero (Mean Centring) and Standard Deviation will become 1 (Scaling by the factor of SD)
- Scaling (Standardization) is important when we use Logistic Regression.
- After scaling (Standardization) also impact of outliers do not changes.
- outliers then are now also outliers.
- So scaling (Standardization) does not do anything to outliers will continue behaving like outliers only after Scaling (Standardization) also .  
So if we find any outliers in our data then we have to handle it before Scaling .