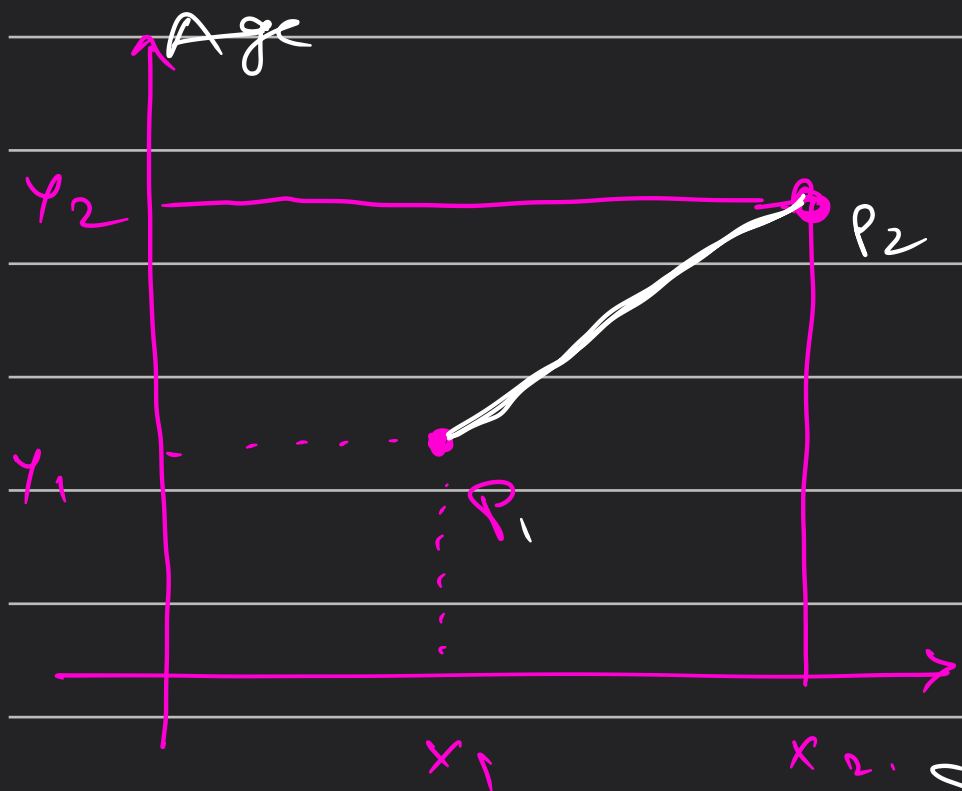# What is feature scaling?

A technique to standardize the independent features present in data in a fixed range.

# Why do we need feature scaling?



Euclidean Distance

$$P_1 - P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$(x_2 - x_1)^2 = 12 < 500000$$

$$(y_2 - y_1)^2 = 589$$

Huge Difference btw both values

⇓

This difference created due to different range causes problem in Machine learning algo specially KNN & logistic Regression ek.

# Standardization [Feature Scaling Technique]

Also called Z-Score Normalization

For the features that being used we create standard values using formula

set $A_i$ be Age $A_i'$ which is standard value would be given by

$$A_i' = \frac{A_i - \boxed{\mu A}}{\boxed{\sigma A}} \quad \begin{array}{l} \rightarrow \text{mean} \\ \rightarrow SD \end{array}$$

value

Through the process of standarization we achieve something called "Mean Centering"

For the converted standardized value the mean becomes 0 and SD becomes around 1

## 📑 Impact of Outliers

Even when values are scaled after standardization, the Outlier values affect the same to the data. There is no difference in the Impact of Outlier values before & after the scaling through standarization

## 📑 When Can we use Standarization for Scaling.

| Algo | Reason |
|---|---|
| K-Means | when Eucliden distance gets compared |
| K-Nearest | distance btw pairs gets impacted by measurement Unit |
| PCA | getting feature with max Variance |
| ANN/ grad Descent | when grad descent is applied we need standarized values. |