Feature Eng
↓
Feature Trans.

FEATURE SCALING

Date

Note: usually we apply feature scaling at the end, means after all other feature transformations.

Agenda:
1. What is feature scaling ?
2. Why do we need feature scaling ?
3. Types of feature scaling.
4. Standardization
5. Normalization
6. Normalization Vs Standardization. (depends on algorithm)

∴ What is feature scaling ?

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

Aim is to bring all the features (independent) on same scale.

Ex. Features --> Age, Height, Weight
(number ↑, cm. ↑, kg. ↑) ← different units

So, these 3 features are now not comparable
To make them comparable we do scale them.

∴ Why do we need feature scaling ?

age	salary	purchase (Y/N)
10	10,000	0
15	25,000	1

Say, this is classification problem having age & salary as independent features & based on that we need to find purchase (Y/N).

Say, we are using KNN algorithm, which is distance based algorithm.

STUDY BUDDIES

#

Date

so, distance = $\sqrt{(15-10)^2 + (25000-10,000)^2}$

This factor will dominate

There won't be much impact of Age feature

so, as a result KNN won't perform well.

Types of feature scaling:

- Standardization (Z-score normalization)
- Normalization.

Standardization:

Age	Standardized_age
10	$\left(\frac{10-30}{10}\right) = -2$
20	$\left(\frac{20-30}{10}\right) = -1$
30	$\left(\frac{30-30}{10}\right) = 0$
40	$\left(\frac{40-30}{10}\right) = 1$

formula: $\left\{ x_i' = \frac{x_i - \bar{x}}{\sigma} \right\}$

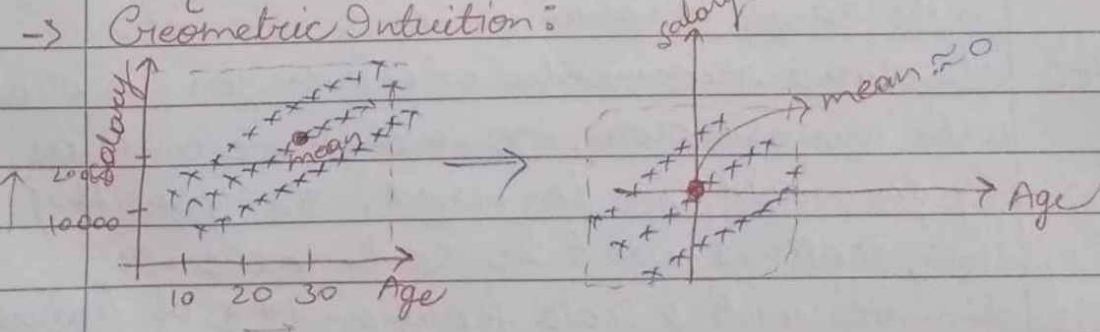
$\{ \bar{x} = \text{mean}, \ \sigma = \text{Std. dev.} \}$.

(say, $\bar{x} = 30$, $\sigma = 10$).

→ Property of standardized value:

{ mean ≈ 0, std. dev. ≈ 1 }.

→ Geometric Intuition:



Before Standardization

After Standardization of Age & salary both.

{ mean centering (mean ≈ 0)
squashed std. deviation ($\sigma \approx 1$)
{ scaling by the factor of std. dev. } }

STUDY BUDDIES

#_____

Date

→ Effect on outliers ?

There is no impact of standardization on outliers. we have to explicitely deal with it.

→ When to use Standardization ?

There is no loss on applying Standardization. But, it will benefit us when applied on distance based algo as, (i) K-means --> uses Euclidean Distances

(ii) KNN --> measures distances b/w pairs of data.

(iii) PCA --> try to get feature with max. variance

(iv) ANN --> Apply gradient descent

(v) Gradient Descent --> θ calculation becomes faster & learning rate is same.

(vi) Deep learning --> Gradient descent.

✳ Normalization:

A technique often applied as part of data preparation for ML. The goal of normalization is to change the values of numeric columns in dataset to use a common scale. without distorting differences in the ranges of values or losing information.

Note: Its always recommended whenever you are dealing with numeric data, try to eliminate units as, (kg for weight, cm. for height) bcz if we keep them, features won't be comparable. So, eliminate units & scale features. (bring all feature to common scale).

Types Normalization

- Min Max Scaling ↑ most used
- Mean normalization
- Max absolute
- Robust scaling

STUDY BUDDIES

#

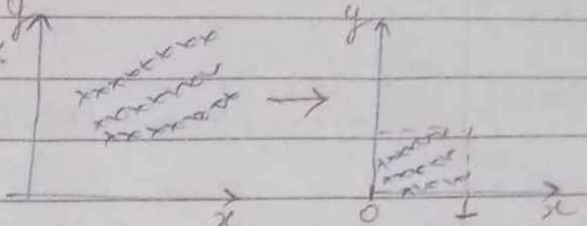
Date

→ Min-Max Scaling :

formula: $x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$ ← when we know min & max.

Range → [0, 1].

• Geometric Intuition:



after scaling x & y, data is squished blw [0, 1].

unit sq., cube(3D), hyper(+3D)

• Note: here shape of distribution after scaling may change, unlike in Standardization.

• Effect on outliers: squishes.

→ Mean Normalization :

formula: $x_i' = \frac{x_i - x_{mean}}{x_{max} - x_{min}}$ → mean centering
↳ scaling blw factor (max - min).

Range → (-1, 1)

Note: { Rarely used (instead of this Standardization is used)
{ There is no separate class for this in Sklearn.

→ Max Absolute Scaling :

formula: $x_i' = \frac{x_i}{|x_{max}|}$ → used when we have sparse data (data with many zeros).

[class in sklearn → MaxAbsScalar].

→ Robust Scalar :

$x_i' = \frac{x_i - x_{median}}{IQR = Q_3 - Q_1}$ ← Robust to outliers (works better with data having outliers)

[class RobustScalar].

Note: can't say which technique works best for data, we have to try & choose.
(Most of the times Standardization is used.)
#

STUDY BUDDIES