DAY-5

Feature Engineering :

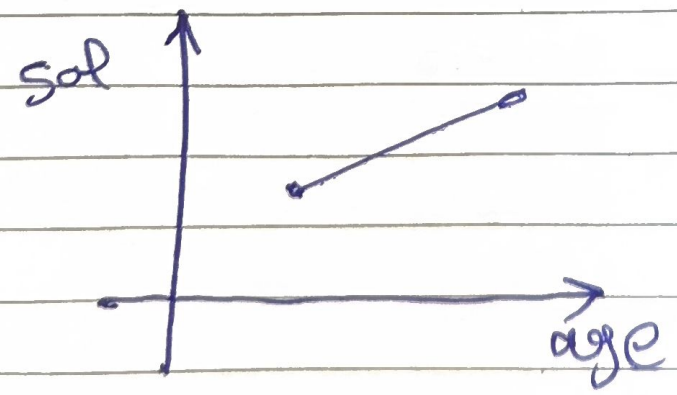
- Feature Transformation
- Feature Construction.
- Feature selection
- Feature Extraction.

⇒ Feature Transformation :

- Missing value imputation.
- Handelling categorical features.
- Outlier Detection.
- Feature Scaling.

1) Feature Scaling :

age	Salary	purchase
50	83000	1
27	48000	0



- for e.g: KNN when we take the euclidean distance the distance may vary too much. for salary. so we scale in the same rang, so that it works as a same unit

Feature
Scaling : 1) Standardization
2) Normalisation.

1] Standardization.

Age	Sal	Age'
27	☺	
15	☹	
33	☺	
63	☺	

$\frac{x_i - \bar{x}}{\sigma}$

$\mu = 0, \sigma = 1$

This data will satisfy

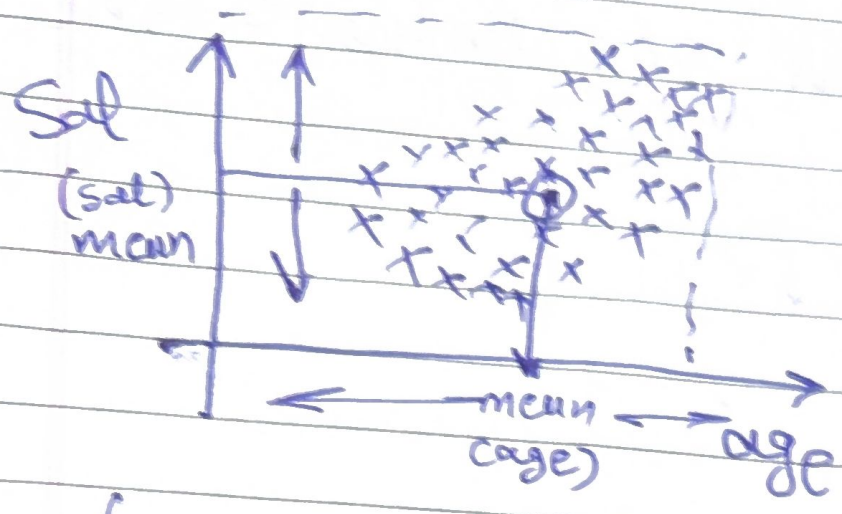
Standardization: $= \frac{x_i - \bar{x}}{\sigma}$ $\mu = 0, \sigma = 1$

$= \frac{27 - 32}{10} =$ ☁

for example

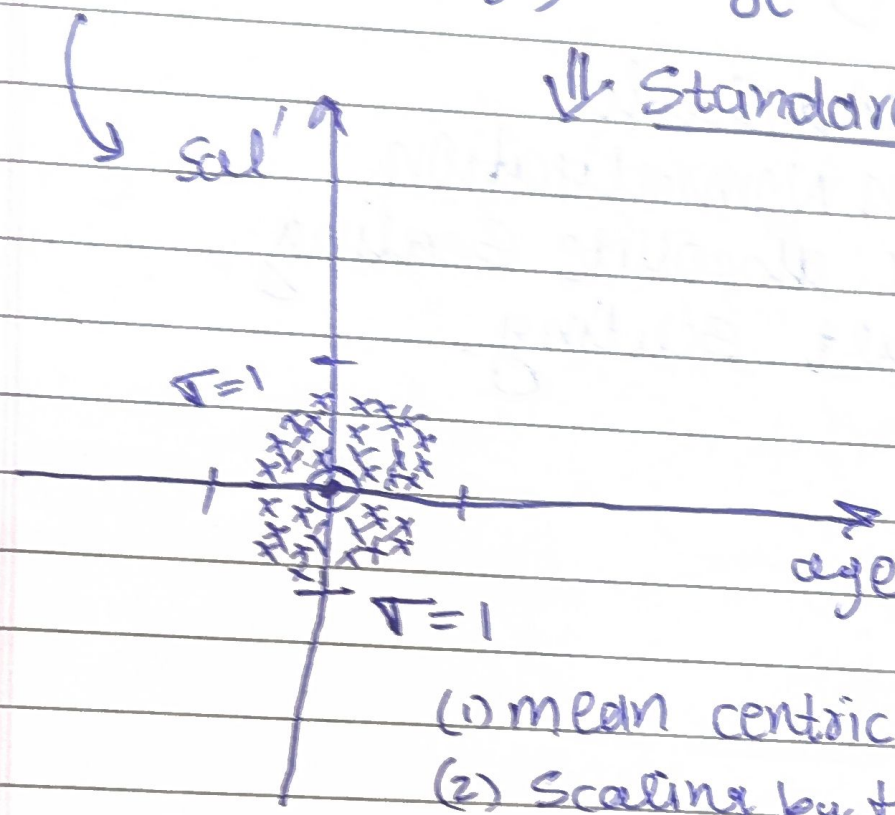
↳ update

Geometric Intuition of Standardization :



⊙ ⇒ mean

⇓ Standardization



⊙ Mean = 0

(1) mean centric
(2) Scaling by the factor of σ.

Normalisation :

=> we have to remove or eliminate the unit factors.

Eg. N, ----

(-1 to +1) or (0 to 1)

Techniques: MinMaxScaler
Mean Normalisation
Max absolute Scaling.
Robust Scaling.

1] MinMaxScaler :

Weight

130
67
78
50
48
30
80
⋮

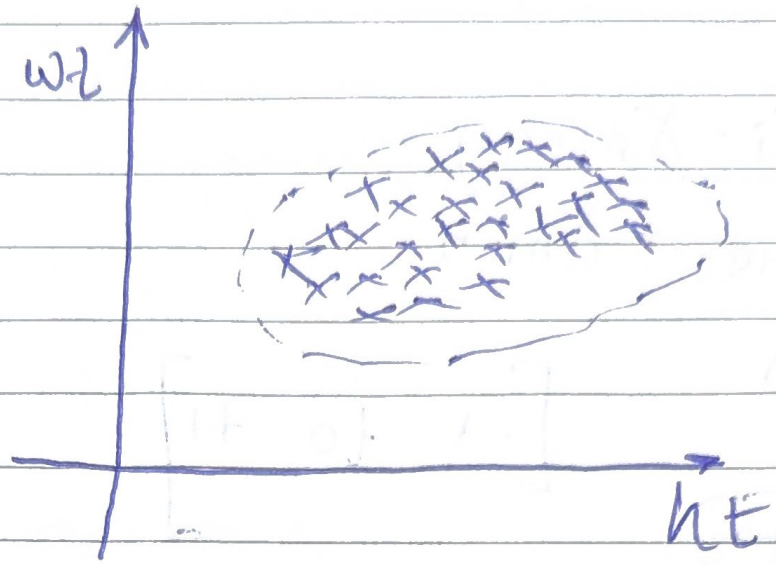
Normalising → (0 - 1)

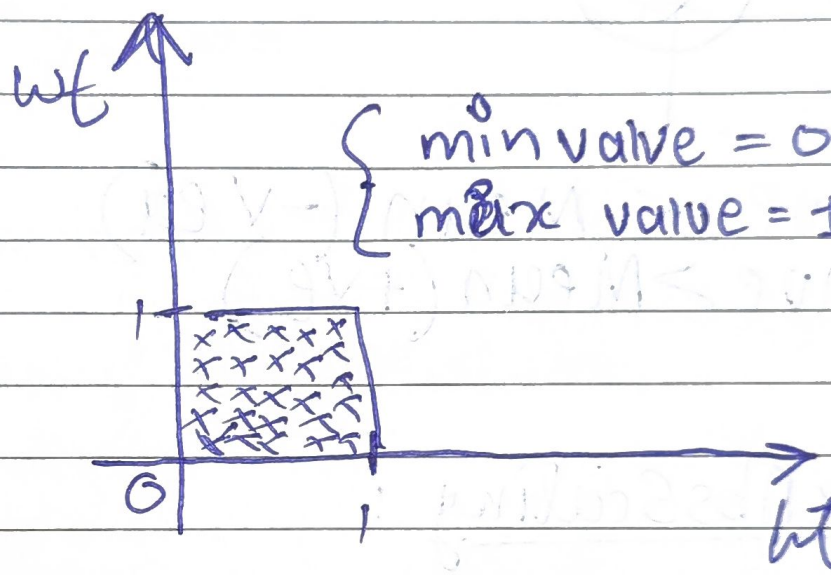
$$X_i' = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

$$(130)\ X_i' = \frac{130 - 30}{130 - 30} = 1$$

$$(50)\ X_i' = \frac{50 - 30}{130 - 30} = 0.2$$

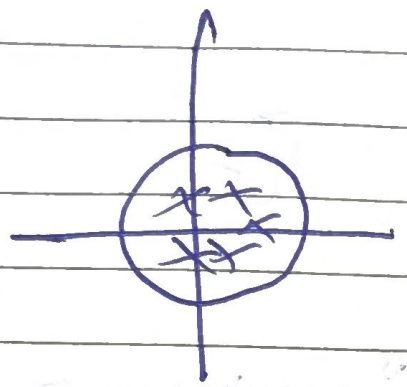
Geometrical Intuition :





2] Mean Normalisation :

$$X_i' = \frac{X_i - X_{mean}}{X_{max} - X_{min}}$$



[-1 to +1]

Value < Mean (-ve)
Value > Mean (+ve)

3] MaxAbsScaling :

$$X_i' = \frac{X_i}{|X_{max}|}$$

↳ MaxAbsScaler
↳ Sklearn

Use : Sparse data.
↳ zyadah zeros
jis data me ho.

4) Robust Scaling

$$X_i' = \frac{X_i - X_{median}}{IQR}$$

↳ Robust Scaler
↳ Sklearn

- Robust to outlier.
  ↳ for lot of outliers data.

Normalisation V/s Standardization.

Some questions to ask

↳ 1) Scaling required ?
2) ~~Must~~
Maximum time standardization
3) Normalisation
↳ CNN → (0-255) ✓