Feature Extraction

Defn

Feature Extraction is a process of selecting and extracting the most important features from raw data

Dataset

ML Application → 1000 features

⇓

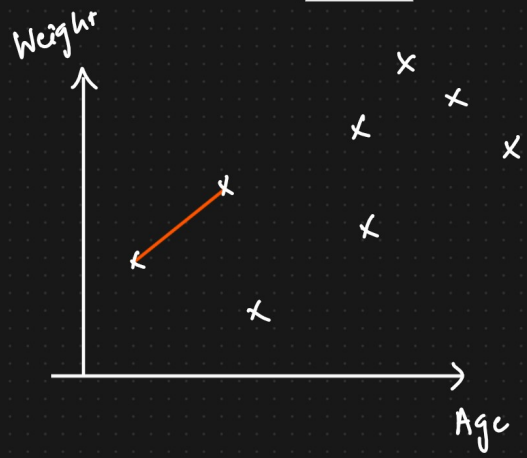
Most Important features

⇓

TRAIN ML Model

① Feature Scaling

Age (years)	Weight (kg)	Height (cms)	BMI
32	70	140	—
28	75	160cm	—
35	80	155cm	—



① Standardization [Machine Learning].

$$Z\text{-score} = \frac{x_i - \bar{x}}{\sigma}$$

$$\left\{ \frac{32-\bar{x}}{\sigma}, \frac{28-\bar{x}}{\sigma}, \frac{35-\bar{x}}{\sigma} \right\}$$

⇓

SND ⇒ $\boxed{\mu=0, \sigma=1}$

(ii) Normalization [Min Max Scaler] [0-1] → Deep Learning

Age

→ 24
→ 25
→ 32
→ 45
50
52

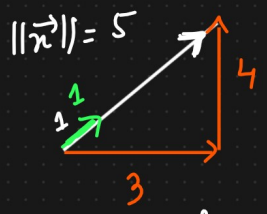
⇒ TRANSFORMATION

$$x_{scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Age " {Min Max Scaler}
↳ [0-1].

(3) Unit Vector = Magnitude of 1 [Feature Scaling]

$\vec{x} = (3,4)$



$\|\vec{x}\| = \sqrt{3^2 + 4^2}$ {pythagoras theorem}.

$= \sqrt{9+16} = \sqrt{25} = 5$ //

$$\hat{u} = \left(\frac{3}{\|\vec{x}\|}, \frac{4}{\|\vec{x}\|}\right) = \left(\frac{3}{5}, \frac{4}{5}\right).$$

$$\|\hat{u}\| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{a^2+b^2}$$

$$= \sqrt{\frac{9}{25} + \frac{16}{25}} = \sqrt{\frac{25}{25}} = \boxed{1}$$ //

② Feature Selection : We just pick the most Important features.

500 features ⟶ Top 10 features

⇓

ML Model Train

① Filter Method : Eg : Correlation

② Embedded Method

③ Wrapper Method

③ PCA (Principal Component Analysis)

f1	f2	O/P

2D ⟶ 1D

1000 D ⟶ 1D {There will be some loss of data}

2D plane

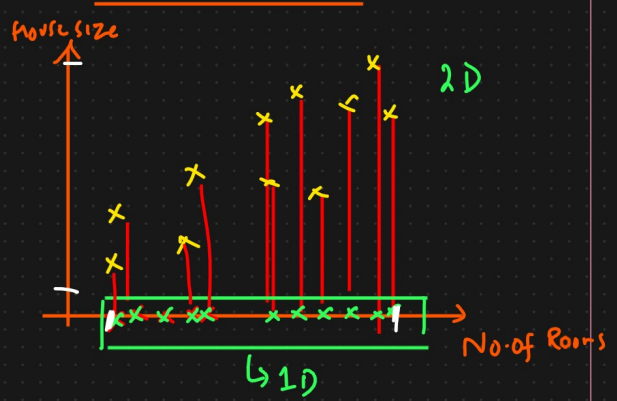
Dataset

I/P feature

No. of Rooms | House Size

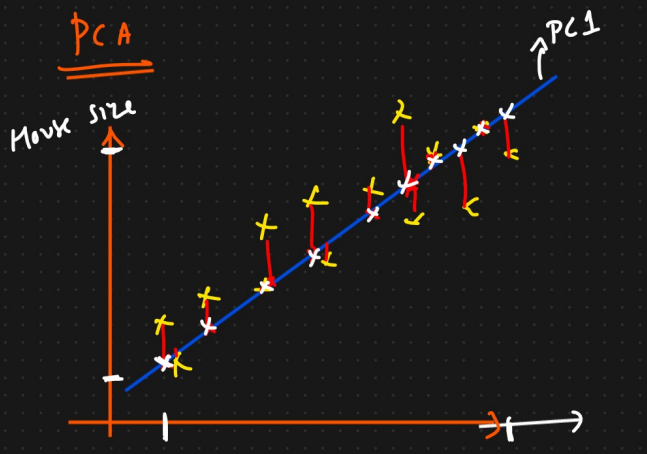
O/P

Price



Loss of Information ↑↑

PCA



2D → 1D {Some loss of Information}

3D → 2D

Higher Dimension → Lower Dimension