

Feature Engineering

Feature Extraction

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

ML Application \rightarrow 1000 features



Most Important features

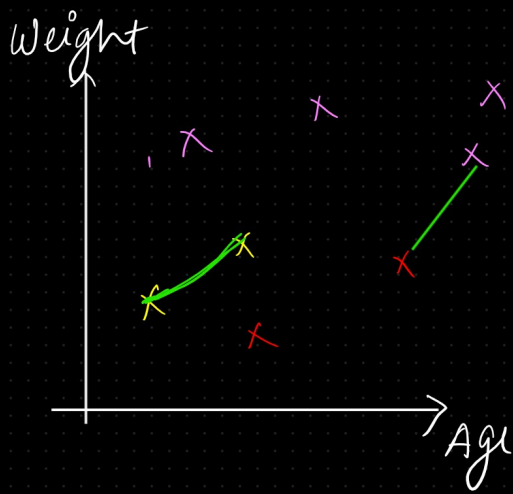


Machine Learning Algorithm

① Feature Scaling

② Feature Selection

<u>Feature Scaling</u>			
Age	Weight (kg)	Height (cm)	BMI
32	70	140 cm	
40	80	160 cm	
35	90	170 cm	



Machine Learning
Model

- ① Standardization
- ② Normalization
- ③ unit vector

Standardization → Machine Learning

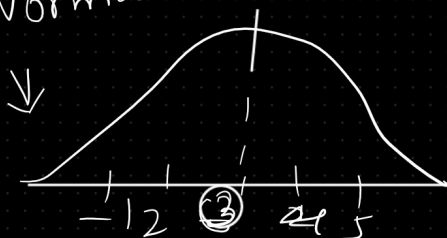
data points

$$Z = \frac{x - \bar{x}}{\sigma}$$

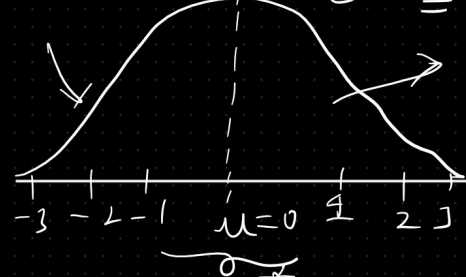
(Annotations: Z → Z-score, x → data points, \bar{x} → sample mean, σ → std. deviation)

$$\mu = 0, \quad \sigma = 1$$

Normal



standardization =



deep learning [Images]

② Normalization [Min Max scaler]

[0 to 1]
[0, 1] → Range

Age

24

25

26

27

28

Transformation



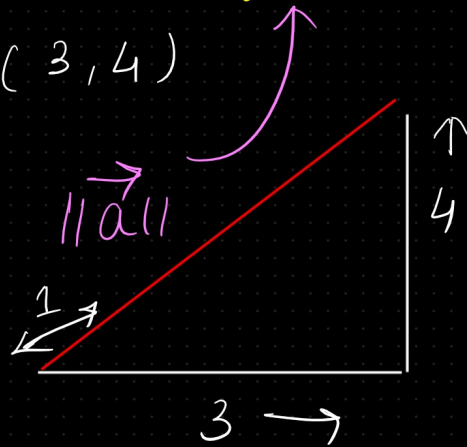
Age' [0, 1]

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

③ Unit Vector $\rightarrow \hat{u} = \text{magnitude} = \frac{1}{\|\hat{u}\|}$

Vector \swarrow direction
 \searrow magnitude

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



$$\begin{aligned} \|\vec{a}\| &= \sqrt{3^2 + 4^2} \\ &= \sqrt{9 + 16} \\ &= \sqrt{25} = 5 \end{aligned}$$

$$\|\vec{a}\| = \textcircled{5}$$

unit vector

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

$$\|\hat{u}\| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{\frac{9+16}{25}}$$

$$\|\hat{u}\| = \sqrt{\frac{\cancel{25}}{\cancel{25}}} = \sqrt{1} = 1$$