PW SKILLS

Covariance And Correlation

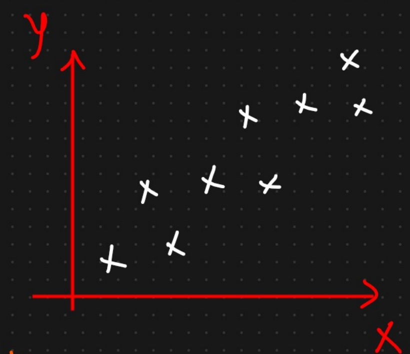
[Relationship between X and Y]

X	Y
2	3
4	5
6	7
8	9

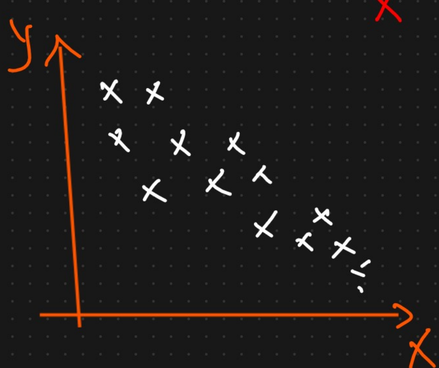
$\to$ X↑ Y↑
X↓ Y↑
$\to$ X↓ Y↓
X↑ Y↓

↓ o/p

↓↑ Size of House $\to$ Price of House ↑↓



X↑ Y↑
X↓ Y↓



X↑ Y↓
X↓ Y↑

Covariance

$$Cov(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

$x_i \to$ Datapoints of x
$\bar{x} \to$ Sample mean of x
$y_i \to$ Datapoints of y
$\bar{y} \to$ Sample mean of y

$$Var(x) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})}{n-1}$$

$= Cov(x,x) \Rightarrow$ Spread

$Cov(x,y)$

$\boxed{\begin{matrix} X\uparrow & Y\uparrow \\ X\downarrow & Y\downarrow \end{matrix}}$ +ve Covariance

$\boxed{\begin{matrix} X\uparrow & Y\downarrow \\ X\downarrow & Y\uparrow \end{matrix}}$ -ve Covariance

X	Y
→ 2	3
→ 4	5
→ 6	7
$\bar{x} = 4$	$\bar{y} = 5$

$$Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$= \frac{\left[(2-4)(3-4) + (4-4)(5-5) + (6-4)(7-5)\right]}{n-1}$$

$$= \frac{4+0+4}{2} = \frac{8}{2} = 4 \text{ +ve value}$$

⇓

Positive Covariance

$$\begin{tikzcd} X \arrow[r, bend left, "+100"] & Y \arrow[l] \arrow[r, bend left, "+300"] & Z \arrow[l] \end{tikzcd}$$

$$\begin{tikzcd} A \arrow[r, bend right, "-100"'] & B \arrow[l] \arrow[r, bend left, "-500"] & C \arrow[l] \end{tikzcd}$$

X & Y are having a positive Covariance

Advantages

① Relationship between X and Y

+ve or -ve value

Disadvantages

① Covariance does not have a specific limit value

② Pearson Correlation Coefficient [-1 to 1]

$$\rho_{x,y} = \boxed{\frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}}$$

① The more the value towards +1 the more +ve Correlated it is (x,y)

② The more the value towards -1 the more -ve Correlated it is (x,y)

③ Spearman Rank Correlation [-1 to 1]

$$r_s = \boxed{\frac{Cov(R(x), R(y))}{\sigma(R_{(x)}) * \sigma(R_{(y)})}}$$

x	y	R(x)	R(y)
1	2	5	5
3	4	4	4
5	6	3	3
7	8	2	1
0	7	6	2
8	1	1	6

Feature Selection

+ve	+ve	+ve	≈0	−ve	
Size of house ↑	No. of Rooms ↑	Location ↑	~~No. of people Staying~~	Haunted	Price ↑