Covariance And Correlation

[Relationship between X and Y]
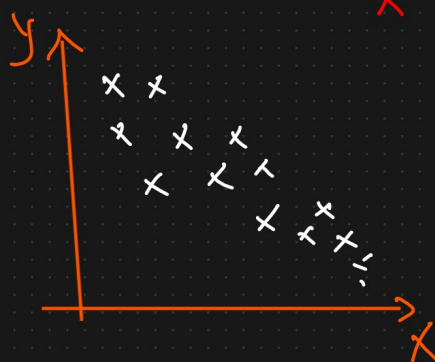
X	Y
2	3
4	5
6	7
8	9

$\rightarrow$

$$X\uparrow \quad Y\uparrow$$
$$X\downarrow \quad Y\uparrow$$
$$X\downarrow \quad Y\downarrow$$
$$X\uparrow \quad Y\downarrow$$

$\downarrow$ O/P

$\downarrow\uparrow$ Size of House $\longrightarrow$ Price of $\uparrow\downarrow$ House
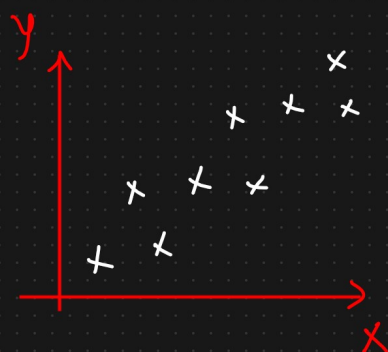
$$X\uparrow \quad Y\uparrow$$
$$X\downarrow \quad Y\downarrow$$

$$X\uparrow \quad Y\downarrow$$
$$X\downarrow \quad Y\uparrow$$

Covariance

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

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

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

$$= Cov(X,x) \implies Spread$$

$x_i \rightarrow$ Datapoints of x

$\bar{x} \rightarrow$ Sample mean of x

$y_i \rightarrow$ Datapoints of y

$\bar{y} \rightarrow$ Sample mean of y

Cov(x,y)

| X↑ Y↑ |
| X↓ Y↓ |

+ve Covariance

| X↑ Y↓ |
| X↓ Y↑ |

−ve Covariance

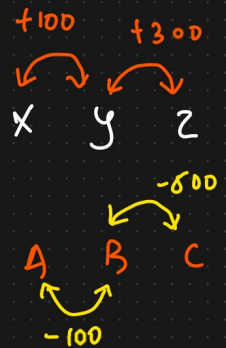| X	Y
→ 2	3
→ 4	5
→ 6	7

$\bar{x}=4$ $\bar{y}=5$

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

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

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

$$\Downarrow$$

Positive
Covariance

+100   +300
X   Y   Z

−500
A   B   C
−100

X & Y are having a positive Covariance

Advantages

① Relationship between X and y
tve or −ve value

Disadvantages

① Covariance does not have a specific limit value

② Pearson Correlation Coefficient 🗨 $[-1 \text{ to } 1]$

$$\int_{x,y} = \frac{Cov(x,y)}{\nabla x \cdot \nabla y}$$ 🗨

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 $[-1 \text{ to } 1]$ 🗨

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

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
Size of
House ↑ 💬

+ve
No. of
Rooms ↑ 💬

+ve
Location ↑ 💬

≈ 0
No. of people
Staying ✗ 💬

−ve
Haunted 💬

Price ↑