Pearson Correlation Coefficient:

This is used for measure the relationship within particular range from (-1 to 1).

→ The more towards +1 then say more positively correlation.

→ The more towards -1 then say more negatively correlation.

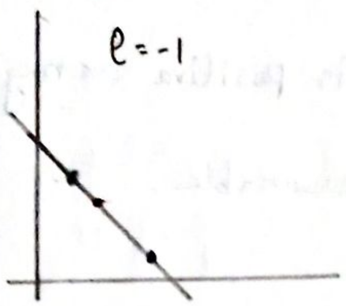
Formula,

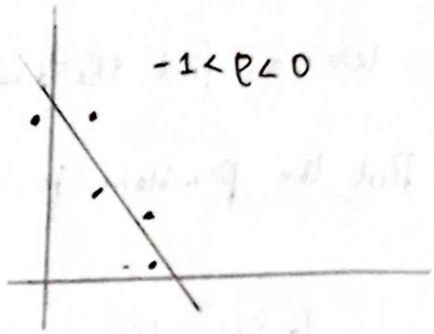
$\sigma_x \rightarrow$ sd of feature x

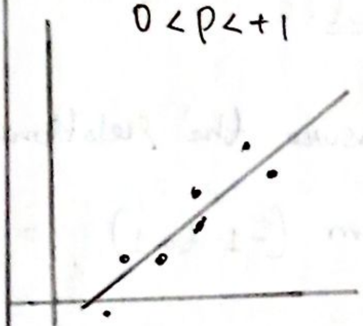
$$\rho(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

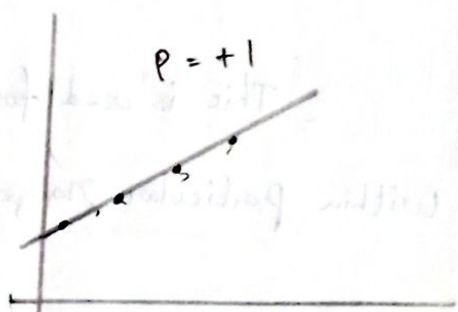
$\sigma_y \rightarrow$ sd of feature y

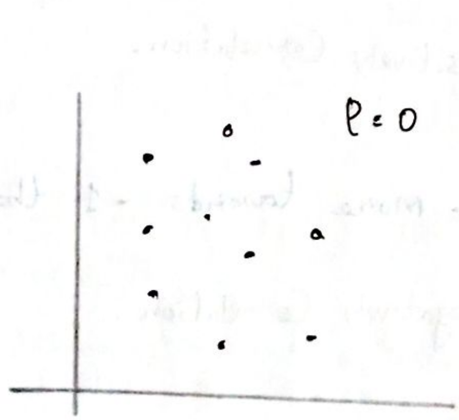
Value = $\{-1 \text{ to } 1\}$ (Ans)









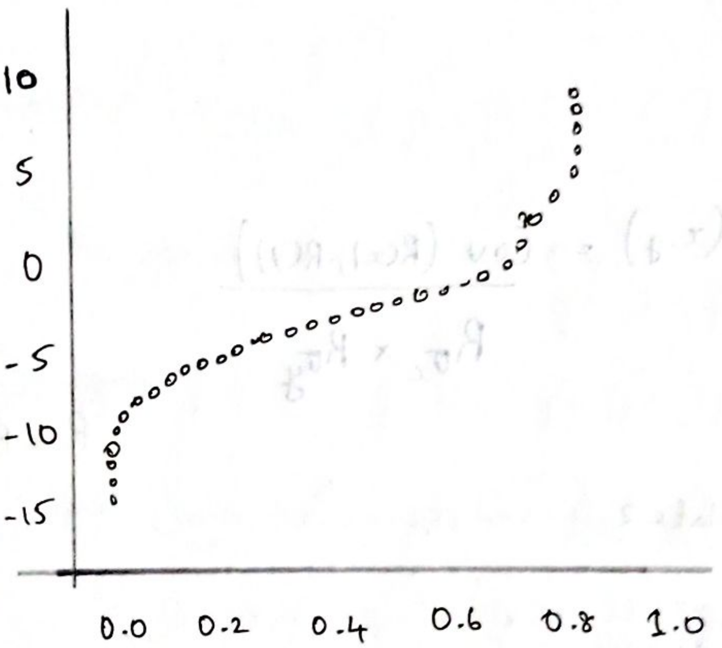


Does not follow straight line we say range from $-1 < \rho < 0$ (or) $0 < \rho < +1$ [Range]

Disadvantage:

→ The Pearson Correlation Coefficient is only works well for Linear properties. Does not fit for non-Linear properties. So, we jump into Spearman rank Correlation.

Spearman's rank Correlation Coefficient:



In this Scenario,

Spearman Correlation = 1

Pearson Correlation = 0.88

Spearman Considered best because, it is non-linear Property. So works well and also show accurate Correlation.

Advantage:

- It is somemore similar to Pearson Correlation, The difference is it works well for non-linear Property as well.

- It also Calculate the relation accurately. (i.e. One point overlab with some distance also Calculate. But, Pearson Cannot).

formula,

$$\text{Spear}(x, y) = \frac{\text{Cov}(R(x), R(y))}{R_{\sigma_x} \times R_{\sigma_y}}$$

$R \rightarrow$ Rank

How to calculate?

Let Consider,

x	y	R(x)	R(y)
170	75	2	2
160	62	3	3
150	60	4	4
145	55	5	5
180	85	1	1

$x \rightarrow$ Height

$y \rightarrow$ weight

Based on the higher value, Rank will be assigned.

The formula only works with $R(x)$, $R(y)$, $R\sigma_x$ and $R\sigma_y$. It neglect the Original feature X and Y while Performing.

The main thing of the Spearman rank Correlation is that Capture non-linear properties as well.