* Covariance

- Covariance measure the direction of relationship between two variables (features)
- A positive covariance means both variables tend to be high or low at the same time (directly proportional)
- A negative covariance means that when one variable is high, the other one tends to low. (inversly proportional.
- A '0' covariance means no relation.
- It is used to measure the distribution of the data points.

→ Variance is also used to measure the distribution of data points, but just for one variable, while covariance is used for multiple variables.

As variance,

$$\sigma^2 = \frac{\Sigma (x_i-\bar{x})^2}{n-1}$$

$$\Rightarrow \sigma^2 = \frac{\Sigma (x_i-\bar{x})(x_i-\bar{x})}{n-1}$$

$$\Rightarrow Cov(x,x) = \frac{\Sigma (x_i-\bar{x})(x_i-\bar{x})}{n-1}$$

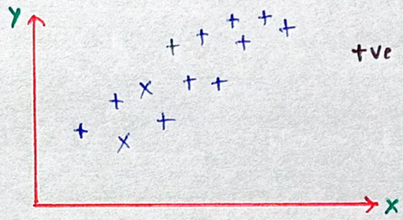
$$\therefore \boxed{Cov(x,x) = Var(x)}$$

And formula for covariance is,

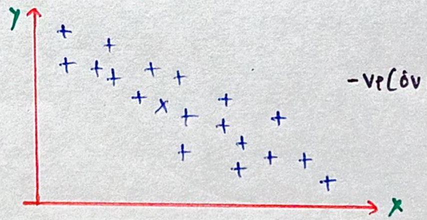
$$\boxed{Cov(x,y) = \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})}{n-1}}$$

→ Variance measures the spread of data along single axis, while covariance measures directional relationship.

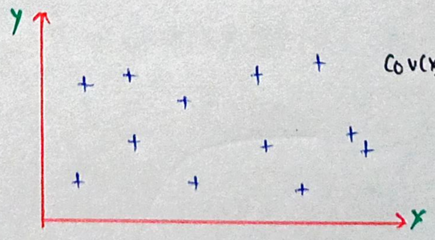
- Few graphical representation of Cov(x,y),



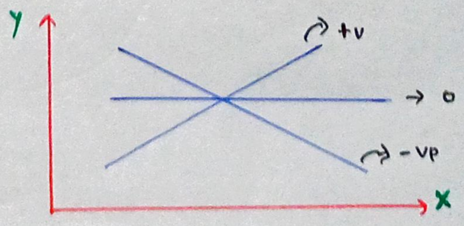
+ve Cov(x,y)



-ve Cov(x,y)



Cov(x,y) = 0



ex- we have two features with certain values.

X	Y
12	40
13	45
15	48
17	60
18	62

$\bar{x} = 15$ $\bar{y} = 51$ (means)

then,

$$Cov(x,y) = \Big[(12-15)(40-51) + (13-15)(45-51) + (15-15)(48-51) + (17-15)(60-51) + (18-15)(62-51)\Big] / 5-1$$

$$= \Big[(-3\times-11) + (-2\times-6) + (0\times-3) + (2\times 11) + (3\times 11)\Big]/4$$

$$= [33+12+0+22+33]/4$$

$$= 25$$

$\therefore$ $Cov(x,y) = 25$, is +ve so their relationship is directly proportional.

→ It is used in feature selection

→ One drawback of covariance is that, it only tells direction of relation and not the strength.
So we cant compare features relation with each other and say it has stronger relation compared to another one as Cov has no scale and its value can go as high and low.