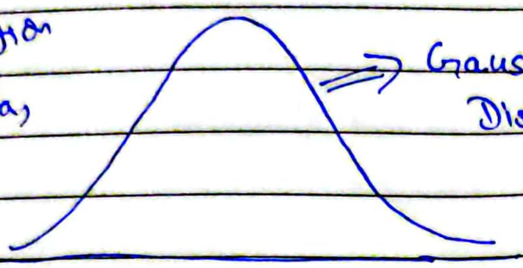
Day - 4

* Central Limit Theorem (CLT)

Population Data, $\boxed{N}$



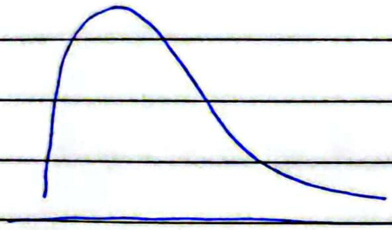
$\Rightarrow$ Gaussian/Normal Distribution $\rightarrow$ Sample Data (n)

$\boxed{n \geq 30}$

$[x_1, x_2 - - - - x_n] \rightarrow S1$

$\downarrow$

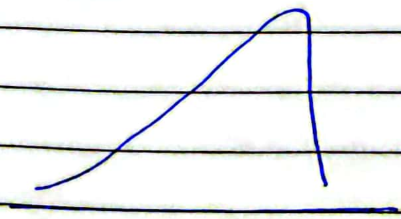
$\bar{x}_1 \rightarrow$ sample 1 mean



$[x_1, x_2, - - - x_n] \rightarrow \bar{x}_2$

$[x_1, x_2, - - - x_n] \rightarrow \bar{x}_3$

$\vdots$



$[x_1, x_2, - - - - x_n]\ \bar{x}_m$

Size of Sample $\leftarrow \boxed{n}$ $\quad\boxed{m} \rightarrow$ num of samples ; $\boxed{n \geq 30}$

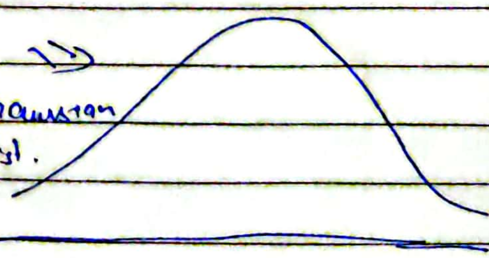
$S_1 \rightarrow \{x_1, x_2, x_3 \_\_\_\_ x_n\} \rightarrow \bar{x}_1$

$S_2 \rightarrow \{x_1, x_2, x_3 - - - x_n\} \rightarrow \bar{x}_2$

$\vdots$

$S_m \rightarrow \{x_1, x_2, x_3 \_\_\_ x_n\} \rightarrow \bar{x}_m$

$\Rightarrow$ Gaussian Dist.



$\Rightarrow$ Larger the num of samples better the result

→ The CLT states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger - no matter what the ~~sample of~~ shape of the population distribution.

S_1, S_2, S_3 ——— S_m → Sample of size m

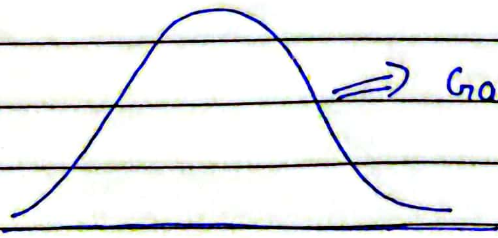
↓ ↓ ↓ ↓

$\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ ——— $\bar{x}_m$ → Sample means

↓

Distribution of these sample means will be normal or gaussian

⇓



⇒ Gaussian Dist.

Q Find size of shark throughout the world?

Sol. Take samples from different regions, $n \geq 30$

↓

Plot sample means distribution

↓

It will follow Gaussian Distribution so we can make assumptions.

Date
Page

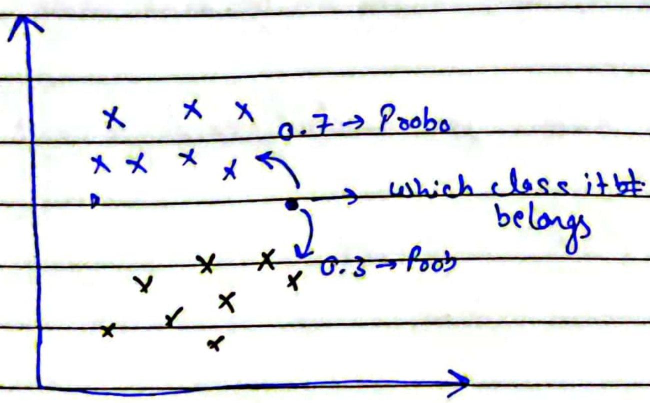
* **Probability** :- It is a measure of the likelihood of an event

Eg :- Tossing a fair coin $P(H) = 0.5$
$P(T) = 0.5$

⇓

Unfair coin → $P(H) = 1$

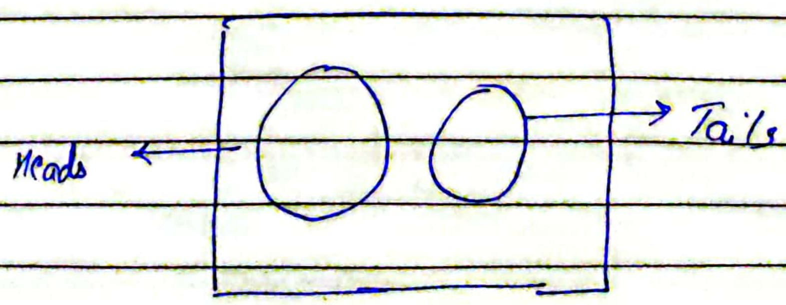
Rolling a Dice $P(1) = \frac{1}{6}, P(2) = \frac{1}{6} --- P(6) = \frac{1}{6}$



→ used in ML

① **Mutual Exclusive Events** :- Two events are mutually exclusive if they cannot occur at same time.

Eg :- Tossing a Coin :-
Rolling a Dice



② **Non - Mutual Exclusive Events :-** Two events can occur at the same time.

Eg :- ⓐ Picking randomly a card from a deck of cards, two events like ~~'A heart of king'~~, 'heart' and 'king' can be selected

ⓑ Bag of marbels

→ **Mutual Exclusive Problems**

① What is the Probability of coins lending on heads or tails

⇓

Addition Rule for mutual exclusive events

$$P(A \text{ or } B) = P(A) + P(B)$$
$$= 0.5 + 0.5 = \underline{1}$$

② What is the probability of getting 1 or 6 or 3 while rolling a dice

$$P(1 \text{ or } 3 \text{ or } 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} = 0.5$$

→ Non-Mutual Exclusive Events Problems :-

① Bag of Marbels : 10 Red, 6 Green, 3 (R&G)

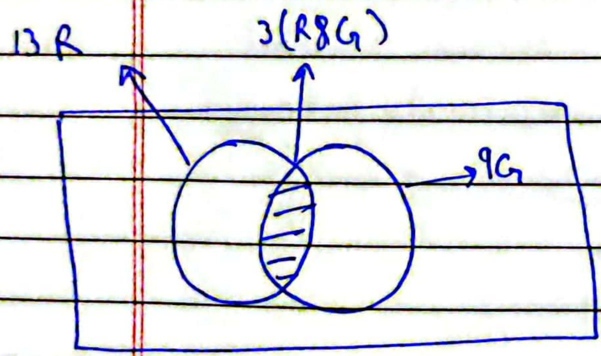
When picking randomly from a bag of marbels what is the probability of choosing a marbel that is red or green?

⇓

Addition Rule for non-mutually exclusive

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$\boxed{\cancel{= \frac{10}{19} + \frac{6}{19} - \frac{3}{19} = \frac{13}{19}}}$$



$$= \frac{13}{19} + \frac{9}{19} - \frac{3}{19}$$

$$= 1$$

② Deck of Cards :- What is the prob of choosing heart ~~and~~ or Queen at the same time

$$P = P(\text{Heart}) + P(\text{Queen}) - P(\text{Heart \& Queen})$$

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$= \frac{16}{52}$$

→ Multiplication Rule

(a) Dependent Events :- Two events are dependent if they affect one another

Bag of marbels → 4 White
3 Yellow

$P(W) \to \frac{4}{7} \longrightarrow P(Y) \to \frac{3}{6}$

1 white marbel taken out

6 because 1 marbel was already taken out.

Conditional Probability

Q what is the Prob. of rolling a "5" and then a "3" with a normal six sided dice?

Sol :- Multiplication rule for independent events

$$P(A \text{ and } B) = P(A) * P(B)$$

$$= \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$$

CamScanner

→ Bag of marbels → 4 orange
3 yellow

what is Probability of drawing "orange" and then drawing a "yellow" marbel from the bag?

⇓

Dependent Events.

Sol :- $P(\text{Orange}) = \frac{4}{7}$ → now 1 orange marbel is taken out

$P(\text{yellow}/O) = \frac{3}{6} = \frac{1}{2}$

↓

Prob. of yellow given orange has happened

$\boxed{P(Y/O)}$ → <u>Conditional Probability</u>

$P(\text{Orange and Yellow}) = P(O) \times P(Y/O)$

$= \frac{4}{7} \times \frac{1}{2} = \frac{4}{14}$

$= \frac{2}{7}$

CamScanner

* Permutations

School of Children Taken to Chocolate Factory

↓

[Dairy Milk, kitkat, Milky Bar, Sneakers, 5 star]

↓

Write 3 names of chocolates you see first

↓

Here 5 options available × 4 options × 3 options

= 60 ways ⇒ Permutations

⇒ $\boxed{\text{with permutation, order matters}}$

[DM, MB, Sneaker]

[5 star, kitkat, MB]

⋮

60 possible arrangements.

^nP_r n = Total num of objects

r = num of selections (here 3 because need to write 3 names)

$$^nP_r = \frac{n!}{(n-r)!} = \frac{5!}{(5-3)!} = 60$$

* Combination

Here, Repeatation will not occur

↓

Only <u>unique</u> combination is possible

$$^{n}C_{r} = \frac{n!}{r!\,(n-r)!} = \frac{5!}{3!\,(2)!}$$

$$= \frac{5\times4\times\cancel{3\times2\times1}}{\cancel{3!}\times 2} = 10$$

---

* Covariance

{feature selection}

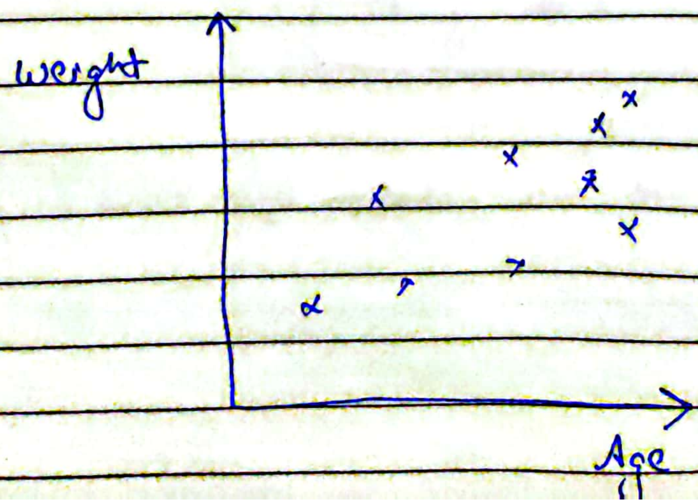
Age	Weight
12	40
13	45
15	48
17	60
18	62

Age ↑ Weight ↑

Age ↓ Weight ↓

⇓

Quantify the relation b/w x & y using mathematical equation



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

$$\text{Variance } (\sigma^2(x, x)) = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

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

$$\Downarrow$$

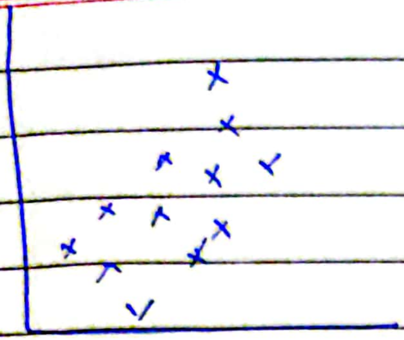
$\boxed{Cov(x, x)}$ → what is the relation of x with X

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

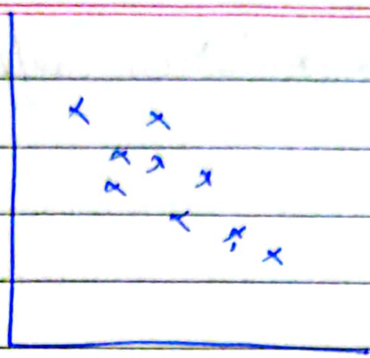
+ve Cov	x ↑	y ↑
	x ↓	y ↓

−ve Cov	x ↑	y ↓
	x ↓	y ↑

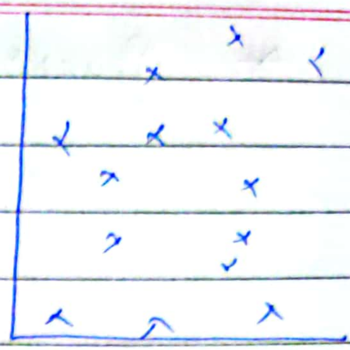
Cov = 0 → no relation b/w X and Y



+ve Cov



−ve Cov



Cov = 0

x	y
10	4
8	6
7	8
6	10

$\bar{x} \to 7.75 \qquad \bar{y} \to 7$

$\text{Cov}(x,y) \to$

$$\text{Cov}(x,y) \to \left[(10-7.75)(4-7) + (8-7.75)(6-7) + (7-7.75)(8-7) + (6-7.75)(10-7) \right]$$

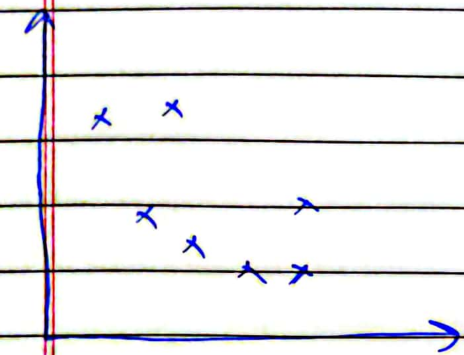
$$= -3.25$$

* Pearson Correlation Coefficient

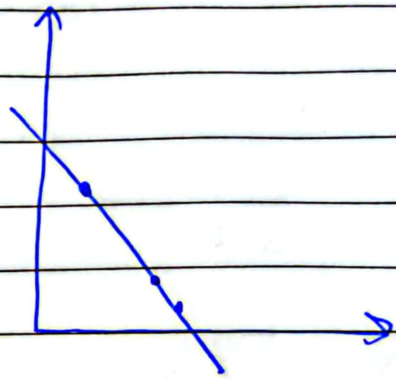
$$\rho(x,y) = \frac{Cov(x,y)}{\sigma_x , \sigma_y}$$

Here we are trying to restrict values b/w $\boxed{-1 \text{ to } 1}$

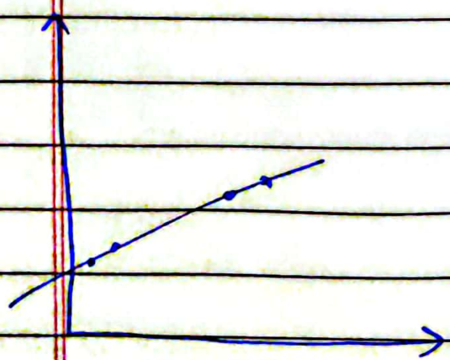
→ More the value towards +1, more +ve correlated
→ More the value towards −1, more −ve correlated.



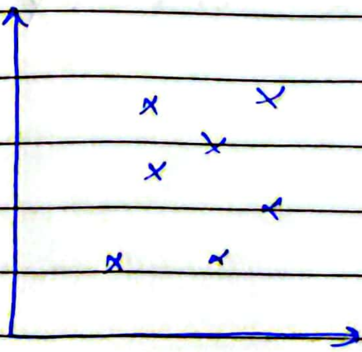
$\boxed{-1 \text{ to } 0}$



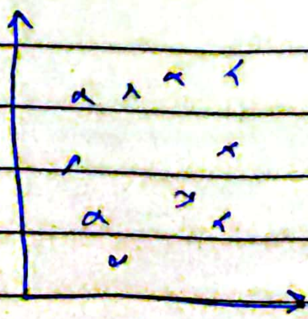
$\rho = -1$



$\rho = +1$



$\rho = 0 \text{ to } 1$



$\rho = 0$

* Spearman Rank Correlation → Used for non-linear data.

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

X	Y	R(X)	R(Y)
10	4	4	1
8	6	3	2
7	8	2	3
6	10	1	4

→ Why this correlation will be used?

x y z → Output → Let output have very ~~high~~ high +ve or -ve ~~covariance~~ correlation

y → output → correlation = 0

In this case we can remove y

now, let say x & z are 95% correlated
↓
This means x and z are approximately same, so we can drop any one of them.