Mutual Information

* Mutual information is a measure of the amount of information that one randomvariable contains about another random variable.
* It is a quantity that how much one randomvariable tells us about the another.
* ie; It is calculated between two random variable.

case 1: X represents the roll of a fair 6 sided die and Y represents whether the roll is even. (0 if even 1 if odd). From this we can say that Y tells us something about the value of X.

Case 2: X represents the roll of one fair die
Z represents the roll of another fair die
Then X and Z share no ~~mutual~~ information
the mutual information b/w a random variable is 0 if and only if two random variable is statistically independant.

example: set A $\{1, 2, 3, 5, 7, 9\}$

set B $\{3, 4, 8, 9\}$

$A \cap B = \{3, 9\}$ This is the mutual information b/w the two sets.

Mathematical Representation:

We are interested in the transfer of information from a transmitter to reciever over a channel.

x_i —— $\boxed{\text{channel}}$ x_j is transmitted over the channel

Before recieving the message at the reciever side the only information we are having is $I(x_j)$

$I(x_j) = -\log_2 P(x_j)$ ⟶ Initial uncertainty or Initial Information.

x_j —— $\boxed{\text{channel}}$ —— y_k

when x_j is transmitted y_k is recieved.

then the information we have is

$I(x_i/y_k) = -\log P(x_i/y_k)$ ⟶ Final uncertainty or final information

Mutual information is the amount of information transfer when x_j is transmitted and y_k is recieved.

$I(x_j ; y_k)$ = Initial uncertainty − final uncertainty

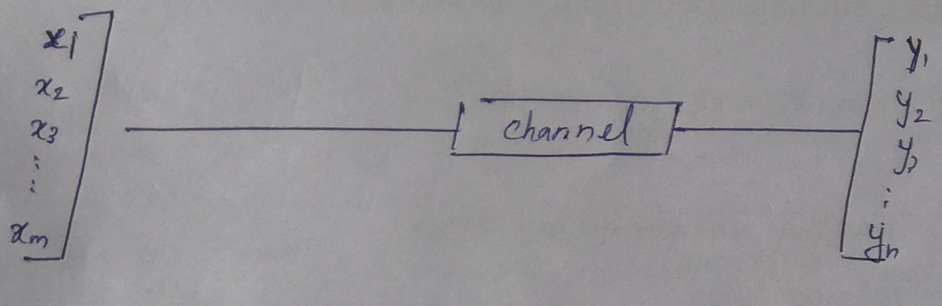
$= -\log_2 P(x_j) - -\log_2 P(x_i/y_k)$

$= -\log_2 P(x_j) + \log_2 P(x_i/y_k)$

↳ this can be written in the form of $\log A - \log B$

$I(x_k ; y_k) = \log \dfrac{P(x_j/y_k)}{P(x_j)}$ this is known as the mutual information b/w x and y

Now Consider we are transmitting a set of message through the channel and at the o/p will get another set



then the mutual information b/w the i/p and o/p is given by (i.e; Mutual information b/w R. Vbls X and Y defined as average of $I(x;y)$)

$$I(X;Y) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j, y_k) \log_2 \frac{(P x_j / y_k)}{P(x_j)}$$

$$\boxed{I(X;Y) = \sum \sum P(x, y)\, I(x,y)}$$

* when the base of the logarithm is 2 the unit of $I(X;Y)$ are bits.
* when the base is e the unit of $I(X;Y)$ are called nats (natural units). $\log_e \rightarrow \ln$.

Properties of Mutual information

1. The mutual information of the channel is symmetric.

ie; $I(X;Y) = I(Y;X)$

$$I(X;Y) = \sum_{j=1}^{m} \sum_{k=}^{n} P(x_j, y_k) \log_2 \left[\frac{P(x_j/y_k)}{P(x_j)} \right] \quad \text{—— (1)}$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j; y_k) \log \left[\frac{P(x_j, y_k)}{P(y_k) \cdot P(x_j)} \right]$$

$$\longrightarrow P\left(\frac{A}{B}\right) = \frac{P(AB)}{P(B)}$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j : y_k) \log \left[\frac{P(y_k/x_j)}{P(y_k)} \right] \quad \text{—— (2)}$$

By comparing the eqns (1) and (2) we can know that in the 2nd eqn instead of x; y is given and instead of y; x is given.

$$I(X:Y) = I(Y;X)$$

2. $I(X;Y) \geq 0 \Rightarrow$ the mutual information is always non negative

ie; we may have some information or no information

3. The mutual information of a channel may be represented in terms of Entropy.

ie; $I(X;Y) = H(X) - H(X/Y)$

Proof:-

$$I(X;Y) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j, y_k) \log_2 \frac{P(x_j/y_k)}{P(x_j)}$$

$\hookrightarrow \log A/B = \log A - \log B$

$$= \sum_{j=1}^{1} \sum_{k=1}^{n} P(x_j, y_k) \left[\log P(x_j/y_k) - \log P(x_j) \right]$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j, y_k) \log P(x_j/y_k) - \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j y_k) \log P(x_j)$$

$\hookrightarrow \sum_{k=1}^{n} P(y_k) \to 1$

$\hookrightarrow$ Sum of probabilities $= 1$

$$= -H(X/y) - \sum_{j=1}^{m} P(x_j) \log P(x_j)$$

$$= -H(X/y) + H(X)$$

$$I(X;Y) = H(X) - H(X/y)$$

thus Mutual information is represented in terms of Entropy.

4. $I(X;Y) = H(X) + H(Y) - H(X;Y)$

$I(X;Y) = H(X) - H(X/y)$ ① $\to$ from property 3.

$H(X,Y) = H(Y) + H(X/y)$ ② $\to$ from Joint Entropy

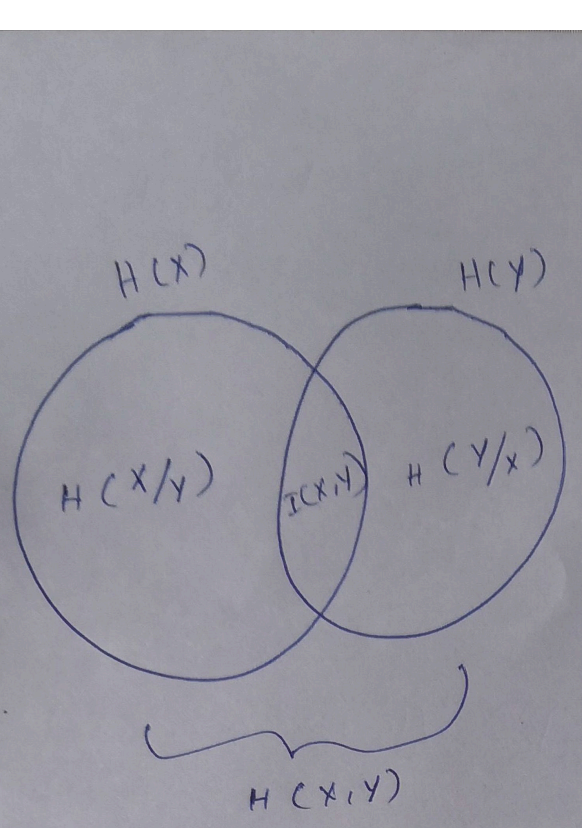
$H(X,Y) - H(Y) = H(X/y)$ — ③

substitute ③ in ①

$I(X;Y) = H(X) - [H(X,Y) - H(Y)]$

$= H(X) - H(X,Y) + H(Y)$

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$



$I(X,Y) = H(X) - H(X/Y)$

$I(X,Y) = H(Y) - H(Y/X)$

$I(X,Y) = H(X) + H(Y) - H(X,Y)$

Channel Capacity:

Maximum value of the mutual information is called channel capacity

Information in symbol

$$C_s = \max I(X,Y) \text{ bits/symbol}$$

Channel capacity can be defined as the max amount of information that can be transmitted by a channel per second

$$C = r\,C_s \text{ bits/sec}$$

r = no: of symbols/sec