WEEK 9

Channel Capacity of Discrete memoryless channel

If the channel is noiseless, then the reception of some symbol y_j uniquely determines the message transmitted. Because of noise, however, there is a certain amount of uncertainty regarding the transmitted symbol when y_j is received. If $P(x_i| y_j)$ represents the conditional probabilities that x_i was transmitted when y_j is received, then there is an uncertainty of $\log [1/P(x_i \backslash y_j)]$ about x_i when y_j is received. When this uncertainty is averaged over all x_i and y_j, we obtain $H(x/y)$, which is the average uncertainty about a transmitted symbol when a symbol is received.

Thus,

$$H(x|y) = \sum_i \sum_j P\left(x_i, y_j\right) \log \frac{1}{P\left(x_i|y_j\right)} \quad bits\ per\ symbol$$

If the channel were noiseless, the uncertainty would be zero. Obviously, this uncertainty, H(x|y), is caused by channel noise. Hence, it is the average loss of information about a transmitted symbol when a symbol is received.
$P(x_i|y_j)$ is a characteristic of the channel and the receiver. Thus, a given channel (with its receiver) is specified by the **channel matrix:**

$$\begin{array}{cc} & \text{Outputs} \\ & \begin{array}{cccc} y_1 & y_2 & \cdots & y_s \end{array} \\ \text{Inputs} \begin{array}{c} x_1 \\ x_2 \\ \cdots \\ x_r \end{array} & \begin{pmatrix} P(y_1|x_1) & P(y_2|x_1) & \cdots & P(y_s|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \cdots & P(y_s|x_2) \\ \cdots & \cdots & \cdots & \cdots \\ P(y_1|x_r) & P(y_2|x_r) & \cdots & P(y_s|x_r) \end{pmatrix} \end{array}$$

So we can get the reversed conditional probabaility using bayes rule as follow:

$$\begin{aligned} P(x_i|y_j) &= \frac{P(y_j|x_i)P(x_i)}{P(y_j)} \\ &= \frac{P(y_j|x_i)P(x_i)}{\sum_i P(x_i, y_j)} \\ &= \frac{P(y_j|x_i)P(x_j)}{\sum_i P(x_i)P(y_j|x_i)} \end{aligned}$$

Thus, if the input symbol probabilities $P(x_i)$ and the channel matrix are known, the reverse conditional probabilities can be computed from the above Eqs. The reverse conditional probability P(xi|yj) is the probability that x_i was transmitted when y_j is received.

If the channel were noise-free, the average amount of information received would be H(x) bits (entropy of the source) per received symbol. Note that $H(x)$ is the average information transmitted over the channel per symbol. Because of channel noise, we lose an average of H(x|y) bits of information per symbol. Therefore, in this transaction, the amount of information the receiver receives is, on the average, I(x; y) bits per received symbol, where:

$$I(x; y) = H(x) - H(x/y) \quad \text{bits per symbol}$$

I(x; y) is called the **Transinformation or the mutual information** of x and y. Because

$$H(x) = \sum_i P(x_i) \log \frac{1}{P(x_i)} \quad bits$$

we have

$$I(x;y) = \sum_i P(x_i) \log \frac{1}{P(x_i)} - \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i|y_j)}$$

and since

$$\sum_j P(x_i, y_j) = P(x_i)$$

It yields

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i)} - \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i|y_j)}$$

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)}$$

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}$$

Using bayes rule, the mutual information can be expressed as:

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(y_j|x_i)}{P(y_j)}$$

or

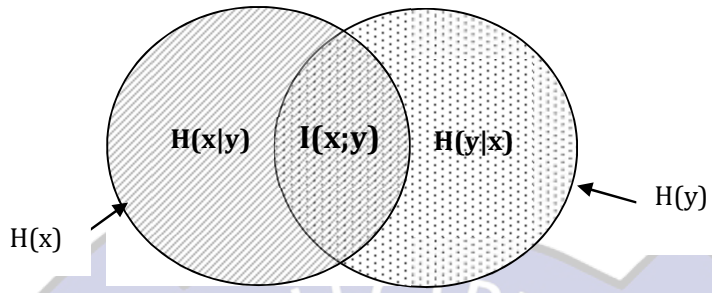
$$I(x;y) = \sum_i \sum_j P(x_i) P(y_j|x_i) \log \frac{P(y_j|x_i)}{\sum_i P(x_i) P(y_j|x_i)}$$

Note that :

$$I(\mathrm{x};\mathrm{y}) = I(\mathrm{y};\mathrm{x})$$
$$= H(\mathrm{y}) - H(\mathrm{y}|\mathrm{x})$$

The quantity H(y|x) is the equivocation of y with respect to x and is the average uncertainty about the received symbol when the transmitted symbol is known.

H(x) - $H(x|y)$ = $H(y)$ - $H(y|x)$



Venn diagram

$$I(X,Y) = H(Y) + \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j)\log_2 P(y_j \mid x_i)$$

But we have

$P(x_i, y_j) = P(x_i)P(y_j \mid x_i)$ *put in above equation yieldes*:

$$I(X,Y) = H(Y) + \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i)P(y_j \mid x_i)\log_2 P(y_j \mid x_i)$$

If the channel is symmetric the quantity:

$$\sum_{j=1}^{m} P(y_j \mid x_i)\log_2 P(y_j \mid x_i) = K$$

Where K is constant and independent of the row number *i* so that the equation becomes:

$$I(X,Y) = H(Y) + K\sum_{i=1}^{n} P(x_i)$$

Hence $I(X,Y) = H(Y) + K$ for symmetric channels

Max of $I(X,Y) = \max[H(Y) + K] = \max[H(Y)] + K$

When Y has equiprobable symbols then $\max[H(Y)] = log_2 m$

Then $I(X,Y) = log_2 m + K$, Or $C = log_2 m + K$

The mutual information of a channel is symmetric; that is:

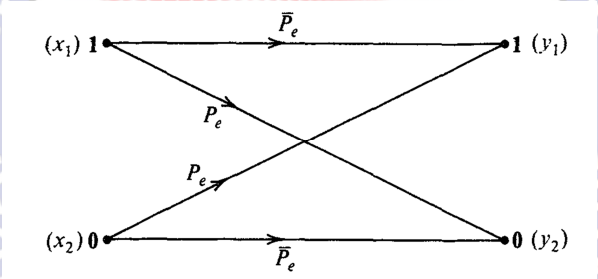
$$I(x ; y) = I(y ; x)$$

where the mutual information I(x ; y) is a measure of the uncertainty about the channel input that is resolved by *observing* the channel output, and the mutual information I(y ; x) is a measure of the uncertainty about the channel output that is resolved by *sending* the channel input.

I(x; y) is a function of the transmitted symbol probabilities $P(x_i)$ and the channel matrix. For a given channel, I(x; y) will be maximum for some set of probabilities. This maximum value is the **channel capacity** C_s,

$$C_s = \max_{P(xi)} I(x; y) \quad \text{bits per symbol}$$

Thus, C_s represents the maximum information that can be transmitted by one symbol over the channel. These ideas will become clear from the following binary symmetric channel (BSC):



Let $P(x_1) = \alpha$ and $P(x_2) = \bar{\alpha} = (1 - \alpha)$. Also,

$$P(y_1|x_2) = P(y_2|x_1) = P_e$$
$$P(y_1|x_1) = P(y_2|x_2) = \bar{P}_e = 1 - P_e$$

Substituting this in the following equations (Mutual Information) :

$$I(x; y) = \sum_i \sum_j P(x_i) P(y_j|x_i) \log \frac{P(y_j|x_i)}{\sum_i P(x_i) P(y_j|x_i)}$$

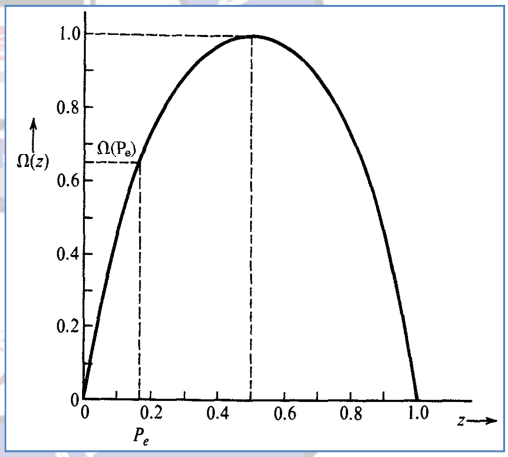
$$I(x;y) = \alpha\bar{P}_e \log\left(\frac{\bar{P}_e}{\alpha\bar{P}_e + \bar{\alpha}P_e}\right) + \alpha P_e \log\left(\frac{P_e}{\alpha P_e + \bar{\alpha}\bar{P}_e}\right)$$
$$+ \bar{\alpha}P_e \log\left(\frac{P_e}{\alpha\bar{P}_e + \bar{\alpha}P_e}\right) + \bar{\alpha}\bar{P}_e \log\left(\frac{\bar{P}_e}{\alpha P_e + \bar{\alpha}\bar{P}_e}\right)$$
$$= (\alpha P_e + \bar{\alpha}\bar{P}_e)\log\left(\frac{1}{\alpha P_e + \bar{\alpha}\bar{P}_e}\right)$$
$$+ (\alpha\bar{P}_e + \bar{\alpha}P_e)\log\left(\frac{1}{\alpha\bar{P}_e + \bar{\alpha}P_e}\right)$$
$$- \left(P_e \log\frac{1}{P_e} + \bar{P}_e \log\frac{1}{\bar{P}_e}\right)$$

Now, by defining:

$$\Omega(z) = z\log\frac{1}{z} + z'\log\frac{1}{z'} \quad , given\ that\ z' = 1 - z$$

$$I(x;y) = \Omega(\alpha p_e + \alpha' p'_e) - \Omega(p_e)$$

The function $\Omega(z)$ vs. z is shown on the right. It can be seen that $\Omega(z)$ is maximum at z= 0.5 (Note that we are interested in the region 0 < z < 1 only.) For a given Pe , $\Omega(Pe)$ is fixed. Hence, it follows that I(x; y) is maximum when $\Omega(\alpha p_e + \alpha' p'_e)$ is maximum. This occurs when



$$\alpha p_e + \alpha' p'_e = 0.5$$

or

$$\alpha p_e + (1-\alpha)(1-p_e) = 0.5$$

This equation is satisfied when: $\alpha = 0.5$
For this value of α, $\Omega(\alpha p_e + \alpha' p'_e) = 1$, and

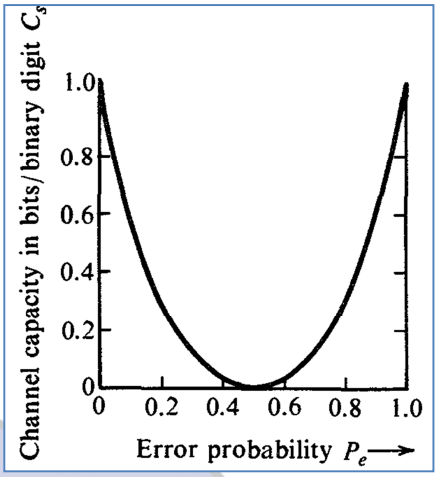
C_s = max I(x; y) = $1 - \Omega(p_e)$

$$C_s = 1 - [p_e \log\frac{1}{p_e} + (1-p_e)\log\frac{1}{1-p_e}]$$

The amount of information that y_j provides about x_i is called the mutual information between $x_i$and y_i. This is given by:

$$I(x_i, y_j) = \log_2\left(\frac{P(y_j \mid x_i)}{P(y_i)}\right)$$

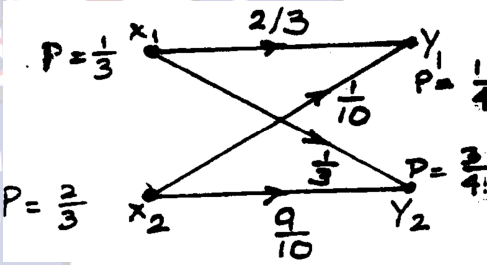
The Figure on the right of this paragraph, shows C_s vs. P_e. From this figure it follows that the maximum value of C_s is unity. This means we can transmit at most 1 bit of information per binary digit. This is the expected result, because one binary digit can convey one of the two equiprobable messages. The information content of one of the two equiprobable messages is $\log_2 2 = 1$ bit. Secondly, we observe that C_s is maximum when the error probability $P_e = 0$ or $P_e =$ l. When the error probability $P_e = 0$, the channel is noiseless, and we expect C_s to be maximum. But surprisingly, C_s is also maximum when $P_e = 1$. This is easy to explain, because a channel that consistently and with certainty makes errors is as good as a noiseless channel. All we have to do is reverse the decision that is made, and we have error-free reception; that is, if 0 is received, we decide that 1 was actually sent, and vice versa. The channel capacity C_s is zero (minimum) when $P_e = 0.5$. If the error probability is 0.5, then the transmitted symbols and the received symbols are statistically independent. If we received **0,** for example, either 1 or **0** is equally likely to have been transmitted, and the information received is zero , and in such cases the channal is said to be useless !



Example: For the following binary channel matrix:

Outputs

Inputs	y_1	y_2
x_1	$\frac{2}{3}$	$\frac{1}{3}$
x_2	$\frac{1}{10}$	$\frac{9}{10}$



Given that $P_x(x_1) = 1/3$ and $P_x(x_2) = 2/3$. Determine H(x), $H(x|y)$, $H(y)$, H(y|x), and I(x; y).

Sol:

$$P(y_1) = p(y_1|x_1)P(x_1)+P(y_1|x_2)P(x_2) = \frac{2}{3}.\frac{1}{3} + \frac{1}{10}.\frac{2}{3} = \frac{13}{45}$$

$$P(y_2) = 1\text{-}P(y_1) = 32/45$$

$$H(x) = P(x_1)\log\frac{1}{P(x_1)} + \mathrm{P}(x_2)\log\frac{1}{\mathrm{P}(x_2)}$$

$$= \frac{1}{3}log3 + \frac{2}{3}\log\frac{3}{2} = 0.918\ bits$$

To find H(x|y), we need to find:

$$P(x_1|y_1) = \frac{P(y_1|x_1)P(x_1)}{P(y_1)} = \frac{10}{13} \quad , \quad P(x_1|y_2) = \frac{P(y_2|x_1)P(x_1)}{P(y_2)} = \frac{5}{32}$$

$$P(x_2|y_1) = \frac{P(y_1|x_2)P(x_2)}{P(y_1)} = \frac{3}{13} \quad , \quad P(x_2|y_2) = \frac{P(y_2|x_2)P(x_2)}{P(y_2)} = \frac{54}{64}$$

$$H(x|y_1) = P(x_1|y_1)\ log\frac{1}{P(x_1|y_1)} + P(x_2|y_1)log\frac{1}{P(x_2|y_1)}$$

$$= \frac{10}{13}\log\frac{13}{10} + \frac{3}{13}\log\frac{13}{3} = 0.779$$

$$H(x|y_2) = P(x_1|y_2)\ log\frac{1}{P(x_1|y_2)} + P(x_2|y_2)log\frac{1}{P(x_2|y_2)}$$

$$= \frac{5}{32}\log\frac{32}{5} + \frac{54}{64}\log\frac{64}{54} = 0.624$$

$$H(x|y) = P(y1)H(x|\ y1) + P(y2)H(x|\ y2) = \frac{13}{45}(0.779) + \frac{32}{45}(0.624) = 0.6687\ bits$$

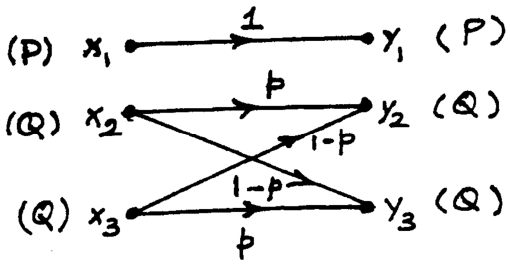
Thus,

$$I(x;y) = H(x) - H(x|y) = 0.918 - 0.6687 = 0.2489\ bits$$

$$H(y) = \sum_i P(y_i)log\ \frac{1}{P(y_i)} = \frac{13}{45}\log\frac{45}{13} + \frac{32}{45}\log\frac{45}{32} = 0.8673\ bits/symbol$$

$$H(y|x) = H(y) - I(x|y) = 0.8673 - 0.2493 = 0.618\ bits/symbol$$

For Ternary channal, we have :



and p(y1) = P , P(y2) = P(y3) =Q

We use the following formula to obtaine the P(x|y) matrix

$$P(x_i|y_j) = \frac{P(y_j|x_1)P(x_i)}{\sum_i P(x_i)P(y_j|x_1)}$$

$$y_j \begin{bmatrix} 1 & 0 & 0 \\ 0 & p & 1-p \\ 0 & 1-p & p \end{bmatrix} \quad x_i$$

$$H(x) = \sum P(x_i) \log \frac{1}{P(x_i)}$$
$$= -P \log P - 2Q \log Q \ , with\ 2Q = 1 - \boldsymbol{P}$$

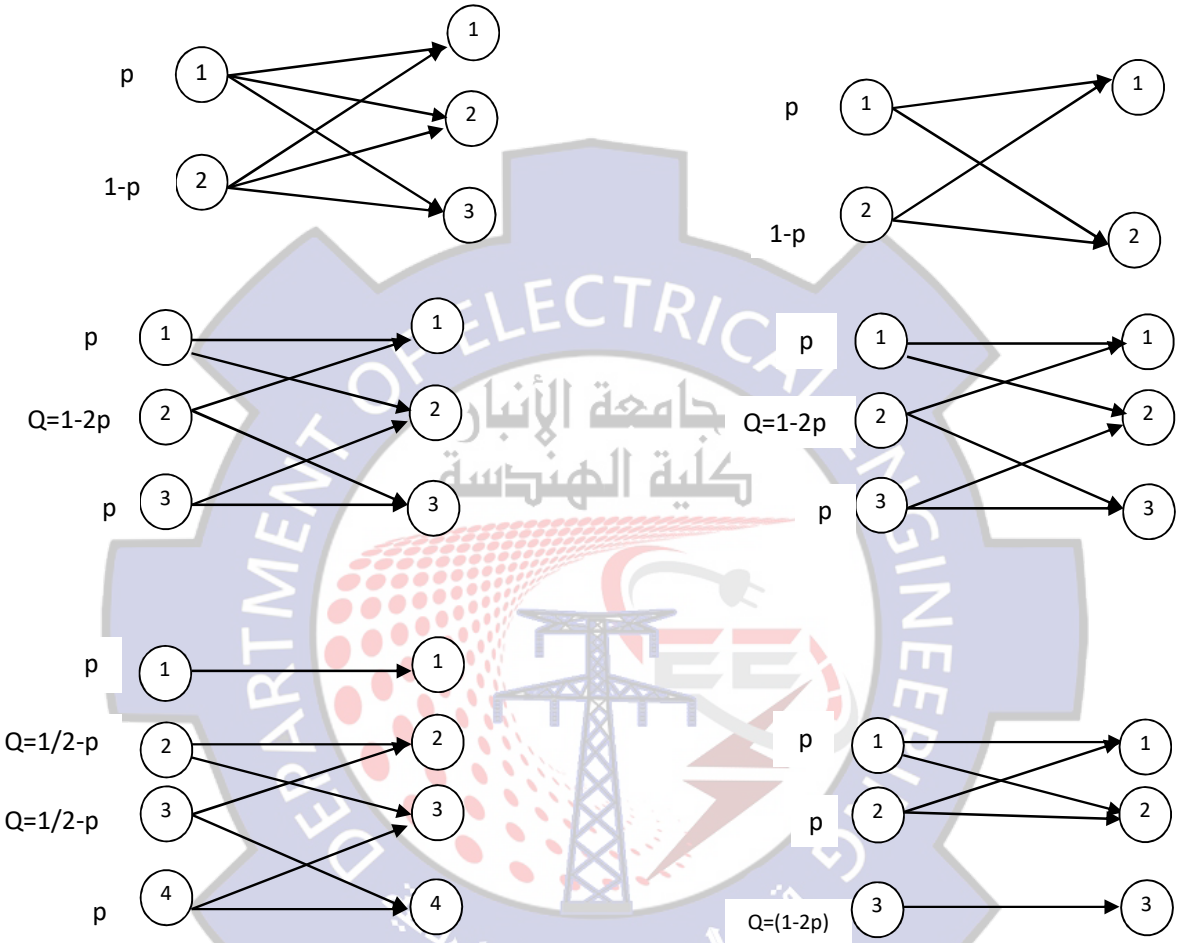
Also,

$$H(x|y) = \sum_i \sum_j P(y_j)P(x_i|y_j) \log \frac{1}{P(x_i|y_j)}$$

And continue the solution as usual.

Assigning the input probability accurately is an important aspect for right solution or finding the net data transferred through the channel matrix, here is some examples:

The channel matrix $P(y_j|x_i)$ is shown:



The conditional probability of various channel types is shown below:

a- $P(Y \mid X) = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$

is a BSC, because it is square matrix and 1st row is the permutation of 2nd row.

b- $P(Y \mid X) = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$

is TSC, because it is square matrix and each row is a permutation of others.

c- $P(Y \mid X) = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$

is a non-symmetric since it is not square although each row is permutation of others.

d- $P(Y \mid X) = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$

is a non-symmetric although it is square since 2nd row is not permutation of other rows.

Marginal Entropies:

Marginal entropies is a term usually used to denote both source entropy H(X) defined as before and the receiver entropy H(Y) given by:

$$H(Y) = -\sum_{j=1}^{m} P(y_j) \log_2 P(y_j) \frac{bits}{symbol}$$

Joint entropy and conditional entropy:

The average information associated with the pair (x_i, y_j) is called joint or system entropy H(X,Y):

$$H(X,Y) = H(XY) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 P(x_i, y_j) \quad bits/symbol$$

The average amount of information associated with the pairs $P(x_i \mid y_j)$ and $P(y_j \mid x_i)$ are called conditional entropies $H(Y \mid X) and\ H(X \mid Y)$, and given by:

$$H(Y \mid X) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 P(y_j \mid x_i) \quad bits/symbol$$

Return to first equation, we have: $P(x_i, y_j) = P(x_i) P(y_j \mid x_i)$, put inside log term

$$H(X,Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 P(x_i) - \sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 P(y_j \mid x_i)$$

But

$$\sum_{j=1}^{m} P(x_i, y_j) = P(x_i)$$

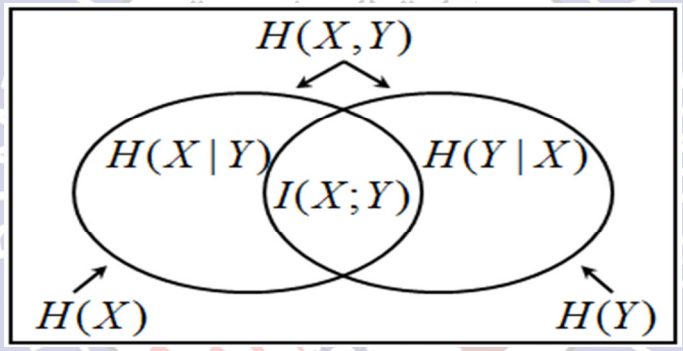
Put it in above equation yields:

$$H(X,Y) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) - \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j)\log_2 P(y_j \mid x_i)$$

So that:

$$H(X,Y) = H(X) + H(Y \mid X)$$

$$H(X,Y) = H(Y) + H(X \mid Y)$$

Mutual information can be equivalently expressed as:

$$\begin{aligned} \mathrm{I}(X;Y) &\equiv \mathrm{H}(X) - \mathrm{H}(X|Y) \\ &\equiv \mathrm{H}(Y) - \mathrm{H}(Y|X) \\ &\equiv \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y) \\ &\equiv \mathrm{H}(X,Y) - \mathrm{H}(X|Y) - \mathrm{H}(Y|X) \end{aligned}$$