Information & Coding Theory

Information Theory] Probability theory

\+

Algebraic Coding Theory] Linear Algebra

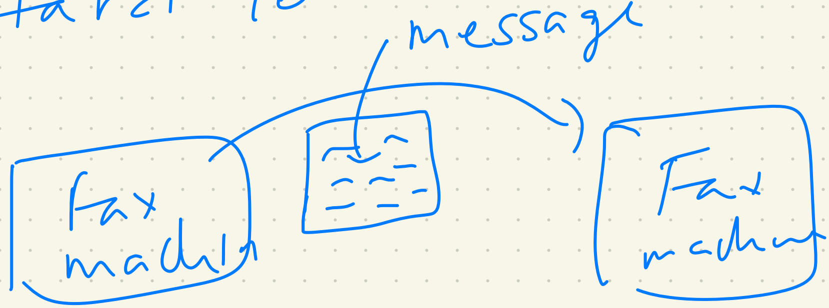
Information

1. Elements of Information Theory – Thomas & Thomas
2. A First Course in Information Th. – Raymond Yeung
3. The theory of information and Coding – McEliece R. J.

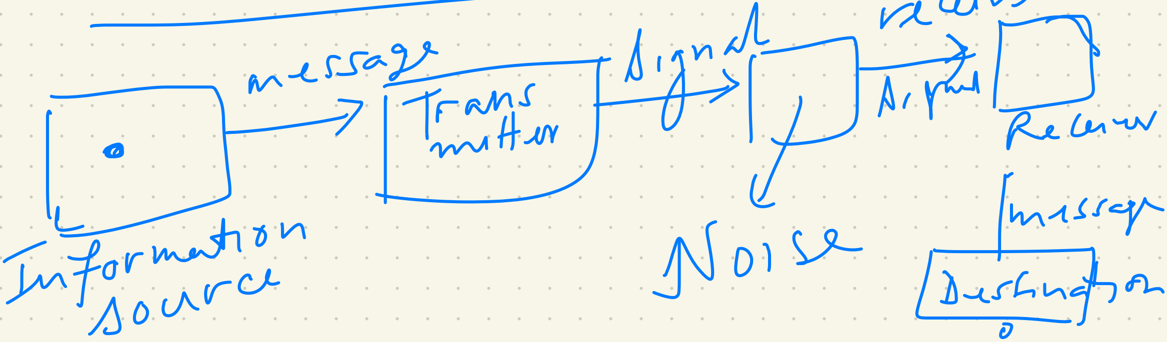
Saturday - 10 - 14 55 am
(instead of Friday)

Information, what is it?

- Hard to answer



1948, Claude E Shannon (1916 - 2001) published a paper title 'The mathematical theory of communication'



Shannon introduced two concept

① "Information is Uncertainty"

An information source is modeled as a 'random variable" or "random process"

② information which is to be transmitted, has to be digital

Information source converted into strings of 0 and 1 or bits (binary digits)

He proved two Theorems

① Source Coding Theorem

— he introduced the concept of 'entropy' as a fundamental measure of information

(2) Channel coding theorem

he introduced the concept of capacity of a noisy channel

Ensemble — A sample space and a probability measure

equivalent to probability space $(\Omega, \mathcal{A}, P)$

$P : \mathcal{A} \to [0, 1]$

(i) $P(A) \geq 0$, $A \in \mathcal{A}$

(ii) $P(\Omega) = 1$

(iii) if $A_1, \ldots, A_K$ are mutually exclusive

$$P(A_1 \cup \cdots \cup A_K) = \sum_{i=1}^{K} P(A_i)$$

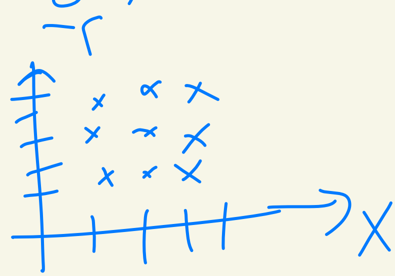
Let U be an ensemble with sample space $\{a_1, a_2, \ldots, a_k\}$,

$$P_U(a_i)$$

Let $\{a_1, \ldots, a_k\}$, $\{b_1, \ldots, b_j\}$ are sample spaces corr to two experiments.

X $\qquad\qquad$ Y

$$P_{XY}(a_i, b_j) = ?$$



$$P_X(a_i) = \sum_{j=1}^{J} P_{XY}(a_i, b_j)$$

marginal prob

$$P(x) = \sum_y P(x,y)$$

$$P(y) = \sum_x P(x,y)$$

Conditional Probability

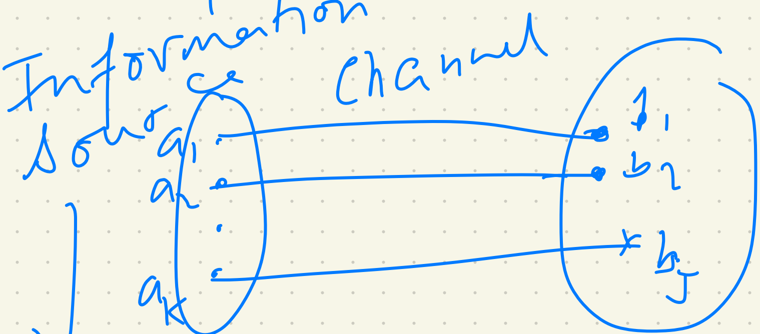
$$P_{Y|X}(b_j \mid a_k) = \frac{P_{XY}(b_j, a_k)}{P_X(a_k)}$$

when $P_X(a_k) \neq 0$

$$P(y \mid x) = \frac{P(x, y)}{P(x)}$$

Lec 2

Information : Shannon's points



random variable $X \to$ input

sender

random variable $Y \to$ output

receiver

XY

$\{a_2, a_1 . a_K\}$

$p_1 \quad p_2 \quad p_K$

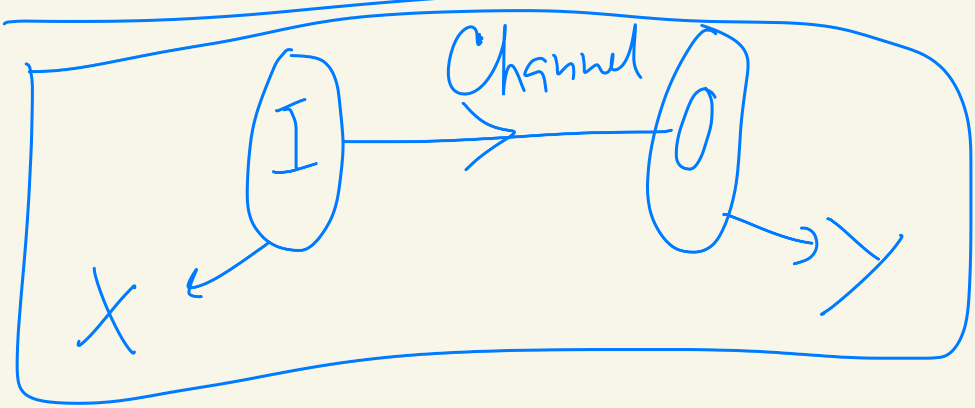
$\{b_1, b_2, \ldots, b_J\}$

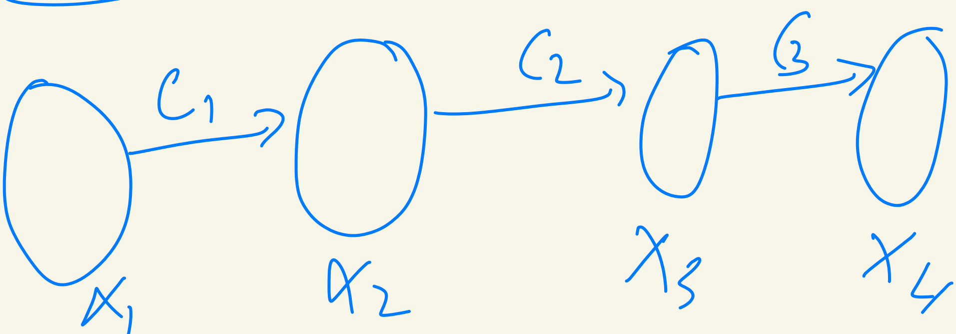
$q_1 \quad q_2 \quad q_K$

Two events $X = a_i$ and $Y = b_j$ are independent

$$P_{XY}(a_i, b_j) = P_X(a_i)\, P_Y(b_j)$$

Statistically independent





Mutual Information

$X \longrightarrow \{a_1, a_2, \quad a_K\}$

$Y \longrightarrow \{b_1, b_2, \quad , b_J\}$

$X, Y \qquad X \times Y \qquad P_{XY}(a_k, b_j)$

$\underline{\underline{x}}$ (input) $\qquad \underline{\underline{y}}$ (output)

$P(x, y) \longrightarrow$ Joint prob dist

posterior prob $\leftarrow P_{X|Y}(\underline{\underline{a_K}} \mid \underline{\underline{b_j}})$

prior $\leftarrow \boxed{P_X(a_K)}, \quad P_Y(b_j)$

Quantify the amount of "information" provided by the event $X = a_K$ by the occurrence of $y = b_j$?

$$I_{X;Y}(a_k, b_j) = \log \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)} \quad (1)$$

If the base of logarithm is 2, the unit of the numerical value is bits, it is called nats when the base is e

Observe $$P_{X|Y}(a_k|b_j) = \frac{P_{XY}(a_k, b_j)}{P_Y(b_j)}$$

Then from equation (1)

$$I_{Y;X}(b_j; a_k) = \log \frac{P_{Y|X}(b_j|a_k)}{P_Y(b_j)}$$

$$I_{Y;X}(b_j; a_k) = \log \frac{P_{XY}(a_k, b_j)}{P_X(a_k)\, P_Y(b_j)}$$

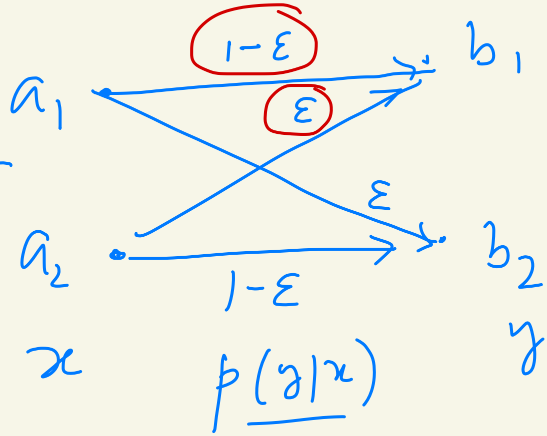
$$= \log \frac{P_{X|Y}(a_k | b_j)}{P_X(a_k)}$$

$$= I_{X,Y}(a_k; b_j)$$

called "mutual" information between the events $X = a_k$ and $Y = b_j$

$$I(x; y) = \log \frac{P(x|y)}{P(x)}$$

Exp Binary Symmetric channel



$0 < \varepsilon < 1$

$$p(b_1|a_1) = 1-\varepsilon = \frac{p(a_1, b_1)}{p(a_1)}$$

meaning: The output b_i is obtained with probability $1-\varepsilon$ when a_i is transmitted, $i=1,2$

Let $P_X(a_1) = \frac{1}{2} = P_X(a_2)$

$$P_{XY}(a_1, b_1) = \frac{1-\varepsilon}{2} = P_{XY}(a_2, b_2)$$

$$P_{XY}(a_1, b_2) = \frac{\varepsilon}{2} = P_{XY}(a_2, b_1)$$

Q = $I_{X;Y}(a_1, b_1) = \boxed{\log 2(1-\varepsilon)}$

$= I_{X;Y}(a_2; b_2)$

$I_{X,Y}(a_1, b_2) = \boxed{\log 2\varepsilon} = I_{X,Y}(a_2, b_1)$

Consider the following cases

(1) $I_{X,Y}(a_1; b_1)$

Observe this value when

(a) $\varepsilon = 0$
noiseless channel

(b) $\varepsilon = 1/2$
completely noisy channel

(c) $\varepsilon < 1/2$
mutual information +ve

If value of ε increases the mutual information decreases

(2) $I_{X,Y}(a_1; b_2) = I_{X,Y}(a_2; b_1)$

(a) $\boxed{\varepsilon \to 0}$

(b) $\varepsilon = 1/2$

(c) $\varepsilon < 1/2$

H.W

Recall

$$I_{X;Y}(\underbrace{a_k ; b_j}) = \log \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)}$$

Focus on this quantity

Obs Mutual information acts like a random variable

Then we define 'average mutual information' which is denoted by

$$I(X;Y)$$

$$= \sum_{k=1}^{K} \sum_{j=1}^{J} P_{XY}(a_k, b_j)\, I_{X;Y}(a_k, b_j)$$

$$= \sum_{k\,j} P_{XY}(a_k, b_j) \log \frac{P_{X|Y}(a_k|b_j)}{P(a_k)}$$

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x|y)}{p(x)}$$

$$\downarrow \qquad \qquad \swarrow$$

$$0 \qquad \qquad 0$$

Then $0 \log 0$ is treated is 0 only

<u>HW</u> Compute the average mutual information of the previous example

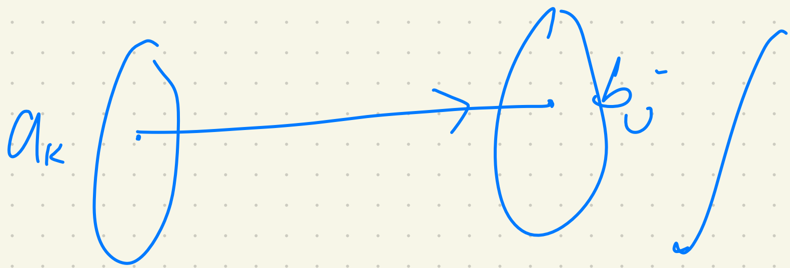
$$I_{X;Y}(a_k; b_j) = \log \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)}$$

<u>Q</u> What happens when $P_{X|Y}(a_k|b_j) = 1$?

if b_j uniquely specifies a_k then

Then $I_{X,Y}(a_k; b_j) = \log \frac{1}{P_X(a_k)}$

$$= -\log \boxed{P_X(a_k)}$$



define as the self-information of $x = a_k$ it is denoted as

$$I_X(a_k) = \log \frac{1}{P_X(a_k)}$$

i,e $\boxed{I(x) = -\log P(x)}$

It only depends on the r.v. X

nonnegative, and the value increases with decreasing $P_X(a_k)$

Q
=

Since Occurance of b_j gurantees the occurrence of a_j, hence $\boxed{I_\alpha(a_k)}$ is the amount of information which removes the uncertainty about occurrence of a_k

— o —