Information Theory and Coding (Lec-02)

Eftekhar Hossain
Lecturer
Dept. of ETE, CUET

Outline

- ➡ Average Self Information
- ➡ Entropy
- ➡ Conditional Entropy
- ➡ Joint Entropy

Average Self Information

- Consider a discrete random variable X with possible outcomes x_i , $i = 1, 2, 3....n$
- The average self information of the event $X = x_i$ is defined as

$$\boxed{H(X) = \sum_{i=1}^{n} P(x_i)I(x_i)}$$

$$\boxed{= -\sum_{i=1}^{n} P(x_i)logP(x_i)}$$

- When the base of the logarithm is 2 the units of $I(x)$ are in bits
- The entropy of X can be interpreted as the expected value of

$$log(\frac{1}{P(X)})$$

- $\boxed{H(X)}$ is called the Entropy

Average Self Information

- The term entropy has been borrowed from statistical mechanics. where it is used to denote the level of disorder in a system.

- We observe that since $0 \leq P(x_i) \leq 1$

$$log(\frac{1}{P(X)}) \geq 0$$

- $H(X) \geq 0$

Example

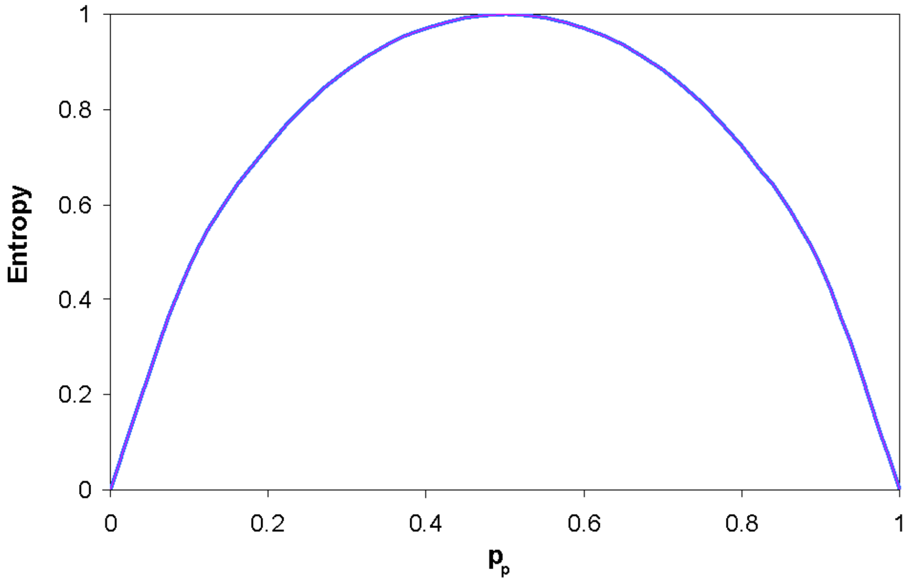
- Consider a discrete binary source that emits a sequence of statistically independent symbols.
- The output is either a 0 with probability p and a 1 with a probability 1 — p.
- The entropy of this binary source is

$$\boxed{H(X) = -(p)log_2(p) - (1-p)log_2(1-p)}$$

Example

$$H(X) = -(p)log_2(p) - (1-p)log_2(1-p)$$

The plot of the binary entropy function vs p is



Entropy of English alphabet

- Consider the English language with alphabet {A. B. .... Z}.
- If every letter occurred with the same probability and was independent from the other letters, then the entropy per letter would be $lo_2 26 = 4.70$
- This is the absolute upper bound.
- However. we know that all letters do not appear with equal probability.
- S. T. A. E are more frequent
- Q. J. Z. J are less frequent

CUET

Entropy of English alphabet

- Consider the English language with alphabet {A. B. .... Z}.
- If we take into consideration the probabilities of occurrences of different alphabets (normalized letter frequency), the entropy per letter, H_L would be

$$\boxed{H_L = 4.14 bits \leq H(X)}$$

- If $\boxed{X^2}$ denotes the random variable of bigrams in the English language. the upper bound on H_L can be defined as

$$\boxed{H_L \leq H(X) \approx 3.56 bits}$$

- Here we consider the probabilities of all pairs.

Entropy of English alphabet

- The logic can be extended to n-grams. Thus the entropy of the language can be defined as

$$\boxed{\lim_{x\to\infty} \frac{H(X^n)}{n}}$$

- Even though the exact value of H_L is difficult to determine, statistical investigations show that for the English language $\boxed{1 \leq H_L \leq 1.5 bits}$
- So each letter in the english text gives at most 1.5 bits of information.
- Let assume the value of H_L is 1.25 bits. Thus the redundancy of the English language is

$$\boxed{R_{Eng} = 1 - \frac{H_L}{log_2 26}}$$

$$\boxed{= 1 - \frac{1.25}{4.70} \approx 0.75}$$

CUET

Entropy of Spoken English

- ➡ Let us now consider the redundancy in spoken english.
- ➡ Suppose an average speaker speaks 60 words per minute and the average number of letters per word is 6.
- ➡ The average number of letters spoken per second in this case is 6 letters/sec.
- ➡ Assuming each letter carries 1.25 bits of information, the information rate of an average speaker is 7.5 bits/sec.
- ➡ If each letter is represented by 5 bits, the bitrate of an average speaker is 30 bit/sec.
- ➡ However, the typical data rate requirement for speech is 32 kilobits/sec.

Conditional Entropy

- The average conditional self information is called the conditional Entropy

$$H(X|Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) log(\frac{1}{P(x_i|y_j)})$$

- The physical interpretation of this definition is as follows
- $H(X|Y)$ Is the information (or uncertainty) in X after Y is observed
- Based on the definition of $H(X|Y)$ we can write

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Conditional Entropy and Mutual Information

$$\boxed{I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)}$$

➡ Since $I(X;Y) \geq 0$, it implies that $\boxed{H(X) \geq H(X|Y)}$

➡ The case $I(X;Y) = 0$ implies that $\boxed{H(X) = H(X|Y)}$ which is possible if and only if X and Y are statistically independent

Conditional Entropy and Mutual Information

$$\boxed{I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)}$$

- Since $\boxed{H(X|Y)}$ is the average amount of uncertainty(information) in X after we observe Y and $\boxed{H(X)}$ is the average amount of uncertainty (self information) of X. $\boxed{I(X;Y)}$ is the average amount of uncertainty (mutual information) about X having observed Y

- Since $\boxed{H(X) \geq H(X|Y)}$ the observation of Y does not increase the entropy(uncertainty). It can only decrease the entropy.

Joint Entropy

- The joint entropy of a pair of discrete random variables (X, Y) with a joint distribution $P(x, y)$ defined as

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

- By using the mathematical definitions of $H(X), H(X,Y)$ and $H(X|Y)$ we obtain the following chain rule

$$\boxed{H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|y)}$$

- And consequently $\boxed{I(X;Y) = H(X) + H(Y) - H(X,Y)}$

Venn Diagram