Entropy and Mutual Information

Sergey Belan

Skolkovo Institute of Science and Technology

14 April 2017

Definitions

Entropy:

$$S(X) = -\sum_{i=1}^{n_X} P(x_i) \log_2 P(x_i)$$

Joint entropy:

$$S(X,Y) = -\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} P(x_i, y_j) \log_2 P(x_i, y_j)$$

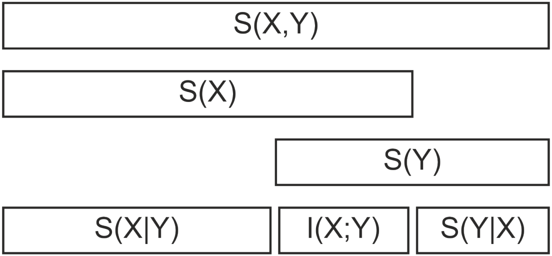
Conditional entropy:

$$S(Y|X) = -\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)}$$

Mutual information:

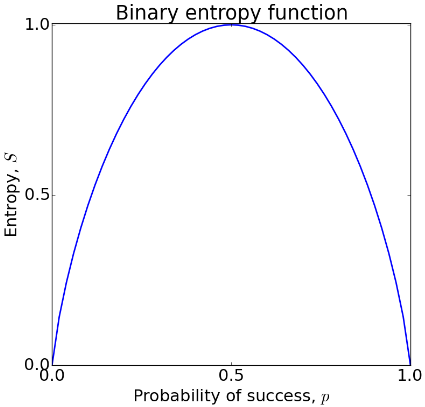
$$I(X;Y) = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_j)P(y_j)}$$

Definitions



$$I(X;Y) = S(X) - S(X|Y) = S(Y) - S(Y|X) =$$
$$= S(X) + S(Y) - S(X,Y)$$

Entropy of Bernoulli Process



$$S_{\text{binary}}(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

Exercise 1: Entropy of English language

The so called Zipf's law states that the frequency of the nth most frequent word in randomly chosen English document can be approximated by

$$p_n = \begin{cases} \frac{0.1}{n}, & \text{for} \quad n \in 1, \ldots, 12367 \\ 0, & \text{for} \quad n > 12367 \end{cases}$$

Under an assumption that English documents are generated by picking words at random, what is the entropy of English per word?

Exercise 2

The joint probability distribution $P(x, y)$ of two random variables X and Y is given in the table. Calculate the marginal probabilities $P(x)$ and $P(y)$, conditional probabilities $P(x|y)$ and $P(y|x)$, marginal entropies $S(X)$ and $S(Y)$, mutual information $I(X;Y)$.

$P(X,Y)$		X			
		x_1	x_2	x_3	x_4
Y	y_1	1/8	1/16	1/32	1/32
	y_2	1/16	1/8	1/32	1/32
	y_3	1/16	1/16	1/16	1/16
	y_4	1/4	0	0	0

Exercise 2. Solution

Marginal probability functions $P(x)$ and $P(y)$

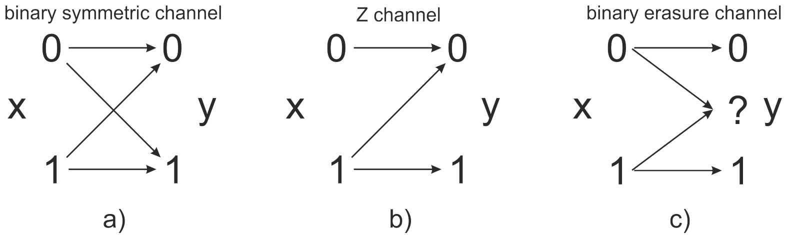
$P(x,y)$		x_1	x_2	x_3	x_4	$P(y)$
Y	y_1	1/8	1/16	1/32	1/32	1/4
	y_2	1/16	1/8	1/32	1/32	1/4
	y_3	1/16	1/16	1/16	1/16	1/4
	y_4	1/4	0	0	0	1/4
$P(x)$		1/2	1/4	1/8	1/8	

Exercise 2. Solution

Conditional probability function $P(x|y)$

$P(x\|y)$		X			
		x_1	x_2	x_3	x_4
Y	y_1	1/2	1/4	1/8	1/8
	y_2	1/4	1/2	1/8	1/8
	y_3	1/4	1/4	1/4	1/4
	y_4	1	0	0	0

Examples of communication channels



Exercise 3. Binary symmetric channel

Consider a binary symmetric channel with probability of error $f = 0.15$ and the following probability distribution of the input symbols: $P(x=0) = 0.9$, $P(x=1) = 0.1$. In other words, the input signal is a Bernoulli process with $p = 0.1$.

1) Calculate the probability distribution of output $P(y)$.
2) Compute the probability $x = 1$ given $y = 1$.
3) Compute the mutual information $I(X;Y)$.
4) What is the capacity of channel for arbitrary f?