ITDM - LEZIONE 04

DEL 08/10/2019

Consider two r.v. X and Y with,

$\mathcal{X} := \{x_1, \dots, x_{N_X}\}$ Possible values for X

$\mathcal{Y} := \{y_1, \dots, y_{N_Y}\}$ Possible values for Y

throughout this lesson we will denote with $P(X,Y)$ their JOINT PROBABILITY and with $P(X|Y) = \frac{P(X,Y)}{P(Y)}$ and $P(Y|X)$ their CONDITIONAL PROBABILITY.

We will now define some of the most important measures of information that deal with two r.v.

CONDITIONAL ENTROPY

The conditional entropy of X given Y is defined as

$$H(X|Y) := \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} P(x_i, y_j) \cdot \log_2 \frac{1}{P(x_i|y_j)}$$

$H(X|Y)$ can be interpreted as being "the remaining uncertainty about X once we have observed Y".

$H(Y|X)$ can be defined analogously.

$H(X|Y)$ has the following properties:

⓪ Generally, $H(X|Y) \neq H(Y|X)$ (ASYMMETRIC)

① $0 \leq H(X|Y) \leq H(X)$

That is, by observing a r.v. like Y we can only DECREASE the uncertainty of X if the two events are correlated.

② $H(X|Y) = H(X) \iff X \perp Y$

X and Y are INDEPENDENT

③ $H(X|X) = 0$

That is, if we observe X, then the uncertainty of X is 0.

JOINT ENTROPY

The joint entropy of X and Y is defined as

$$H(X,Y) := \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} P(x_i, y_j) \cdot \log_2 \frac{1}{P(x_i, y_j)}$$

$H(X,Y)$ has the following properties:

⓪ $H(X,Y) = H(Y,X)$ (SYMMETRIC)

① $\max\{H(X), H(Y)\} \leq H(X,Y) \leq H(X) + H(Y)$

② $H(X,Y) = H(X) + H(Y) \iff X \perp Y$

The ⇐ part follows from noting that, if $X \perp Y$, then $P(x_i, y_j) = P(x_i) \cdot P(y_j)$.

Mutual Information

The mutual information of X and Y is defined as follows

$$I(X,Y) := \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} P(x_i, y_j) \cdot \log_2 \frac{P(x_i, y_j)}{P(x_i) \cdot P(y_j)}$$

$I(X,Y)$ can be used to measure the information that is common between X and Y. Notice that if $X \perp Y$, that is if X and Y are independent and do not share any information then the arg of the $\log_2$ is 1 and thus the mutual information is 0.

$I(X,Y)$ has the following properties:

⓪ If $X \perp Y \Rightarrow I(X,Y) = 0$

(?) TODO

① $0 \leq I(X,Y) \leq \min \{H(X), H(Y)\}$

② $I(X,X) = H(X)$

③ $I(X,Y) = I(Y,X)$ (SYMMETRIC)

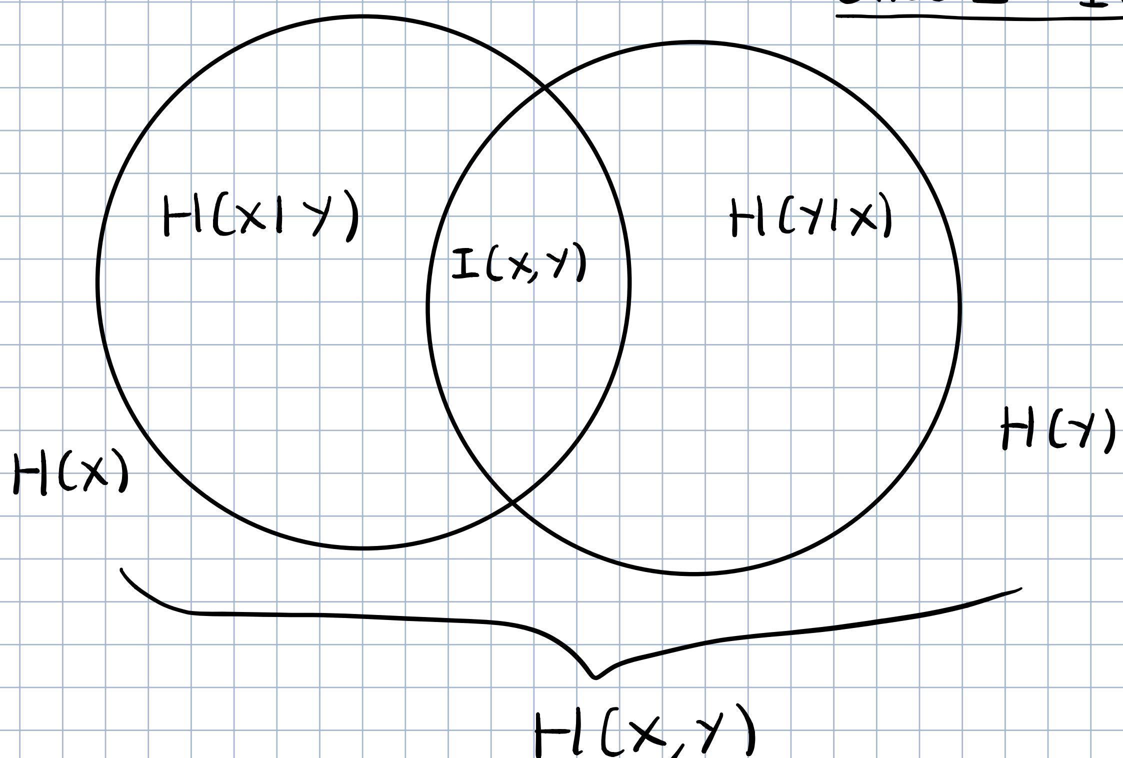
RELATIONS BETWEEN MEASURES OF INFORMATION

The following formulas show how the previously defined measure of information are related to each other:

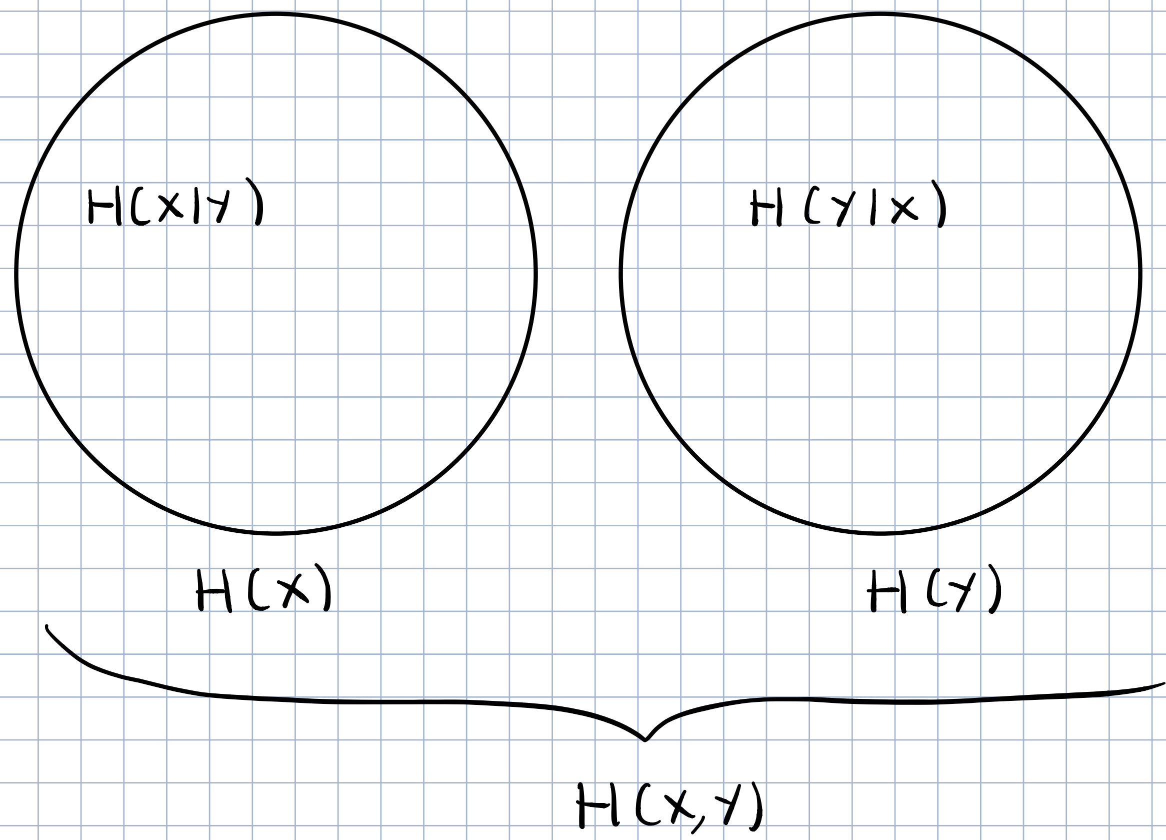
- $H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$
- $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
- $I(X,Y) = H(X) + H(Y) - H(X,Y)$

We can also use VENN-DIAGRAMS to construct a model of information in which to express the various relations:

CASE I: $I(X,Y) \neq 0$



CASE II : $I(X,Y)=0$



USAGE OF MEASURES OF INFORMATION

The measures introduced so far can be used to analyze the correlation between FEATURES in a given DATASET.

A FEATURE is a characteristic of the object we are trying to study. Every object has a particular set of measures for each feature we are interested in.

We can use the concept of RANDOM VARIABLES to model the different features in a dataset.

We can use the measures introduced so far to do various things with our data. For example we could try to reduce the REDUNDANCY of the information present in our dataset by removing features that are too similar to other features present in our dataset.

EXTENSION TO CONTINUOUS CASE

All of the measures defined can be extended to the continuous case by following these "rules of thumb":

1) (SUMS) $\sum\sum \longrightarrow \int\int$ (INTEGRALS)

2) (PMF) $P(x_i, y_j) \longrightarrow f(x_i, y_j)$ (PDF)