Mutual Information

https://dbis.ipd.kit.edu/download/paper-27.pdf

📍 **Ziel:** Du verstehst den Unterschied zwischen Korrelationskoeffizient und Mutual Information und weißt wie berechnet man den Mutual Information. Teile über Effizienz sind nicht relevant.

▼ What did authors try to accomplish?

- Efficently compute the MI (Mutual Information) with data that changes (nearest-neighbor based estimation of MI on dynamic data)

▼ What were the key elements of the approach?

- Mutual Information
  - captures both linear and arbitrary nonlinear dependencies
  - is defined on the probability density of the data ⇒ exact computation impossible on samples
  - MI estimators yield good results even for small samples
  - ▼ Definition

Mutual Information. Introduced by Shannon [26], the notion of *entropy* is a measure for the expected information from observing the value of a random variable X, noted as $H(X)$. The expected information for observed values of two random variables X and Y is the natural extension *joint entropy* $H(X, Y)$. This gives way to the notion of *Mutual Information*

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \tag{1}$$

which describes the information shared between both variables. Using the definition of entropy for continuous random variables in Equation 1 yields the differential definition of MI

$$I(X;Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx\, dy \tag{2}$$

where $p(x)$, $p(y)$ and $p(x, y)$ are the marginal and joint probability density functions of X and Y, respectively [8]. Using the natural logarithm, MI is then measured in the *natural unit of information* (nat).

- Structure
  1. Computational complexity of nearest-neighbor based estimators

2. Dynamic data structures
3. Near-Optimal computation time
4. Systematic experimental evaluation

- DEMI if the data size is small.
- ADEMI can be used for very data-intensive tasks

▼ What can you (possibly) use yourself?

- Mutual Information
- DEMI, ADEMI for dynamic data

▼ What other references do you want to follow?

- Janett Walters-Williams and Yan Li. 2009. Estimation of mutual information: A survey. In International Conference on Rough Sets and Knowledge Technology (RSKT'08). 389–396
- Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. Phys. Rev. E 69 (2004), 16. Issue 6.
- Fabian Keller, Emmanuel Müller, and Klemens Böhm. 2015. Estimating mutual information on data streams. In Proceedings of the 27th International Conference on Scientific and Statistical Database Management (SSDBM'15). ACM

Mutual information is a distance between two probability distributions. Correlation is a linear distance between two random variables. Although both of them are a measure of relationship between features, the MI is more general than correlation coefficient (CE) sine the CE is only able to take into account linear relationships but the MI can also handle non-linear relationships.

You can have a mutual information between any two probabilities defined for a set of symbols, while you cannot have a correlation between symbols that cannot naturally be mapped into a R^N space.

On the other hand, the mutual information does not make assumptions about some properties of the variables... If you are working with variables that are smooth, correlation may tell you more about them; for instance if their relationship is monotonic.