The Poincaré-Boltzmann Machine: passing the information between disciplines

Information Cohomology methods for learning the statistical structures of data

Pierre Baudot*, PhD[1], Mathieu Bernardi, MscI[1]
[1]Median Technologies, France
in col. with Daniel Bennequin, Monica Tapia, Jean-Marc Goaillard. *corresponding author e-mail: pierre.baudot@mediantechnologies.com

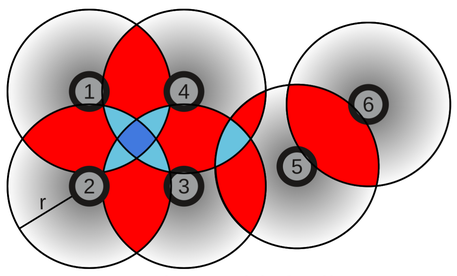
Abstract

Information cohomology is a branch of topological data analysis that allows us to quantify directly statistical dependencies and independences in a given dataset. Theorems establish Shannon entropy as a first cohomology class and mutual information as coboundaries on finite probability space endowed with a random variable chain complex structure. We present some simplicial subcase applications to supervised and unsupervised learning in different contexts: transcriptomic and digits and medical CT image classification.

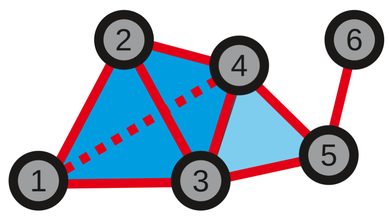
Principles and Theory

We consider random variables as partitions of atomic probabilities and the associated poset given by their lattice. The basic cohomology is settled by the Hochschild coboundary, with a left action corresponding to information conditioning. The first degree cocycle is the entropy chain rule, allowing to derive the functional equation of information and hence to characterize **entropy uniquely as the first group of the cohomology**. (minus) Odd multivariate mutual informations (MI, I_{2k+1}) appears as even degrees coboundary, and the introduction of a second trivial or symmetric action coboundary gives even MI (I_{2k}) in the odd degrees. **Mutual statistical independence is equivalent to the vanishing of all k-MI (I_k=0)**, leading to the conclusion that the I_k define refined measures of statistical dependencies and that the cohomology quantifies the obstructions to statistical factorization. We develop the computationally tractable subcase of on the simplicial (Boolean) sub-lattice, represented by **entropy H_k and information I_k landscapes.** The **H_k and I_k have linearly idenpendent gradients** in an open dense set of the probability simplex. (see [1,2,3,4] for theorems).

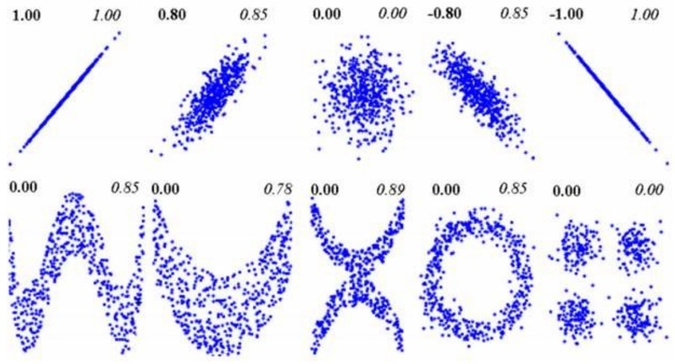
Persistence analysis principle (for comparison)



Data points with balls

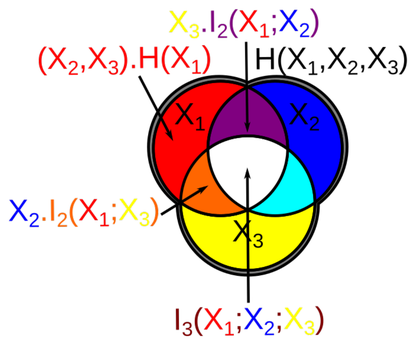


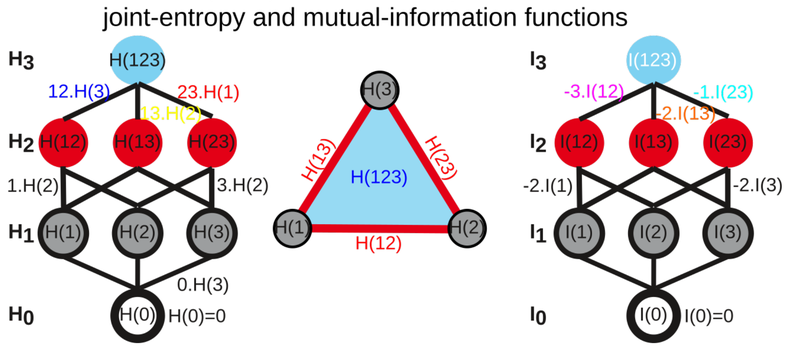
Čech 3-complex

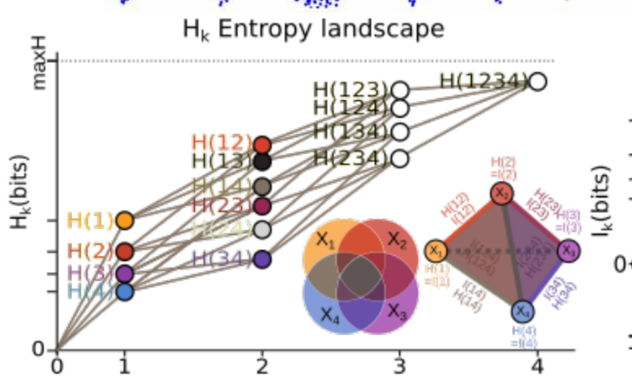


Mutual Informations generalize correlations and Cov.

(semi) lattice and simplex of joint-entropy and mutual-information functions







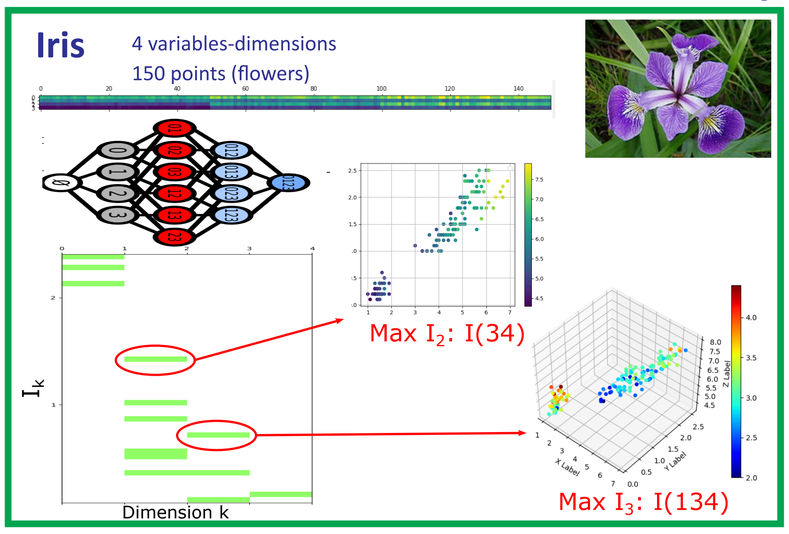
The marginal I_1 component defines a self-internal energy functional U_k, and **$I_{k,k>1}$** define the contribution of the **k-body interactions to the free energy functional G_k** given by the KL divergence between marginals and the joined variable (the "total correlation"). The set of information paths in simplicial structure is in bijection with the symmetric group.

Supervised learning: Let X_1 be the label variable to learn, then supervised learning of X_1 is defined by the sublattice, information landscapes and complexes for which all chains contain X_1, given by the 2^{n-1} H_k, I_k which (sub)gradients are independent in an open dense subsets of the probabilty subsimplex $\Delta_{X/X1}$ (subsimplex obtained by conditionning on the parametters E_{X1}). Then, the maximal depth of a Deep Neural Net acheiving the classification given the data is the dimension of the information simplicial complex (holds for supervised and unsupervised).

Unsupervised

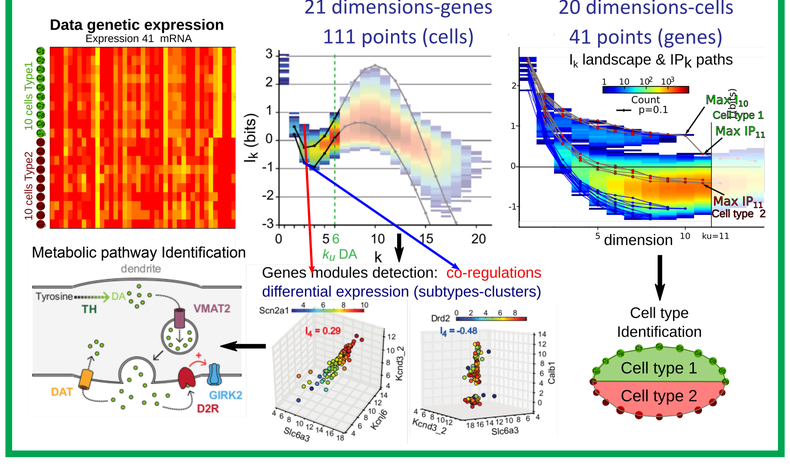
Iris

4 variables-dimensions
150 points (flowers)



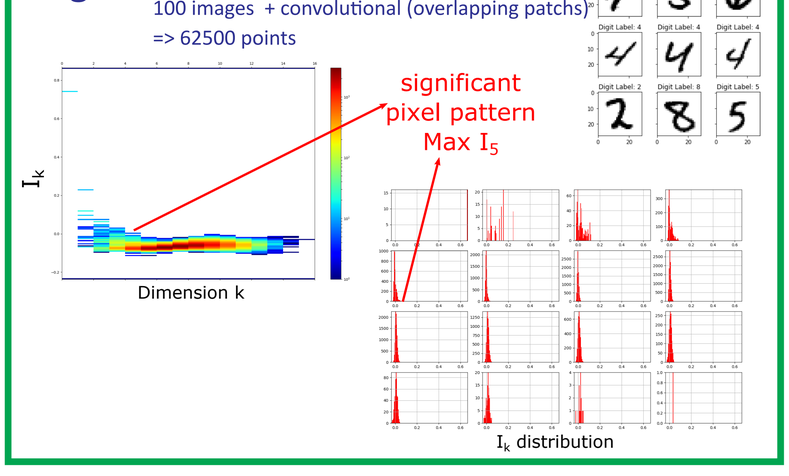
Transcriptome [5]

21 dimensions-genes
111 points (cells)



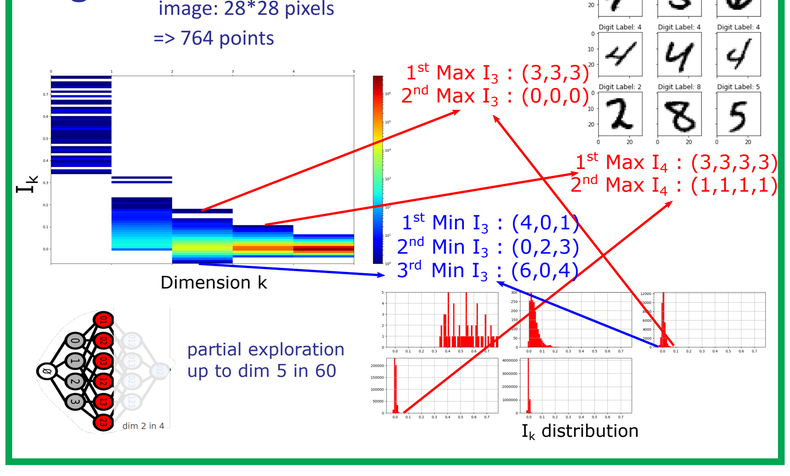
Digits

4*4 pixels patchs = 16 dimensions
100 images + convolutional (overlapping patchs)
=> 62500 points



Digits

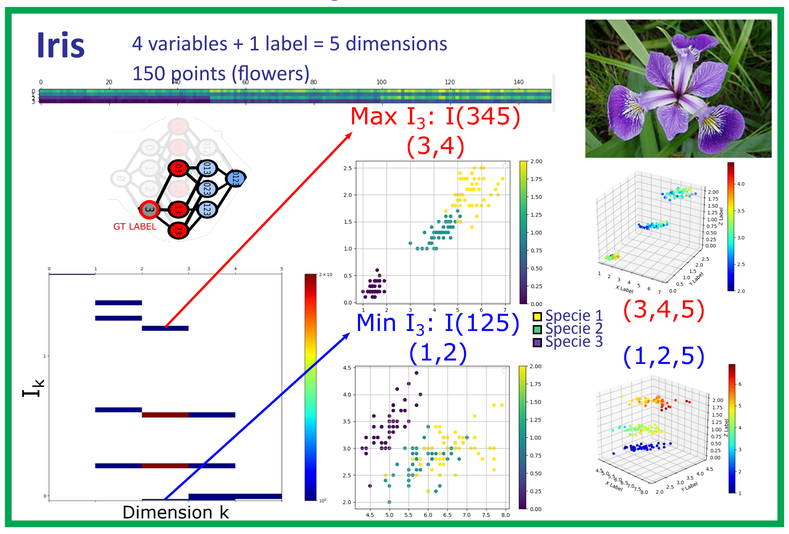
Transpose : 60 dimensions-images
image: 28*28 pixels
=> 764 points



Supervised

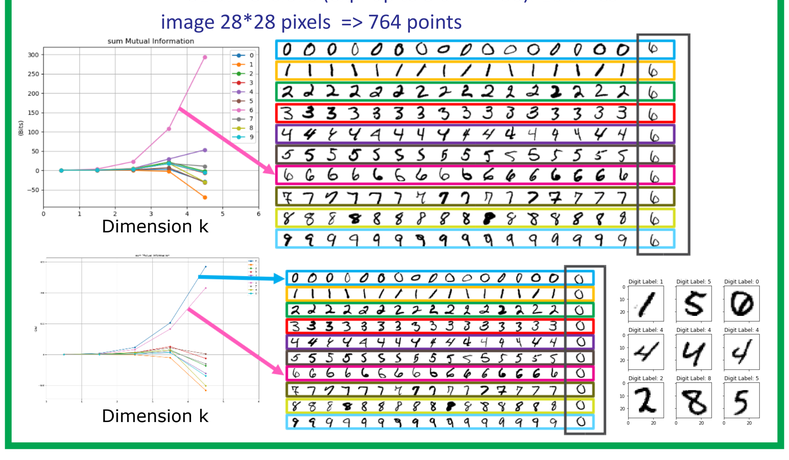
Iris

4 variables + 1 label = 5 dimensions
150 points (flowers)



Digits

Transpose : 30 images of the same digit + 1 label
=> 31 dimensions (exp: up to 5 dim in 30): 10 times
image 28*28 pixels => 764 points



Conclusions

1. New methods for topological data analysis intrinsicaly based on statistics. Encouraging results on data to be confirmed on larger training set and comparted to deep networks.
2. Generalization and formalisation of Deep Neural Networks with algebraic topology. New cohomological formalization of supervised learning as a subcase of supervised learning, for which the backpropagation (or natural gradient) is implemented by the information chain rule and is forward (cohomology).
3. Computationaly expensive $O(2^n)$ or $C(k,n)$ in partial exploration: current development of parrallell and GPU processing of the programs.

Bibliography

[1] P. Baudot and D. Bennequin. The homological nature of entropy. Entropy, 17(5):3253–3318, 2015.
[2] J.P. Vigneaux. Topology of statistical systems: a cohomological approach to information theory. PhD Thesis, 2019.
[3] P. Baudot, Tapia M., Bennequin, D, Goaillard J.M., Topological Information Data Analysis. 2019 arXiv:1907.04242
[4] P. Baudot, The Poincaré-Boltzmann Machine: from Statistical Physics to Machine Learning and back. 2019 arXiv:1907.06486
[5] M. Tapia & al; Neurotransmitter identity and electrophysiological phenotype are genetically coupled in midbrain dopaminergic neurons. Sci; Reports. 2018.