MUTUAL INFORMATION OF A FULLY CONNECTED NEURONAL NETWORK LEARNING PROCESS

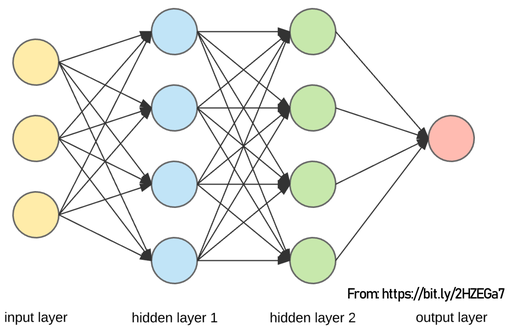
Natalia Copete, Diego A. Garzón, J. M. Pedraza, J. Forero-Romero

Universidad de los Andes
Colombia

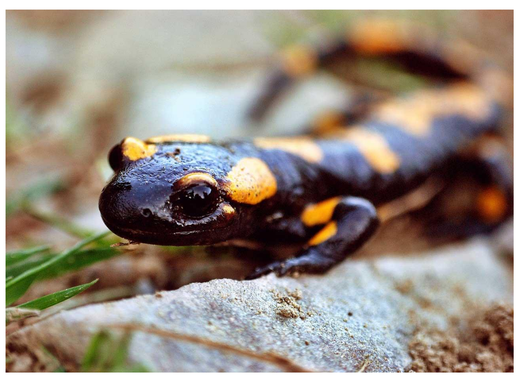
FULLY CONNECTED NEURAL NETWORK (FCNN)

An artificial neural network (ANN) is a group of connected processors that produce a sequence of real-valued activations in a way that is analogous to the brain. It is one of the most wide-spread approaches in Machine Learning (ML) community and it has showed a good performance in diverse classification tasks, such as chemical reactivity, imaging identification, language translation, protein structure, among others. In general terms, the structure of an ANN consists of an input and an output layer of neurons, and a certain number of hidden layers in between. In analogy to its biological counterparts,

ANNs possess their own axon connections: the neural weights. Every weight is a number that accomplishes the task of representing the connection between two pairs of neurons, by relating mathematically the output of one, to the input of the other. The basis of the learning process of an ANN is, in part, to optimize the values of those numbers.



Schneidman et al. found the importance of high order interactions in the neural network of retina cells of larval *Ambystoma tigrinum*, which works with a weak pairwise correlation and a strong global correlation.



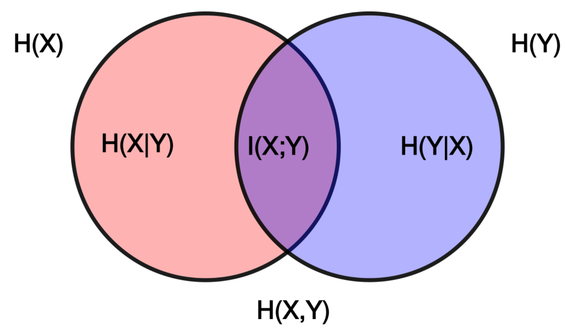
INFORMATION THEORY

The information theory is the study of the value of the information and the process of its communication. Nowadays, it is one of the most versatile theory with implications from the mechanical statistics to the financial world. Its more important quantity is the entropy, H, which describe the amount of data a stochastic source of data can generate.

$$H(x) = -\sum_{x \in X} p(x) \log p(x)$$

$$I(X;Y) = \frac{1}{n-1} \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

The mutual information,(MI) between two or more quantities stablish how much information you can obtain about one quantity if you know the other one. This two quantities are extremely correlated and let us stablish how much redundant is a quantity in terms of the other. This quantity has implications from RNA sequencing to stastical mechanics



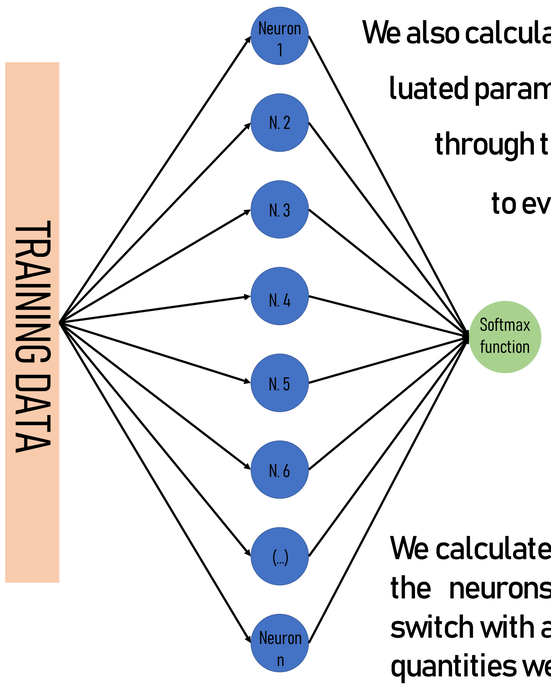
In biological studies, it has been found that weak correlations among pairs of neurons coexist with strong correlations in the states of population as a whole. In the present study we aimed to answer the question :

¿Does the FCNN behave in a similar way?

We have observed that, for a number of neurons bigger than 3, the mean normalized MI between pairs of neurons is significantly lower than the normalized global MI (all neurons), on a particular layer and during the entire learning process. Besides, as the FCNN learns, the mean local MI tends to decrease whereas the global MI tends to increase.

METHODOLOGY

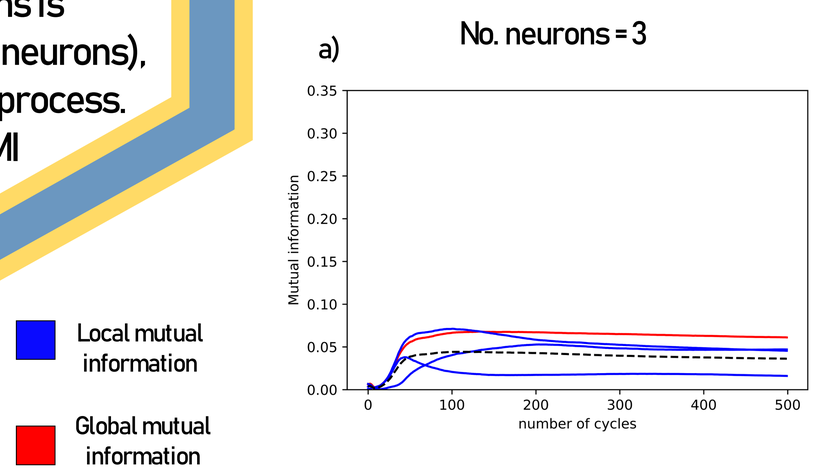
We trained FCNN of one layer of N neurons with a sigmoid activation function for the classification of handwritten numbers from 0 to 9 with a classification based on a softmax function and a optimizer type Adam in TensorFlow.



We also calculated MI and evaluated parameters of the FCCN through the training process. In order to evaluate the proficiency of the algorithm

We calculated local IMN (pairwise) and global IMN (all the neurons) assuming that the neurons were a switch with a 0.5 threshold , and analyze how this two quantities were related.

RESULTS



Here we show two graphs (a and b) to evidence the different behavior from a local (pairwise) versus a global perspective in two different layers: one composed by 3 neurons and the other by 16. Also, the (c) graph aims to show that mutual information obeys the same comportment for different switch thresholds. Finally, d and e graphs show that between all possible pairs of neurons in a particular layer, there is one pair that has a significantly bigger value of MI, which suggests the redundancy of one neuron on the layer

REFERENCES

- Elad Schneidman, Michael J. Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature, 440(7087):1007–1012, April 2006.
- Jurgen Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61:85–117, January 2015.
- Ball, K. R.; Grant, C.; Mundy, W. R.; Shafer, T. J. A Multivariate Extension of Mutual Information for Growing Neural Networks. *Neural Networks* **2017**, *95*, 29–43.
- MacKay, D. J. C.; Kay, D. J. C. M. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press, 2003.
- MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges http://yann.lecun.com/exdb/mnist/ (accessed May 1, 2019).