3IA Côte d'Azur
Institut interdisciplinaire d'intelligence artificielle

Can neural networks have homology?

Edouard Balzin

3IA postdoc, chair: Carlos Simpson

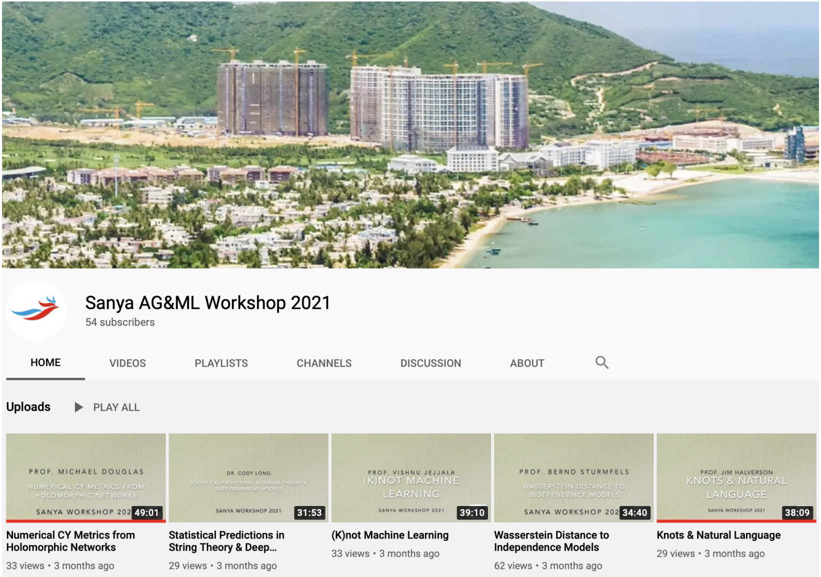
May 7, 2021

Last time Boris Shminke spoke of how to use ML and NN to work with mathematical objects.

This talk focuses on another project of ours, that tries to use ideas from what was recently pure mathematics to understand NN and ML.

Besides Boris, it is also joint with Sai Muttavarapu (M1 intern), Matan Prasma (Berlin)

Exchange between pure math and AI is picking up



Of course, neural networks have been studied mathematically for quite a while, from the analytical viewpoint (Cybenko's Universal Approximation Theorem), or with statistical tools.

Recently, there have been attempts to apply category theory, topology, homological algebra as means of figuring out what is going on in the realm of ML.

One hope is that we are dealing with "new physics" here: the observations of string theory gave birth to domains like Mirror Symmetry. Maybe the AI can do something similar?

Another hope is of course that studying NNs with different mathematics will lead to new practical consequences.

What is homology? I

Powerful idea, started at classifying spaces. Associate algebraic invariants:

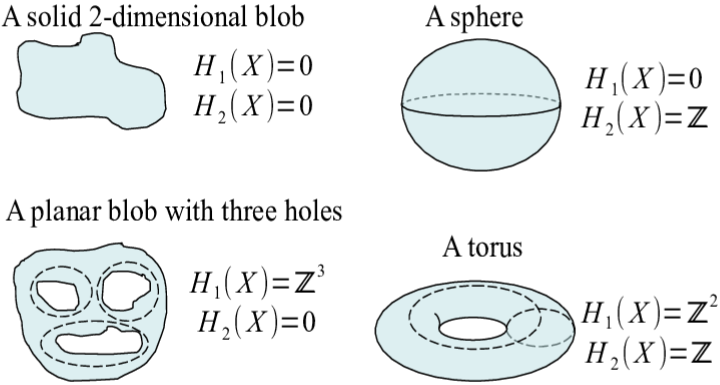


Image source: Walker, Brenton (2016, DOI 10.1109/INFOCOM.2016.7524493)

What is homology? II

Studying spaces algebraically became the discipline of algebraic topology.

However, in mathematics, most objects can be encoded as spaces, in a certain sense.

Consequences, last 30 years:

- Algebraic Geometry $\implies$ Derived Algebraic Geometry,
- Category Theory $\implies$ Higher Category Theory,
- Mathematical Physics $\implies$ Homological Mirror Symmetry, Topological Quantum Field Theories.

So, what about DNN?

Our project started when M.P. found a paper that tried to associate homology to a deep neural network in a supervised learning context.

The paper (cited at the end) claimed that their homology groups correlate with learning process, and detect stability of the model with respect to adversarial attacks.

Trying to run the code, we instantly encountered the following issue:

ML papers: claims vs reality

Correlation metric

Let F be a DNN (feed-forward, fully-connected).

Let $n_1(F), n_2(F) : \mathbb{R}^m \to \mathbb{R}$ be two neurons of F. We can consider, for a dataset $\{x_i\}_I \in \mathbb{R}^m$ (usually a training dataset), the following number:

$$d(n_1(F), n_2(F)) := \sqrt{1 - |c(n_1(F), n_2(F))|},$$

where

$$c(n_1(F), n_2(F)) = \frac{\sum(n_1(F)(x_i) - \overline{n_1(F)(x)})(n_2(F)(x_i) - \overline{n_2(F)(x)})}{\sqrt{\sum(n_1(F)(x_i) - \overline{n_1(F)(x)})^2 \sum(n_2(F)(x_i) - \overline{n_2(F)(x)})^2}}$$

or 0 if the denominator of the above expression is 0.

Lemma 1. *Let $X(F) = \{n_j(F)\}$ the set of all neurons at all layers. Then d as defined above makes X into a metric space.*

Homology of finite metric spaces

The pair $(X(F), d)$ is already an interesting object that changes with gradient descent. It is still quite complicated to analyse.

The idea is then to treat the set of neurons with the metric d as a dataset, and apply the technology of persistent homology.

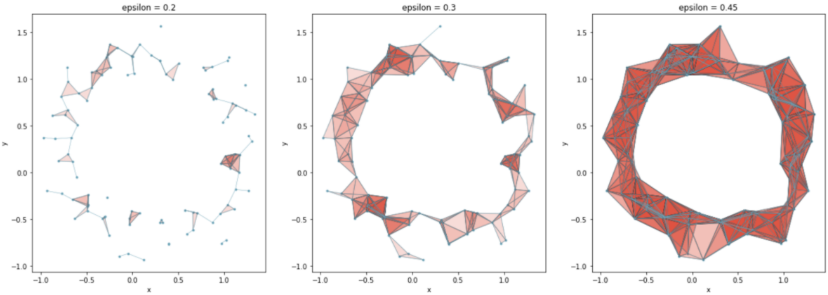
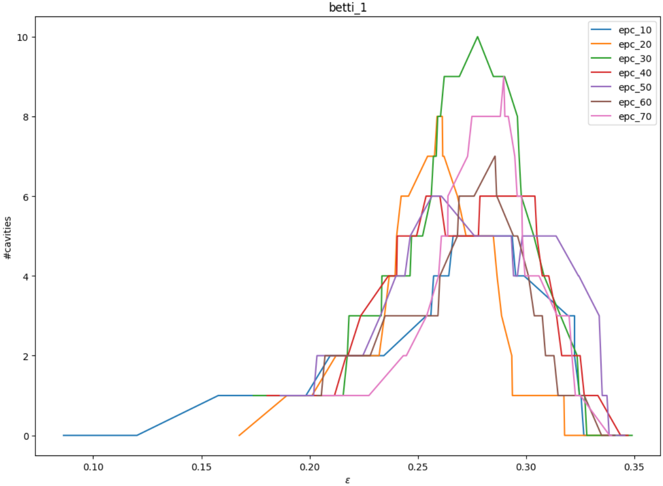
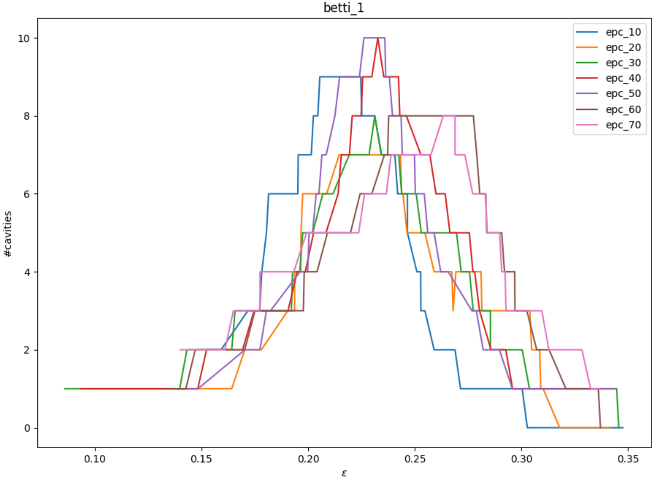


Image source: *https://eric-bunch.github.io/blog/topological-data-analysis-and-persistent-homology*

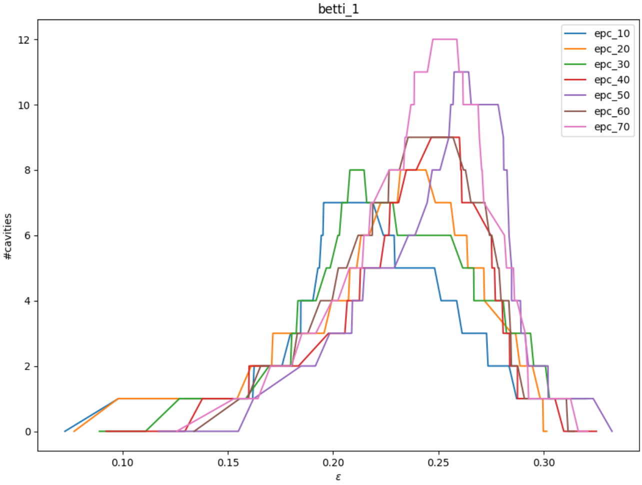
Example: classification problem of $\mathbb{D}^4 \subset [0,1]^4$, $\mathbb{R}^4 \to \mathbb{R}^{64} \to \mathbb{R}^{64} \to \mathbb{R}^2$, ReLU and finally Softmax.



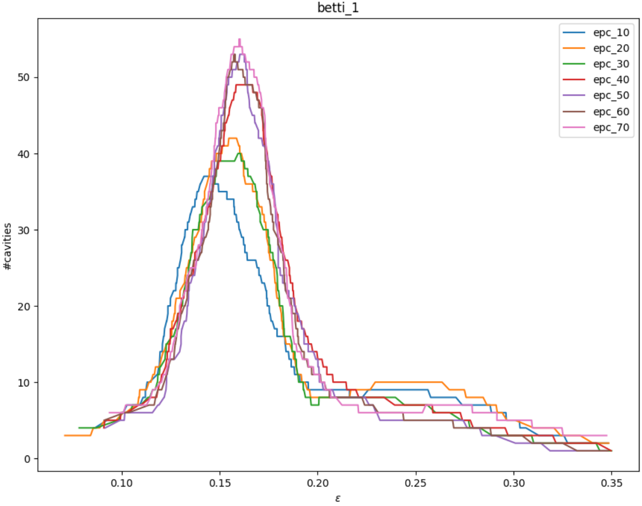
Example: classification problem of $\mathbb{D}^4 \subset [0,1]^4$,
$\mathbb{R}^4 \to \mathbb{R}^{128} \to \mathbb{R}^{128} \to \mathbb{R}^2$, ReLU and finally Softmax.



Example: classification problem of $\mathbb{D}^4 \subset [0,1]^4$,
$\mathbb{R}^4 \to \mathbb{R}^{128} \to \mathbb{R}^{128} \to \mathbb{R}^{128} \to \mathbb{R}^2$, ReLU and finally Softmax.



Example: classification problem of $\mathbb{D}^4 \subset [0,1]^4$,
$\mathbb{R}^4 \to \mathbb{R}^{512} \to \mathbb{R}^{512} \to \mathbb{R}^2$, ReLU and finally Softmax.



It appears that the homology is continuous in some sense, both for deformations within the same architecture (training) and for changes of architecture.

Perhaps the homological description of the process abstracts from the architecture of the particular classifier network, leading to some new kind of object?

How to formalise any of this?

If we have $F_n \to f$, does this permit to define a sort of homology $H_*(F_n, \varepsilon) \to H_*(f)$? What is the meaning of this homology?

What other constructions are sensible variations of the neuron metric space? Maybe (persistent) homology is superfluous?

Why are there no β_2 in examples above?

The code used for the computations is a modified version of:

***

Ciprian Cornelanu *DNN-topology*,
`https://github.com/cipriancorneanu/dnn-topology`

C. A. Corneanu, M. Madadi, S. Escalera, and A. M. Martinez, *What does it mean to learn in deep networks? and, how does one detect adversarial attacks?* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4757–4766, 2019

***

This code is incredibly inconvenient to run new numerical experiments and further develop the ideas. Work with B.S. and S.M. – writing something more adapted.