Neural Semigroups and Shapes of Learning

Edouard Balzin

AI and Mathematics

- Chair holder: Carlos Simpson
- 3IA axis: Core Elements of AI
- Postdoc: Edouard Balzin
- Contract: 9/2020-8/2022

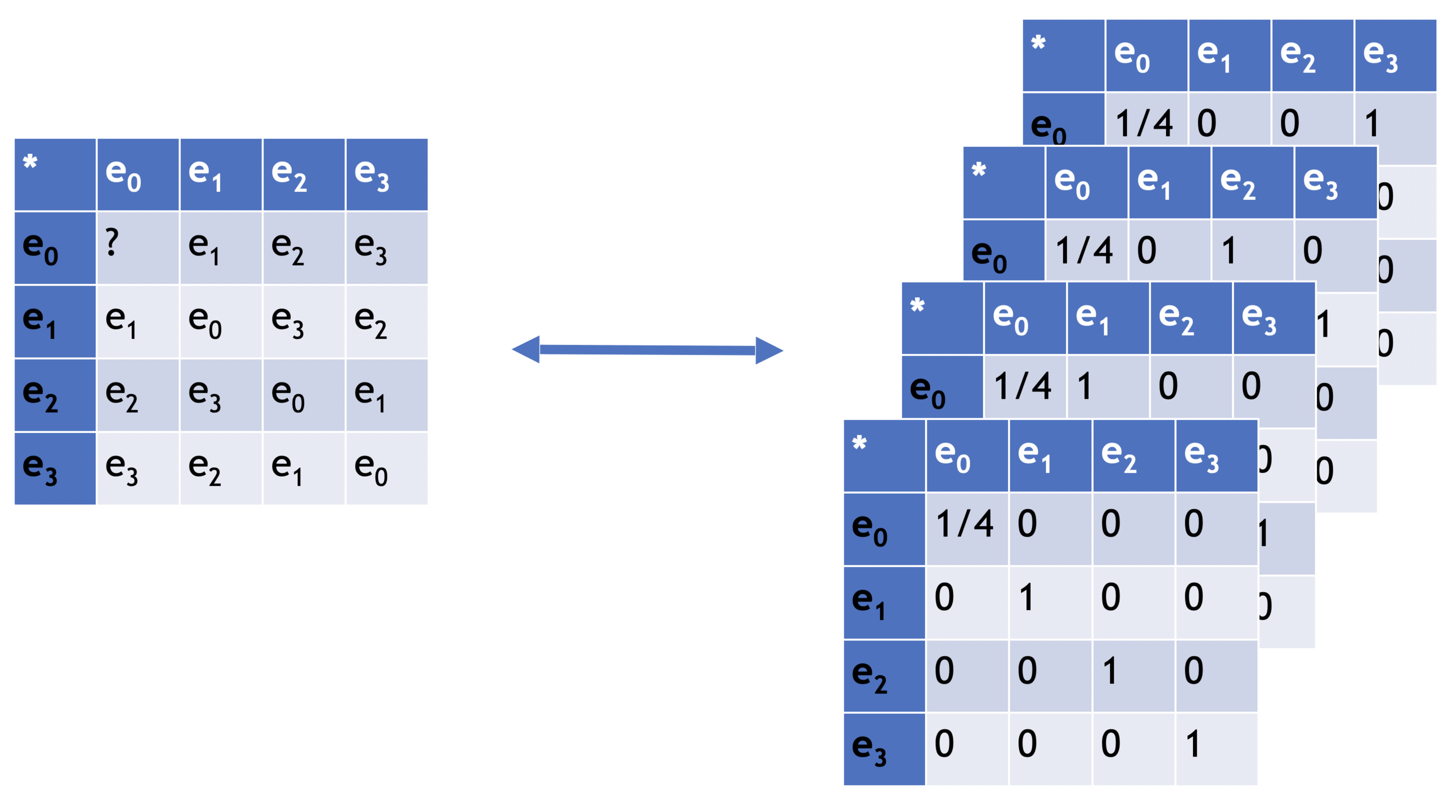
A Neural Network for Semigroups

- Based on *E. Balzin and B. Shminke, "A Neural Network for Semigroups", arXiv:2103.07388.*
- My question here: can neural networks understand things about mathematical objects?
- A semigroup is a set S together with an operation $\cdot : S \times S \to S$ that is associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. Examples: groups, monoids, can be obtained from categories, from finite automata.
- Assume S finite, then can numerate the elements and represent $\cdot$ by a multiplication table:

	e_0	e_1	e_2	e_3
e_0	e_0	e_1	e_2	e_3
e_1	e_1	e_0	e_3	e_2
e_2	e_2	e_3	e_0	e_1
e_3	e_3	e_2	e_1	e_0

	f_1	f_2	f_3	f_4	f_5
f_1	f_1	f_1	f_1	f_1	f_1
f_2	f_1	f_1	f_1	f_1	f_1
f_3	f_1	f_1	f_2	f_1	f_2
f_4	f_1	f_1	f_1	f_2	f_1
f_5	f_1	f_1	f_2	f_1	f_2

  How many are there tables for given $|S|$ up to isomorphism / anti-isomorphism?
- Surprisingly many: already around two billion for $|S| = 8$. Classification is unknown in general. Can a neural network understand something about such data?
- in our paper: try filling partial multiplication tables using NN. Given $\cdot : S \times S \to S$, present $S = \{1, ..., n\}$ and define $C_{ij}^k = 1$ if $i \cdot j = k$ and 0 otherwise. C_{ij}^k is a tensor of dimension n^3.
- What if we have a partial table? Can consider $C_{ij}^k = 1/n$ for unknown products $i \cdot j$.



- Idea: take a NN that accepts tensors of $\dim = n^3$, use it as a denoising autoencoder that "denoises" the missing cells. What loss function?
- First idea: supervised learning from known semigroup tables.
- Second idea: learning with associator loss

$$\mathrm{AL}(x, y) := \mathrm{KL}\left(\sum_{m=1}^{n} y_{ij}^m y_{mk}^l, \sum_{m=1}^{n} y_{im}^l y_{jk}^m\right)$$

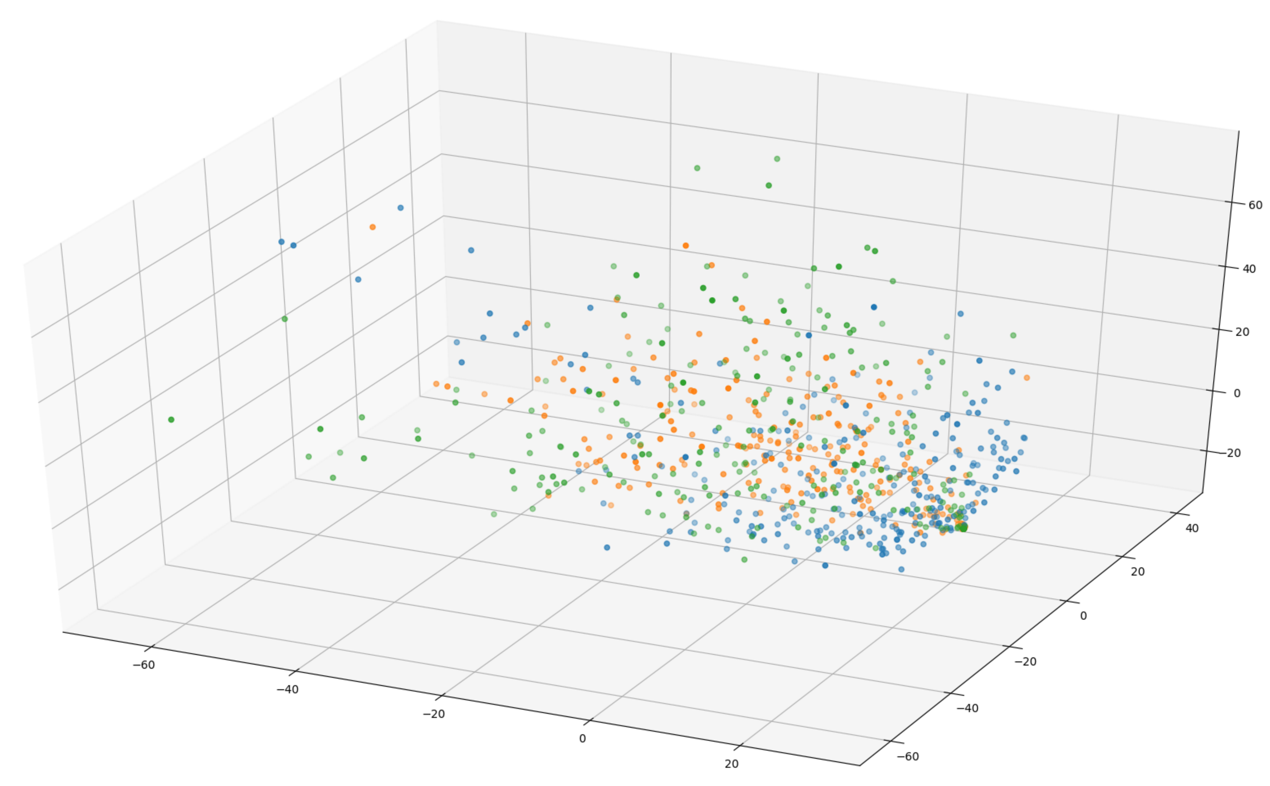
- the paper tests this network for $n = 5$ (around 200k tables). Go ask Boris!

Going further

- Work in progress with B. Shminke and Z. Bulic.
- Generalise the semigroup NN to an unsupervised learning problem of generating semigroups.
- Can such a NN generator find all known semigroups? What will it find for unknown cardinalities?
- Autoencoders provide "feature sets" on inner layers. What about "generalised features" of semigroups?
- What about other algebraic structures?

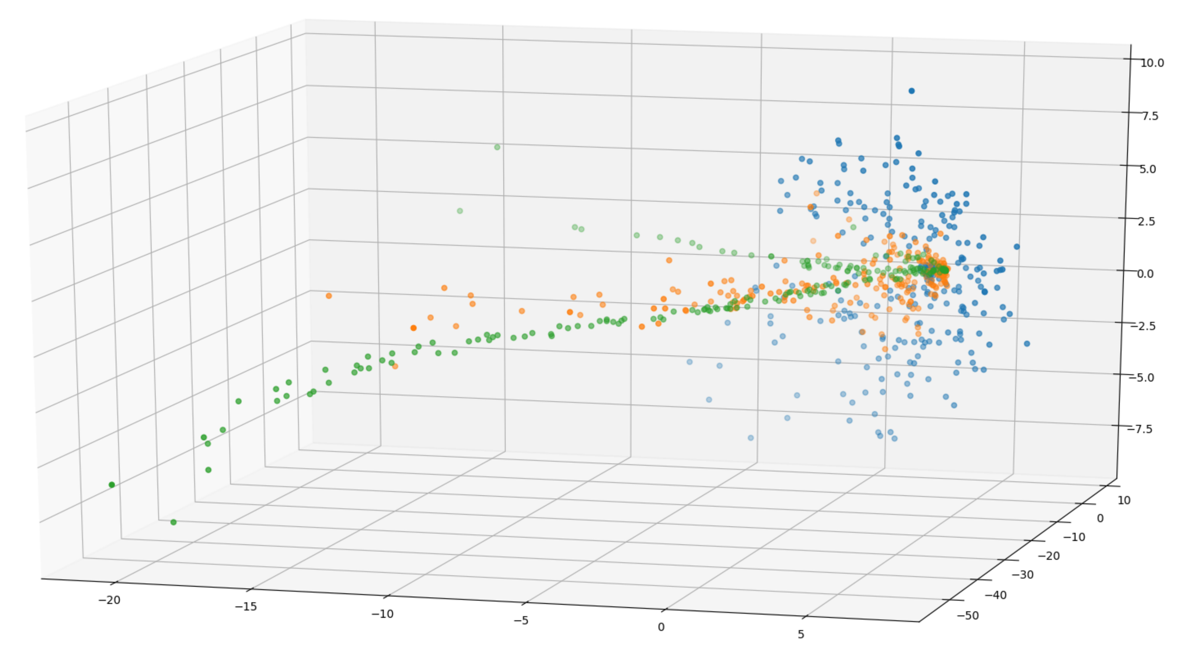
Topology of supervised learning

- Builds on *Corneanu et al, "What Does It Mean to Learn in Deep Networks? And, How Does One Detect Adversarial Attacks?"* `https://github.com/ciprianconeanu/dnn-topology`
- Given a neural network $D : \mathbb{R}^I \to \mathbb{R}^O$, consider the set $N(D)$ of all its neurons. The values of a neuron $n : \mathbb{R}^I \to \mathbb{R}$ are called activation values.
- Fixing an ordered finite set $X \subset \mathbb{R}^I$ (for example a training/testing set) allows to write a vector $(n(x_1), ..., n(x_n)) \in \mathbb{R}^X$.
- This provides a map $N(D) \to \mathbb{R}^X$. The set of neurons is given a shape. What does it look like, how to study it? Does it tell anything useful about learning?
- Our goals: Study of topological structures arising on/from $N(D)$ for fully connected NN trained on synthetic and realistic data like MNIST.

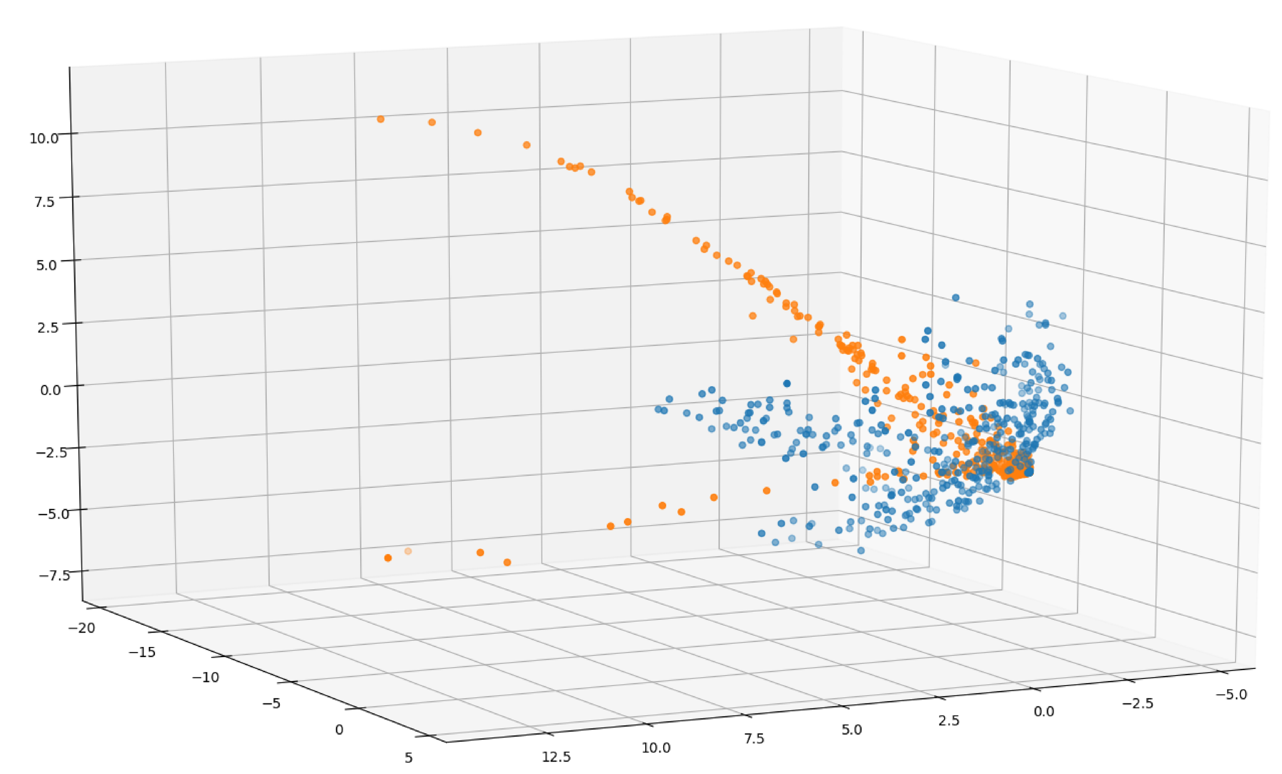


A three layer fully connected NN trained to classify MNIST viewed using 3D MDS

- Synthetic case: classifying points in the 4-dimensional unit ball using different architectures. Witness some sort of "shape invariance":



Three layers of 256 neurons



Two layers of 384 neurons

- Problem: how to formalise this apparent invariance? Corneanu et al tried to define homology of NN, is it of any use here? Ask me for details!