Learnable Simplicial Complexes

Towards the topological statistical learning

Mr. Dmitrii Pasechnyuk-Vilensky
http://dmivilensky.ru

Mr. Daniil Doroshenko

جامعة محمد بن زايد للذكاء الاصطناعي

Московский государственный университет имени М.В. Ломоносова

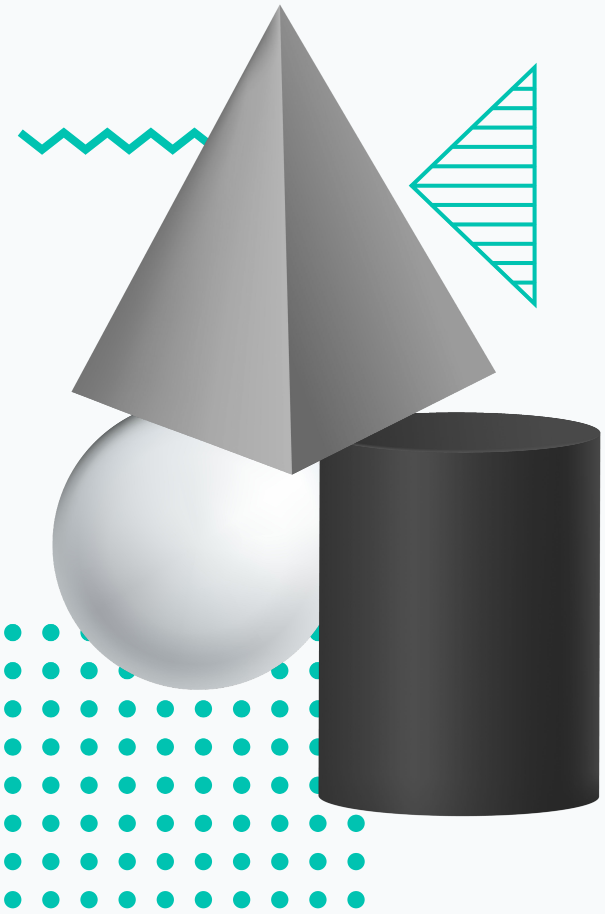
20/04/2025

Table of contents

01 Motivation

What are the assumptions and the basement of neural networks, and are they outdated?

- Data belong to the finite-dimensional vector space or to a Riemannian manifold embedded in it → no freedom in algebraic structure of date and trivial topologies
- Embeddings of the data must be separable in some metric space → lack of interpret- and explainability
- Structure of model must mimic the biological mind → lack of mathematical studies of neural networks



What do simplicial complexes can suggest?

Dynamicity and independence

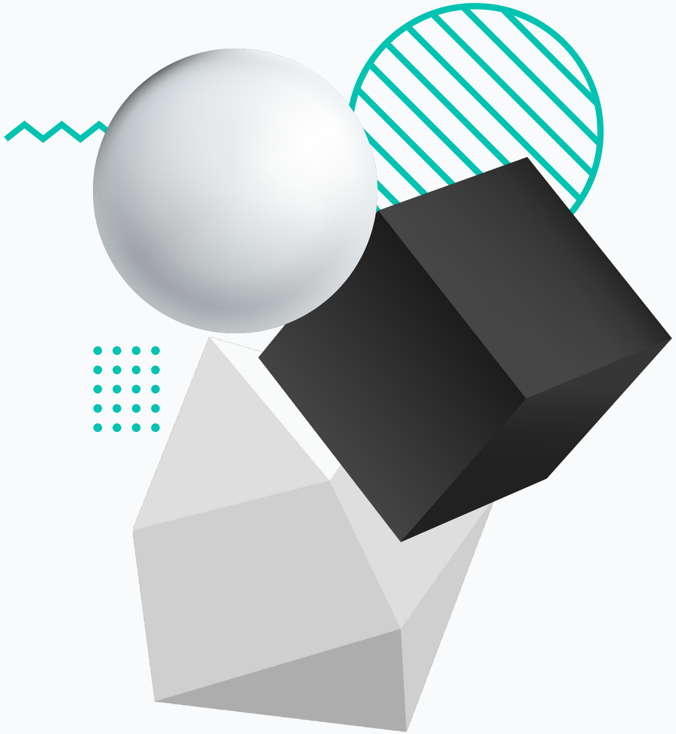
Data may have arbitrary structure initially, but we provide them with a structure, noting the patterns

Hierarchy and complexity control

The resulting embedding is defined on a space with structure, which one can explain and modify if needed

Fine mathematical object to study

The behaviour of the model depends on the structure of the complex, not on the optimization artifacts



02 Activation Topology

why

To construct (>0)-skeletons inductively
To join simultaneously activated objects
(if the edges were activated with similar
intensities connect them with a triangle)

how

(We assume that $|K_n| \in \mathbb{N}$)

$$\partial_{n+1}(\Sigma \in K_{n+1}) = \sigma_2 - \sigma_1 \in C(K_n)$$

$$w(\Sigma) = (\Phi \circ w \circ \delta_{n+1})(\Sigma)$$

where w is linear on $C(K_n)$, $\Phi(x) = \exp(-\gamma x^2)$

$$\frac{\partial w}{\partial w(\sigma)} = \frac{\partial \Phi}{\partial x} \circ \pi_\sigma \circ \partial_{n+1}$$

03 Diffusion

why

To eliminate (>0)-skeletons inductively

To activate the whole group if any member has been activated

how

$$\Delta_n = \partial_{n+1} \circ \partial_{n+1}^* + \partial_n^* \circ \partial_n$$

Or its persistent version for nested complexes $K \hookrightarrow L$.

Computational complexity is (Mémoli et al., 2012):

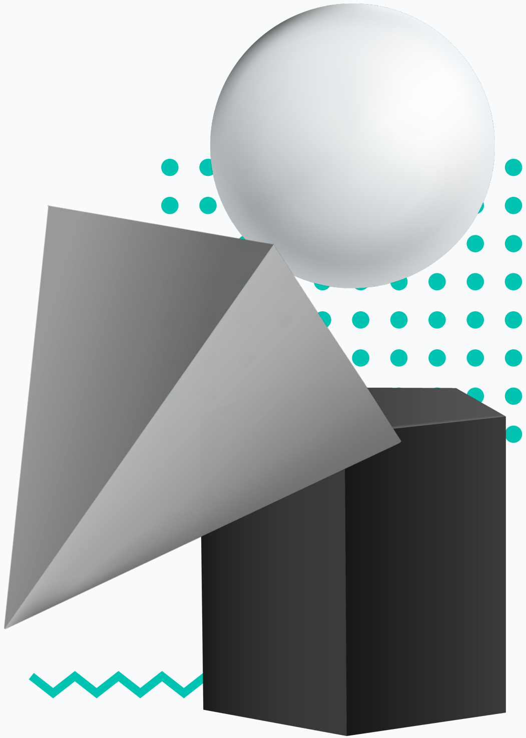
$$O(|K_n|^2 + (|L_n| - |K_n|)^3 + |K_n|(|L_n| - |K_n|)^2 + |L_{n+1}|)$$

Heat equation: $$\frac{\partial w(\sigma \in K_n)}{\partial t} - \Delta_n w(\sigma) + f(\sigma) = 0$$

$$(f(\sigma) = w_0(\sigma), \text{ or } f(\sigma) = 0)$$

Euler's scheme: $$w(\sigma) \leftarrow w(\sigma) + \eta(\Delta_n w(\sigma) - f(\sigma))$$

$$\frac{\partial w_\infty(\sigma)}{\partial w_0(\sigma')} = (\deg_n \sigma - (id - \pi_\sigma) \circ \Delta_n)^{-1} \frac{\partial f(\sigma)}{\partial w_0(\sigma')}$$



04 Persistency Encoding

why

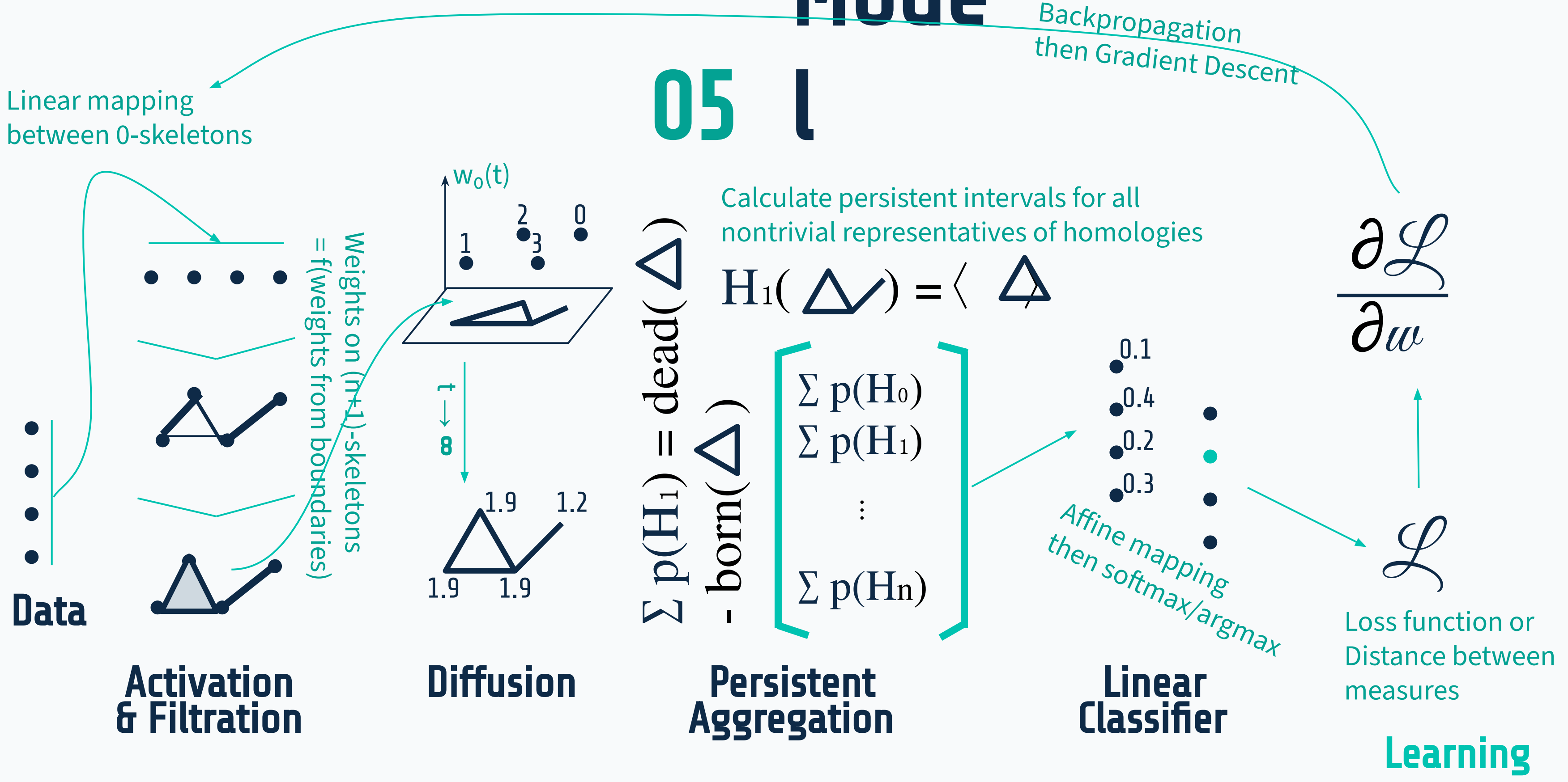
To transform the complex into its representation in R^n based on homology invariants

how

- Define a filtration on 0-, 1-skeletons based on weights (simplex is in subcomplex if its weight $w_{ij} > \theta$, parameter)
- For each clique and non-trivial cycle (being not a boundary) define ‘bf’ and ‘df’ (birth and death) θ-s for 0-, 1-skeleton, correspondingly
- Calculate statistics of [‘bf’, ‘df’] and ‘df’-’bf’ for the sets of cliques and cycles and form a vector

Mode

05 l



Backpropagation
then Gradient Descent
Linear mapping
between 0-skeletons
Weights on (m+1)-skeletons
= f(weights from boundaries)
$w_0(t)$
1
2
3
0
t → ∞
1.9
1.2
1.9
1.9
Calculate persistent intervals for all
nontrivial representatives of homologies
$H_1(\ ) = \langle\ $
$\sum p(H_1) = \text{dead}(\triangle) - \text{born}(\triangle)$
$\sum p(H_0)$
$\sum p(H_1)$
⋮
$\sum p(H_n)$
0.1
0.4
0.2
0.3
Affine mapping
then softmax/argmax
$\frac{\partial \mathcal{L}}{\partial w}$
$\mathcal{L}$
Loss function or
Distance between
measures
Data
Activation
& Filtration
Diffusion
Persistent
Aggregation
Linear
Classifier
Learning

06 Perspectives

Learnable Simplicial Complex

Algebraic Structure of Data

One can treat each data point separately, even assume that each value belongs to its own group

- Graph-product of groups
- Coxeter group if $\mathbb{Z}/2\mathbb{Z}$

Dynamics in Hierarchy modelling

Given the complex, one can vary chain complex, e.g. $C(K, \mathbb{Z}/2\mathbb{Z})$ leads to sign-less boundaries, difference is replaced with summation. Besides, one can replace Laplace operator with any other discrete differential operator

Study of Emerging topologies

If maximal complex is general enough, the properties of activated subcomplexes will depend strongly on the dataset. Same time, the generalization ability will depend strongly on properties of maximal complex

Practical Applications

If the underlying complex converges, one can check if there is a structure in it which is not reduced but is not activated by data, and infer unobserved data from it. Besides, any hierarchical data may be modelled using the suggested model