MEGL - SP24

JJ. FB.
CL. DH.

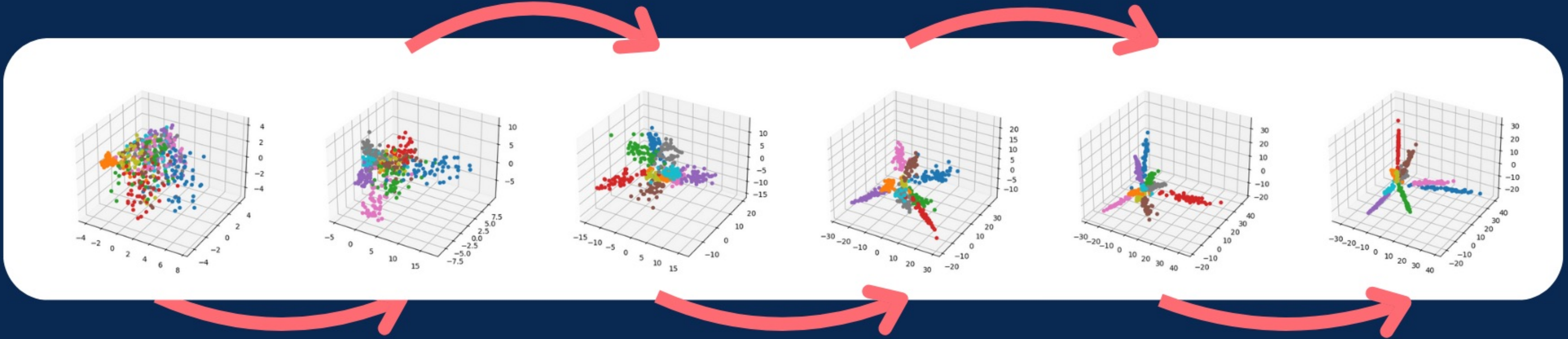
THE TOPOLOGY OF NEURAL NETWORKS

December 6th 2024

UNDER THE DIRECTION OF DR. SCHWEINHART
GRADUATE MENTOR: S POTHAGONI

RESEARCH TOPIC

The premise of our research is to study how neural networks change the data passed through them using tools from topological data analysis.



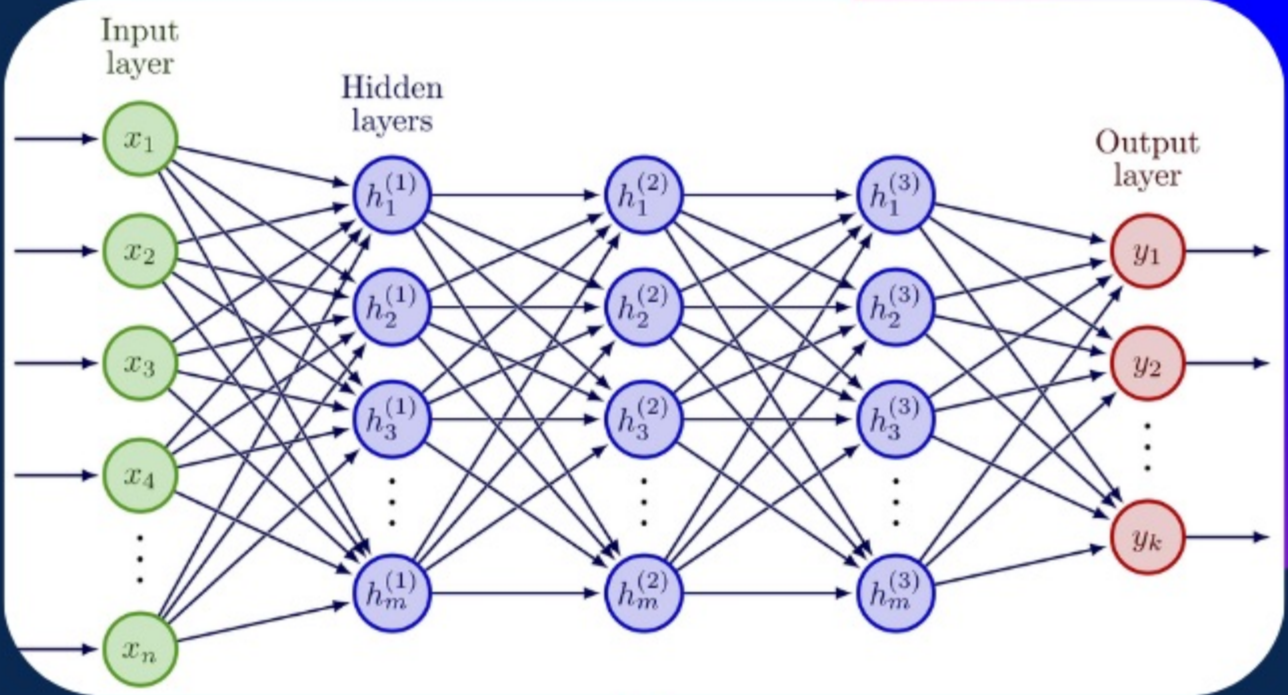
WHAT ARE
NEURAL NETWORKS ?

BUT WHAT IS A NN *MATHEMATICALLY* DOING?

MATHEMATICALLY,

$$g \circ f_n \circ f_{n-1} \circ \cdots \circ f_1(x)$$

- Inputs and outputs are vectors
- Architecture is a composition of affine and non-linear transformations
- g(x) is a regression function used for classification

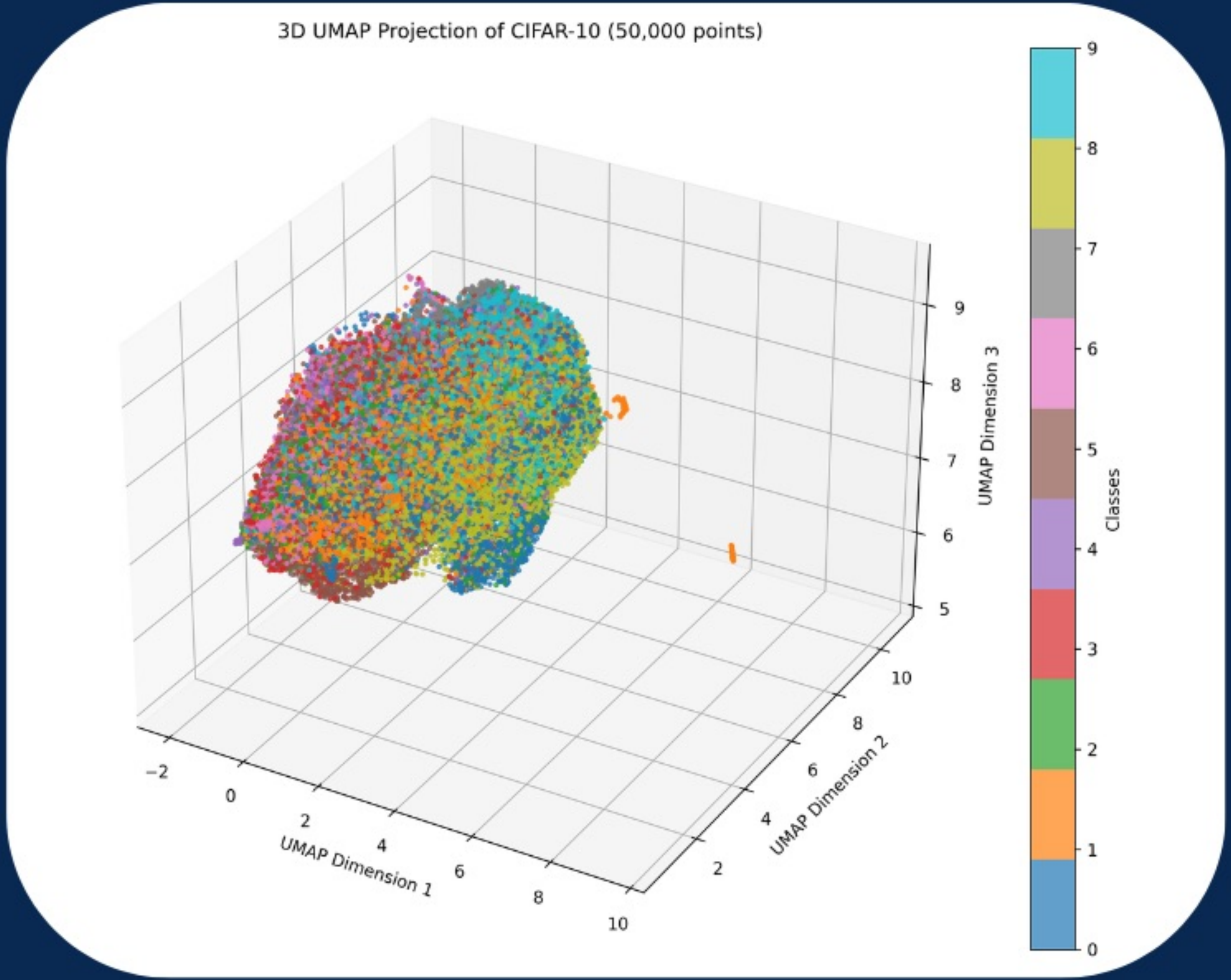


$$a(x) := \frac{1}{1 + e^{-x}} \quad a(x) := \tanh(x) \quad a(x) := \mathbf{max}(0, x)$$

$$f_i(x) = a(W_i x + b_i)$$

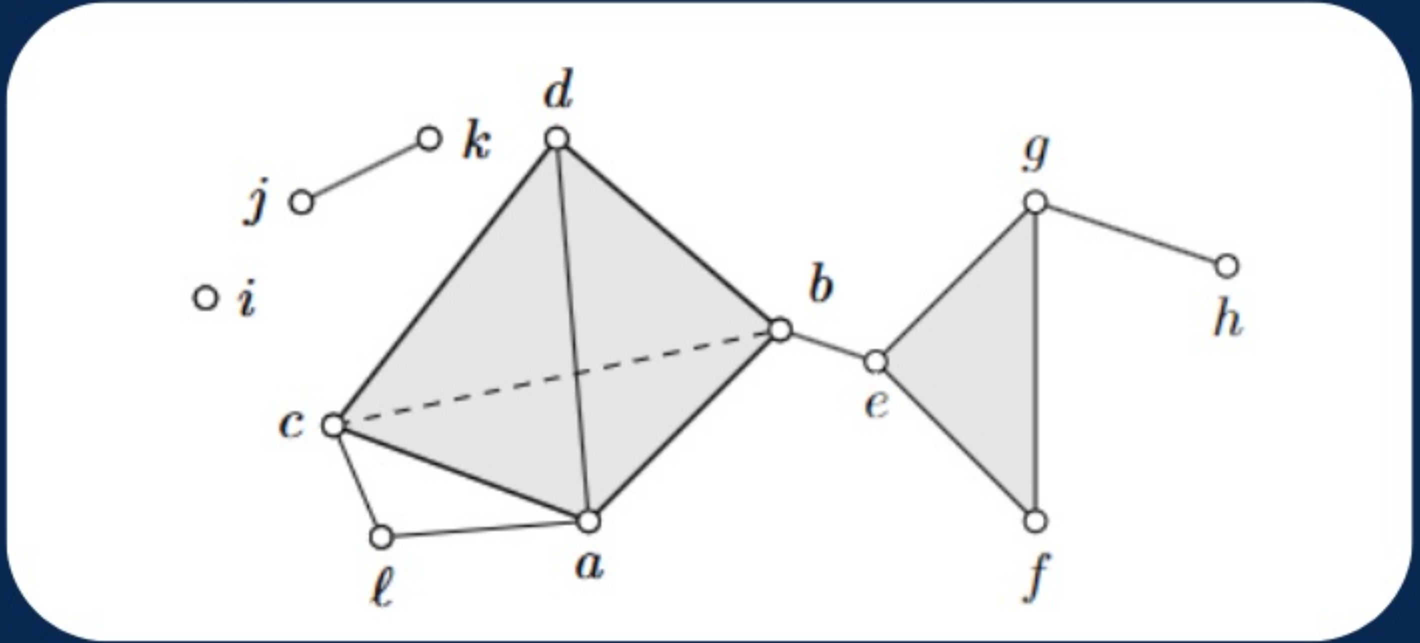
TOPOLOGICAL
DATA ANALYSIS

- “Data has shape,” and shape has meaning. [1]
- Is it possible to measure the complexity of the manifold?



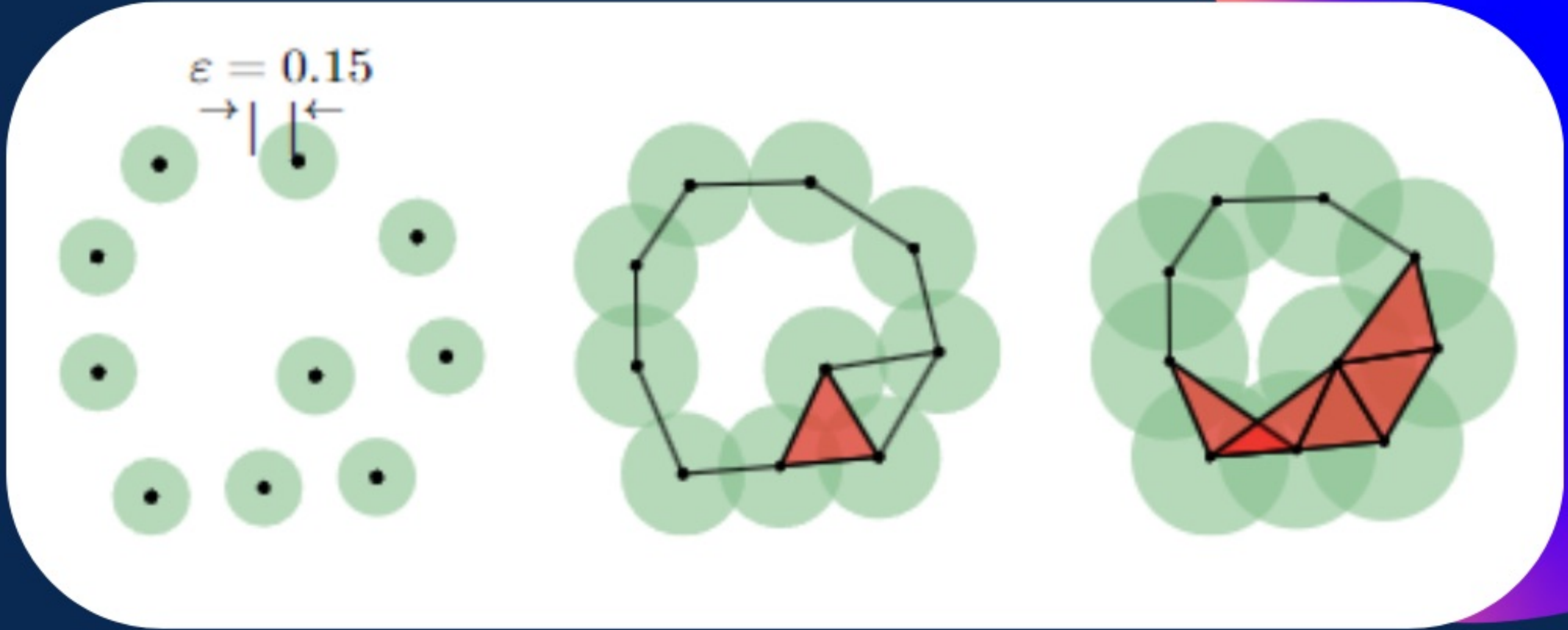
SIMPLICAL
COMPLEX

- How to create a manifold from point cloud datasets?
- Create a set containing the n-dimensional connections between points to triangulate the data manifold



PERSISTENT HOMOLOGY
RIPS COMPLEX

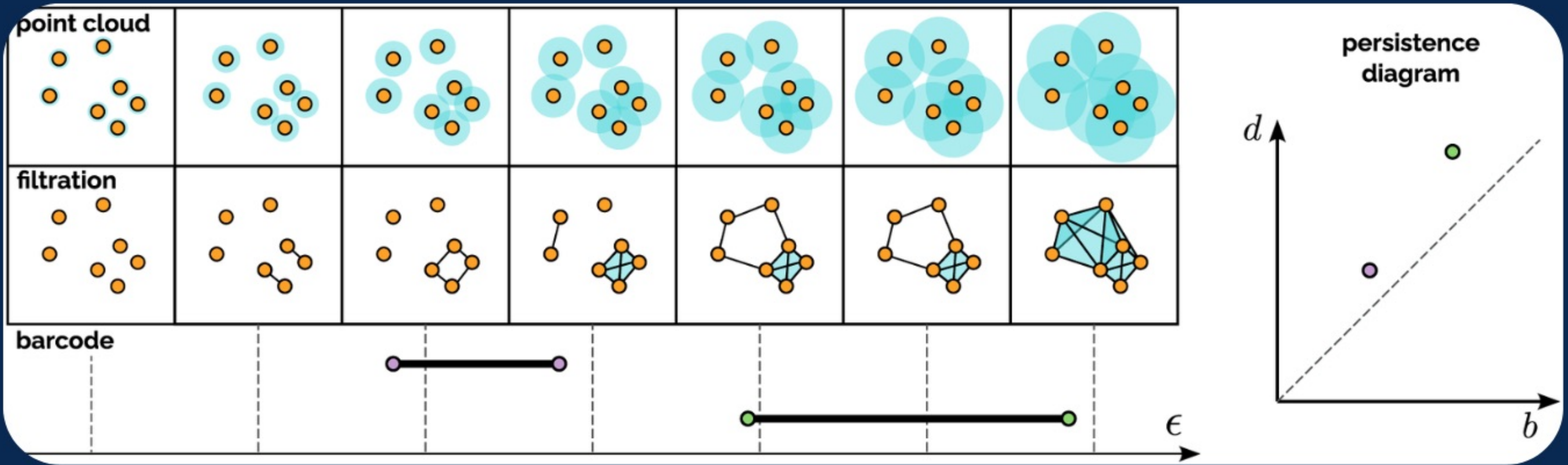
- Allows us to create simplicial complexes from a point cloud with a growing parameter ε.
- As ε grows, the structure of our data transforms and certain features *persists* longer than other features.
  - We call this process *filtration*



PERSISTENT HOMOLOGY

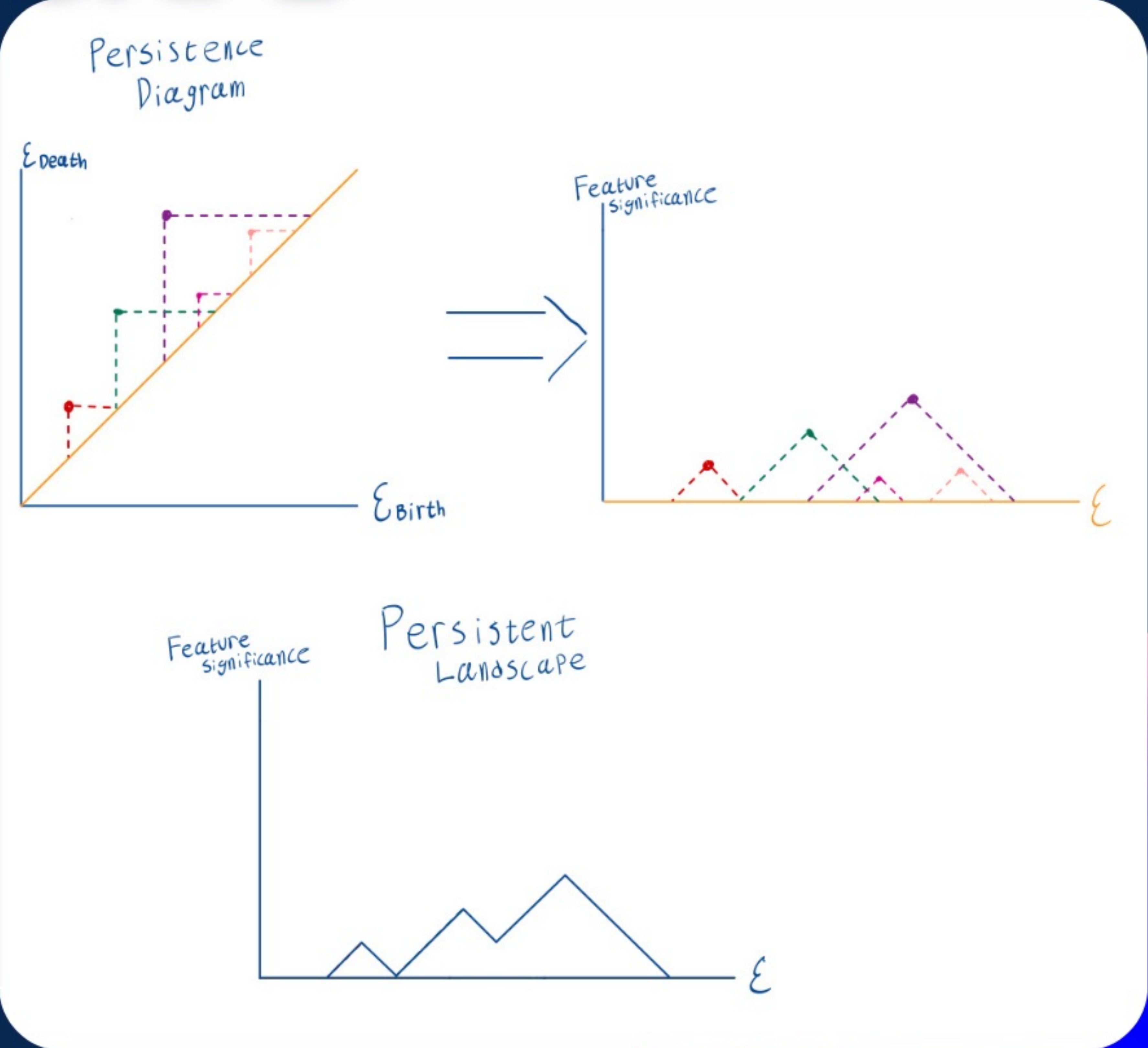
PERSISTENT DIAGRAMS

- Most of the time, we're dealing with hundreds to up to thousands of data points. So we use *Persistent Diagrams* [3] to easily condense, visualize, and analyze our data.
- Maps time of birth and death of topological features
  - Features give insight into shape characteristics of the data



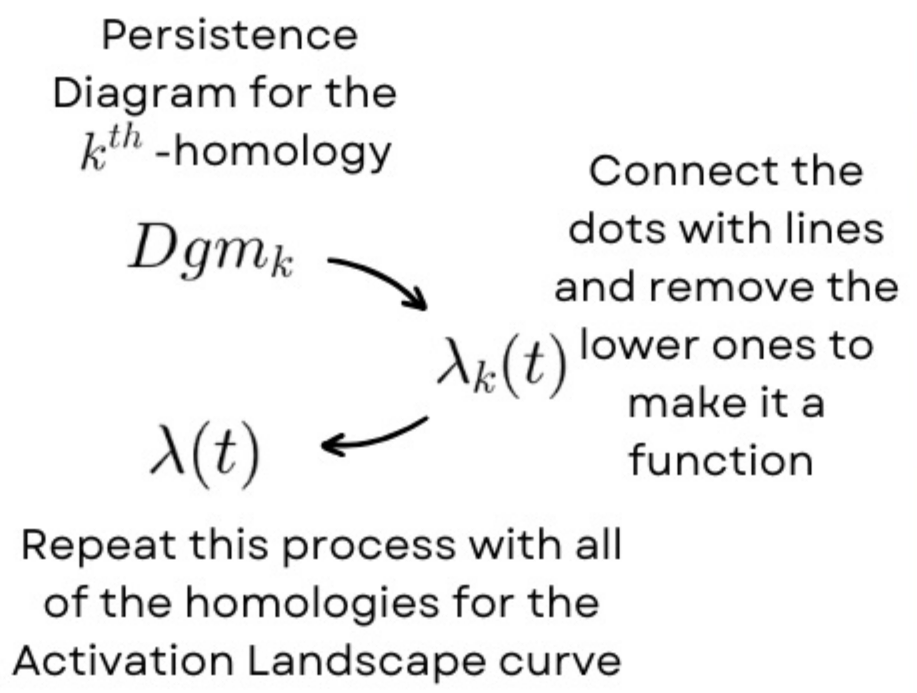
PERSISTENT HOMOLOGY
PERSISTENT LANDSCAPE

- We can then transform our persistent diagram into a *persistent landscape*
- This process also *vectorizes* our data by mapping to a *Function Space.*
- We can now use Linear Algebra to study the topology of the data.



ACTIVATION LANDSCAPES

- By calculating the topological complexity, we can interpret the structure of our data
- To do this, we convert the persistence diagram into a function called an activation landscape
- The norm of this function is the TC [2]



We can define the TC of a dataset as the norm of this function

$$||\lambda(t)||_{L^2}^2 = \int_0^\infty (\lambda(t))^2 dt$$

TOPOLOGICAL COMPLEXITY

- We start by creating a new function which plots the TC of our dataset at each layer then normalizing the domain to be the unit interval (0,1).
- This measures how the topological complexity of the data changes over the whole NN.

Neural Network

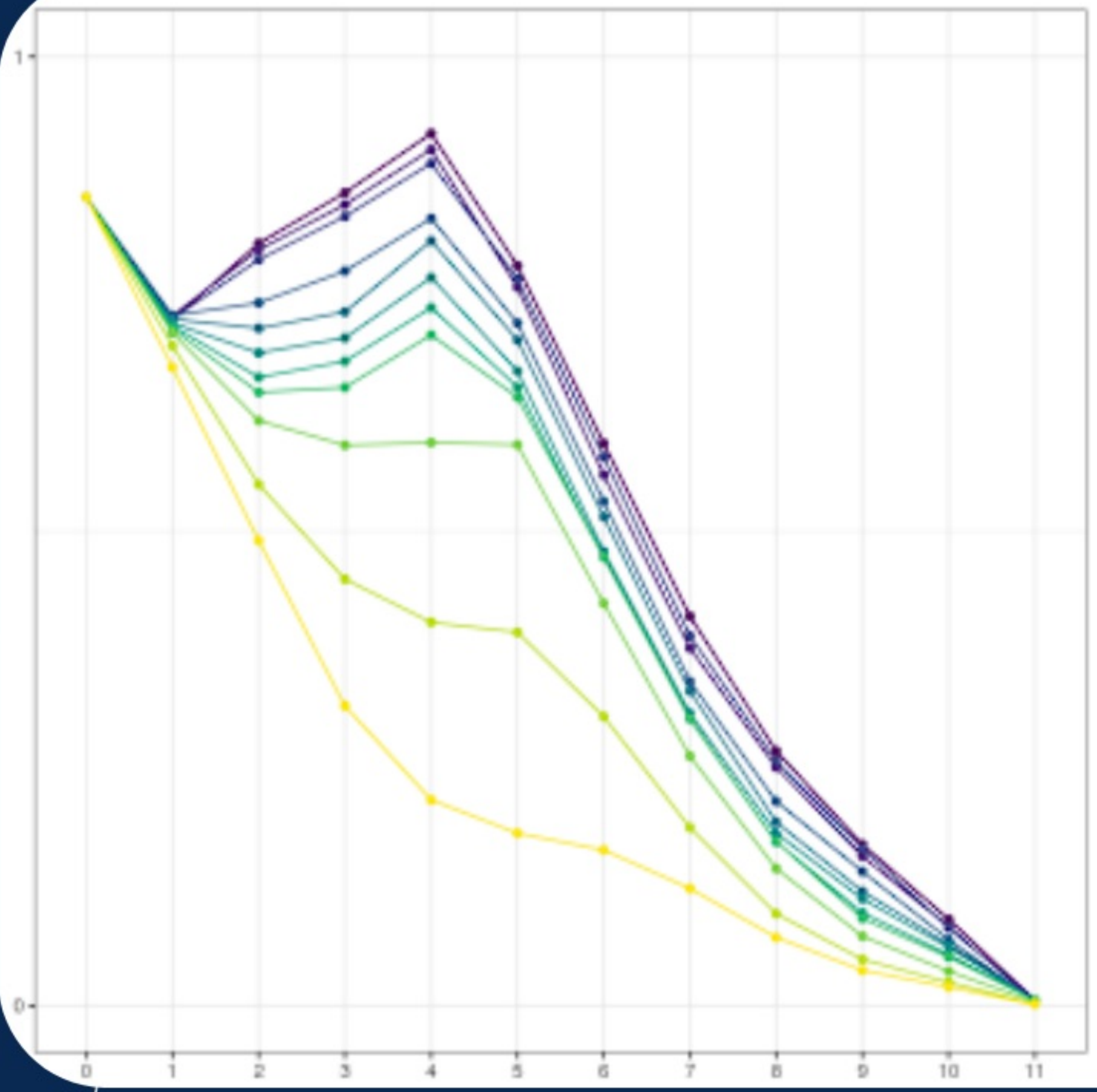
$$g \circ f_k \circ f_{k-1} \circ \cdots \circ f_2 \circ f_1(x)$$

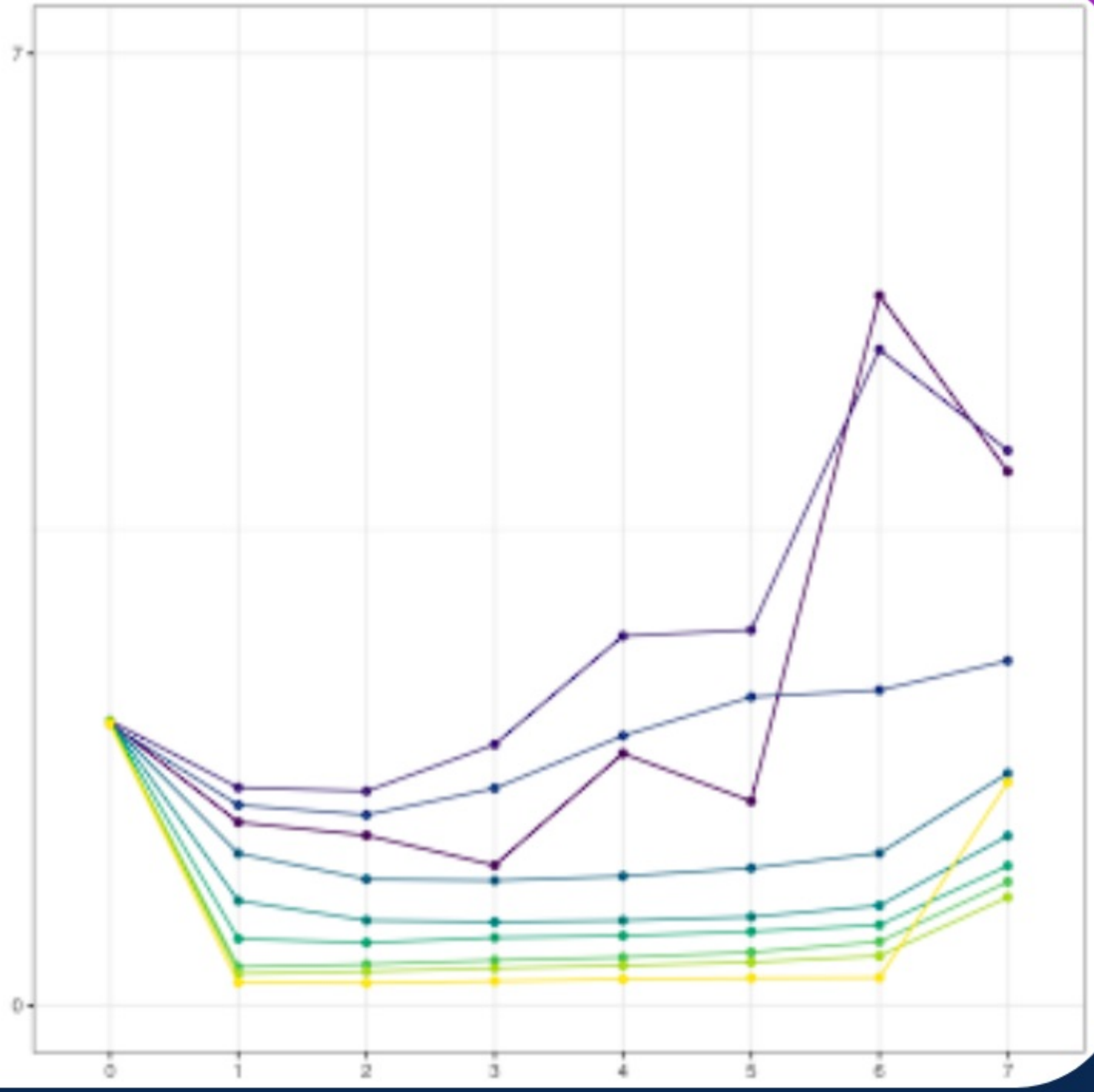
$$TC_i := ||\lambda_i(t)||_{L^2}$$

TC of dataset at the ith layer connected with lines

Norm of this function normalized to the unit interval (0,1)

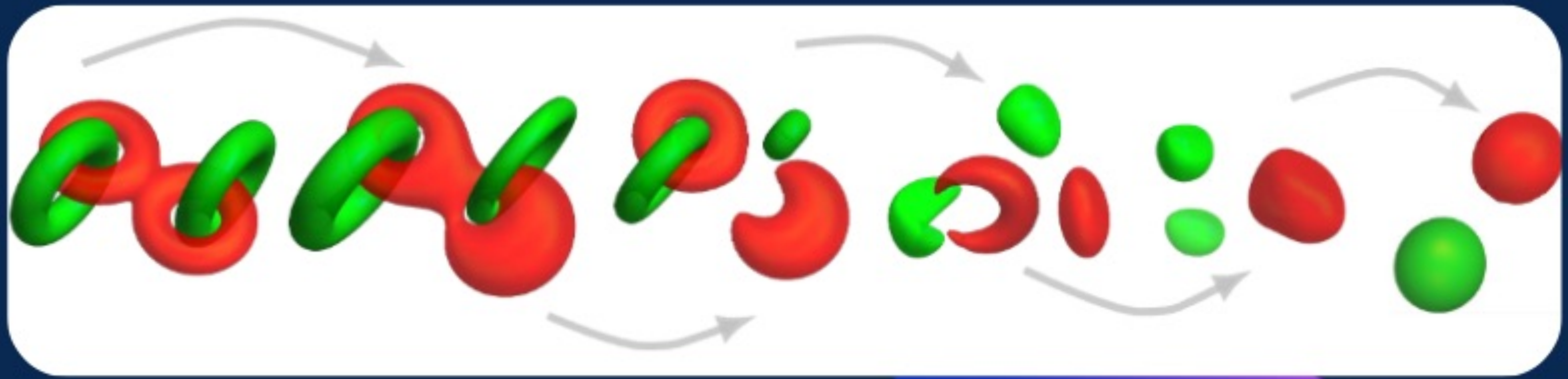
$$||\hat{T}(t)||_{L^2} = \sqrt{\int_0^\infty (\hat{T}(t))^2 dt}$$





AVERAGE NORMS OF THE ACTIVATION LANDSCAPES USING A LARGER NUMBER OF TRAINING ACCURACY THRESHOLDS WITH CONSECUTIVE LAYERS CONNECTED BY LINE SEGMENTS [2]

CURRENT **CONJECTURE**



Let $\mathcal{D} = \{v_i\}_{i=1}^n$ with $v_i \in \mathbb{R}^d$, and let $N(x) = g \circ f_N \circ \cdots \circ f_1(x)$ represent an N-layer fully trained neural network.

Notation: The k-th activation layer of the network is denoted as $N^{(k)}(x) = f_k \circ \cdots \circ f_1(x)$ for $k < N$. Starting from an initialized model $N_0(x)$ at epoch $t = 0$, we say the model is trained when $N_t(x) \to N(x)$ as $t \to T$, where T is the total number of training epochs.

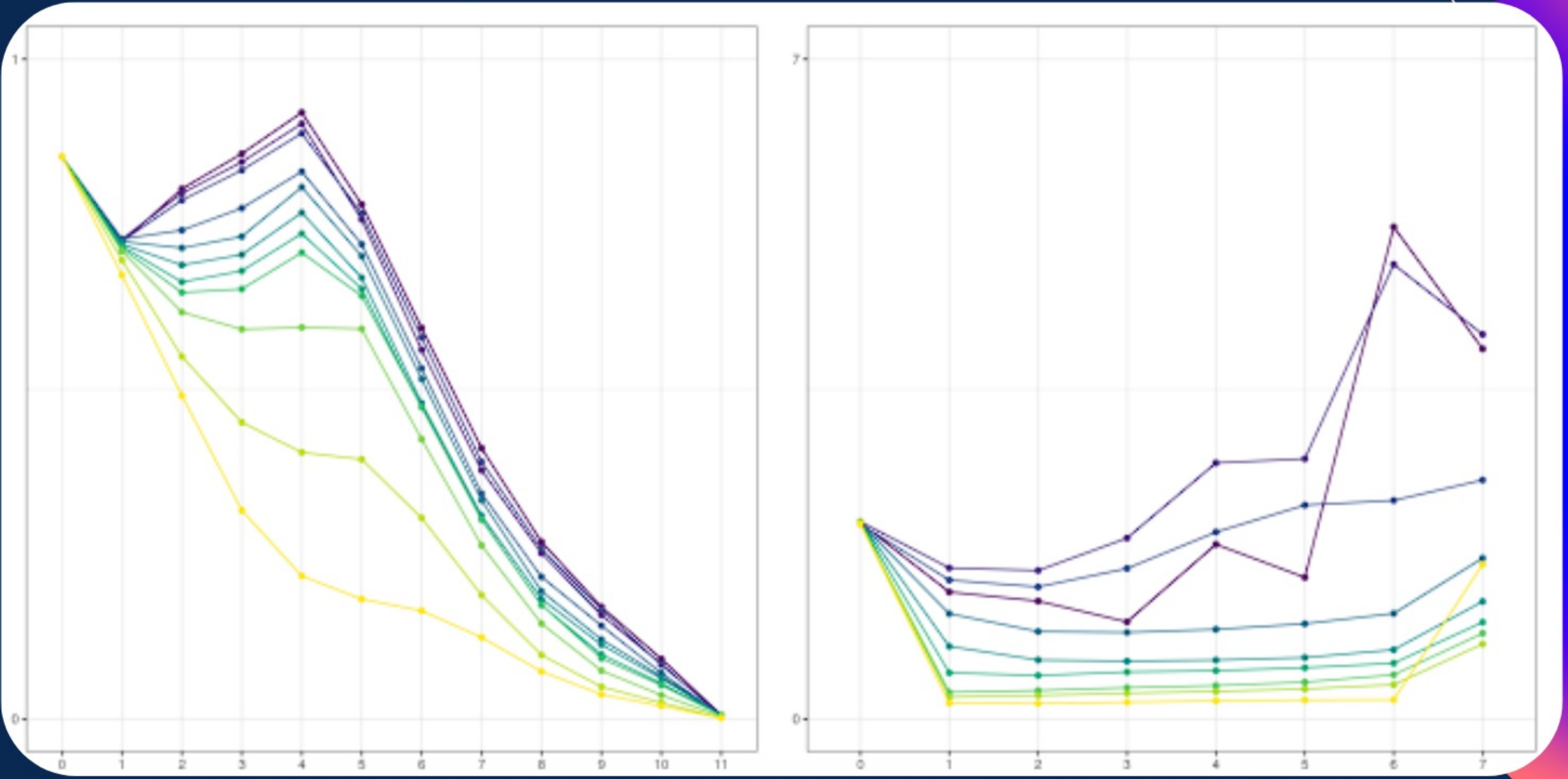
Conjecture: Let $\mathcal{D}_t^{(k)}$ denote the transformed data up to the k-th layer after t training epochs, and let $\mathrm{TC}_t^{(k,p)}$ represent the p-th dimensional topological complexity of $\mathcal{D}_t^{(k)}$. Then for some $K < N$, if $N_t(x) \to N(x)$, there exists a training epoch t_i such that

$$\mathrm{TC}_{t_i}^{(k,0)} > \mathrm{TC}_{t_i}^{(k,p)} \quad \text{for all } p > 0$$

and for some $K \leq k \leq N$.

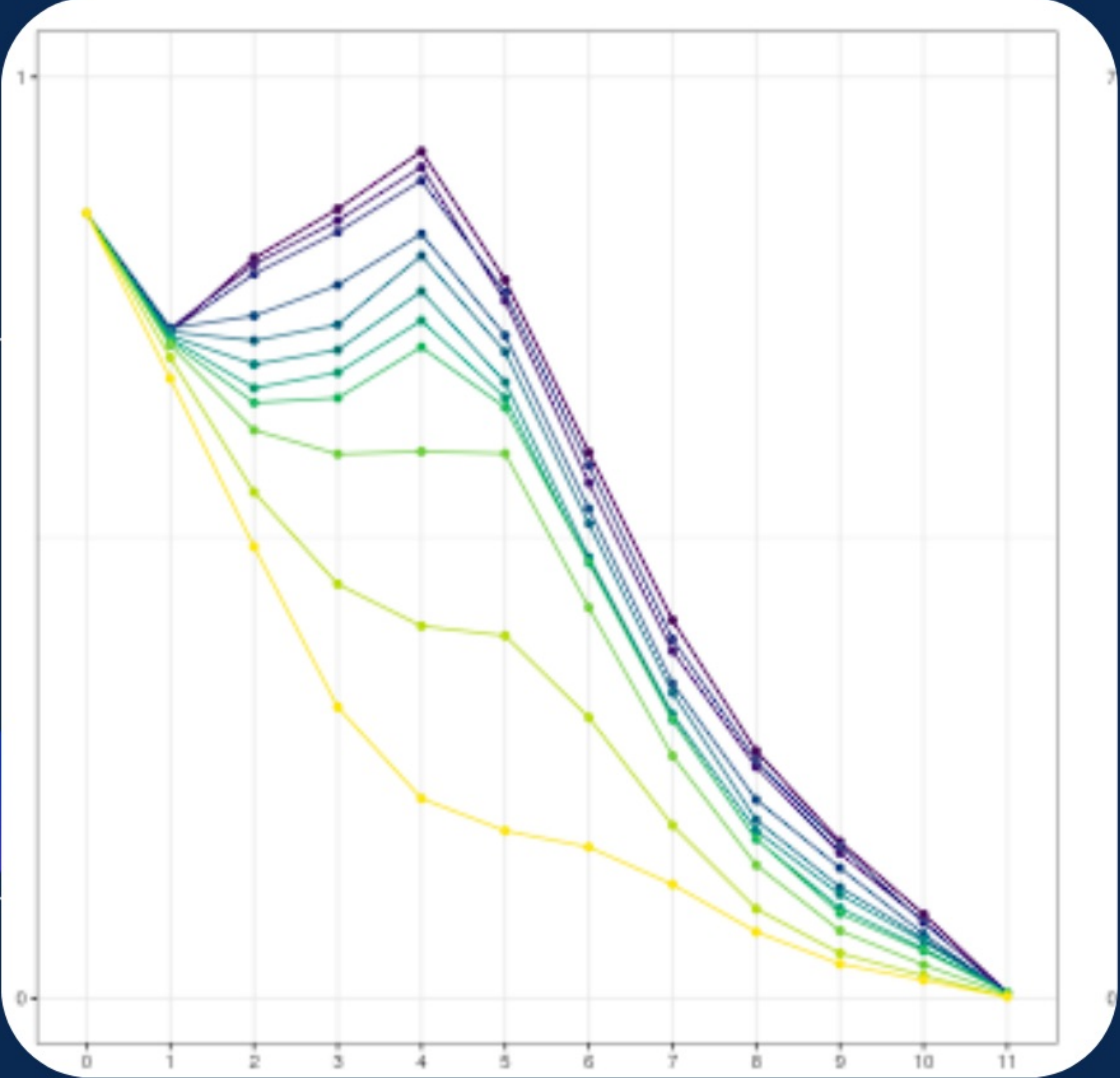
EXPERIMENTS

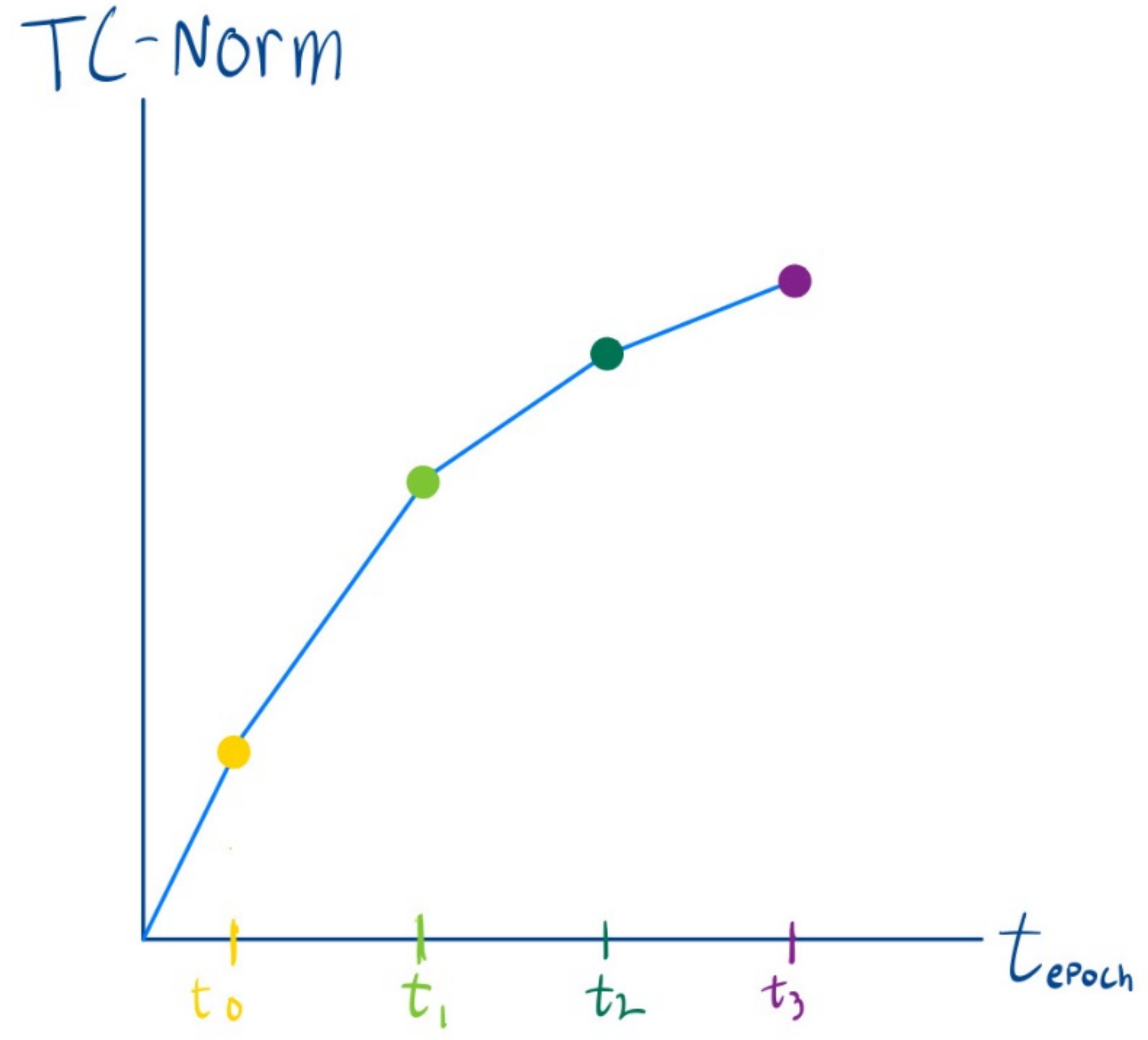
MOTIVATION BEHIND EXPERIMENT



AVERAGE NORMS OF THE ACTIVATION LANDSCAPES USING A LARGER NUMBER OF TRAINING ACCURACY THRESHOLDS WITH CONSECUTIVE LAYERS CONNECTED BY LINE SEGMENTS [2]

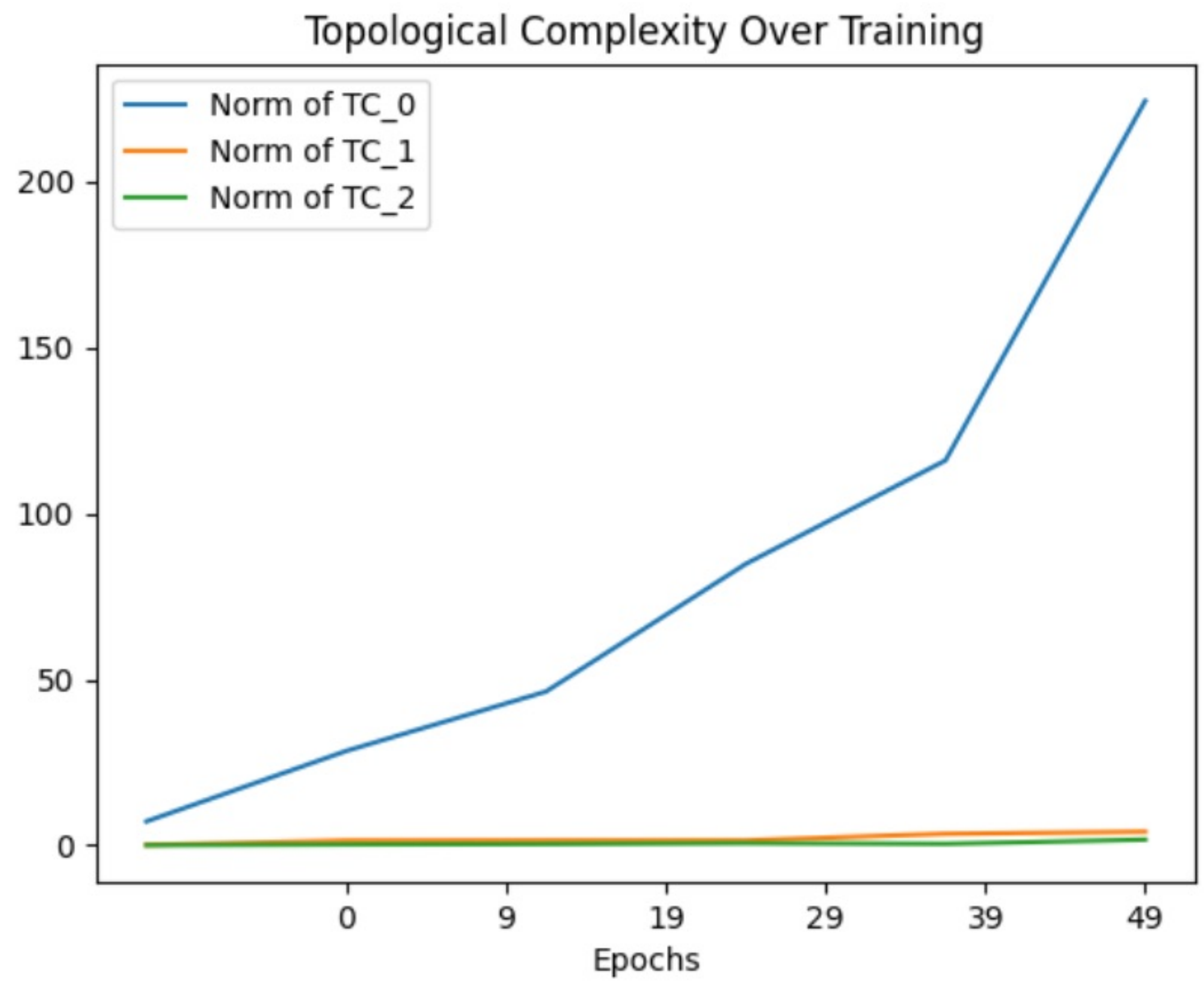
MOTIVATION BEHIND EXPERIMENT





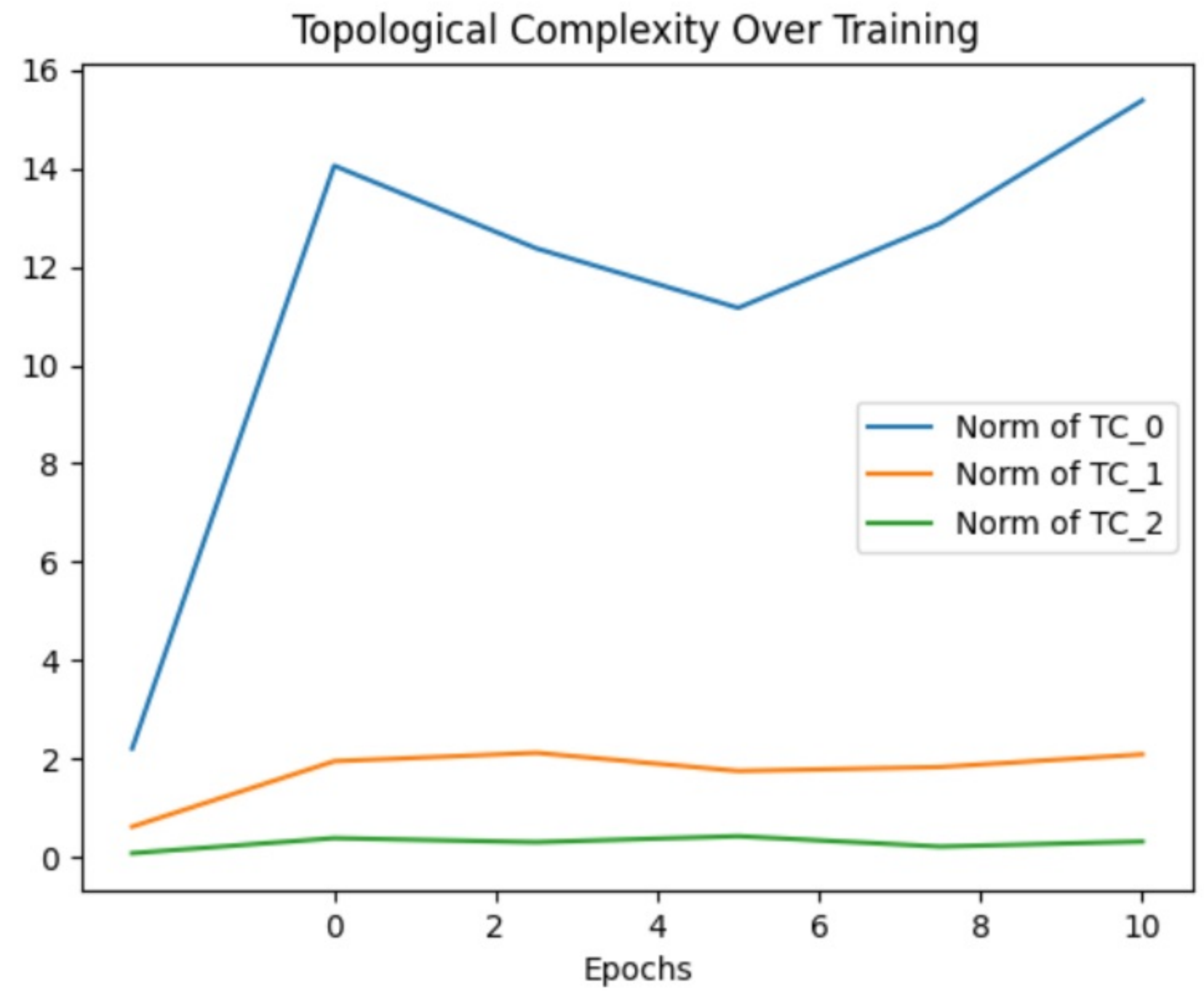
CUSTOM 6-LAYER **MODEL**

- Trained off of the CIFAR-10 Dataset
  - Trained to be 80% accurate
- Computed on the Hopper Compute Cluster
-

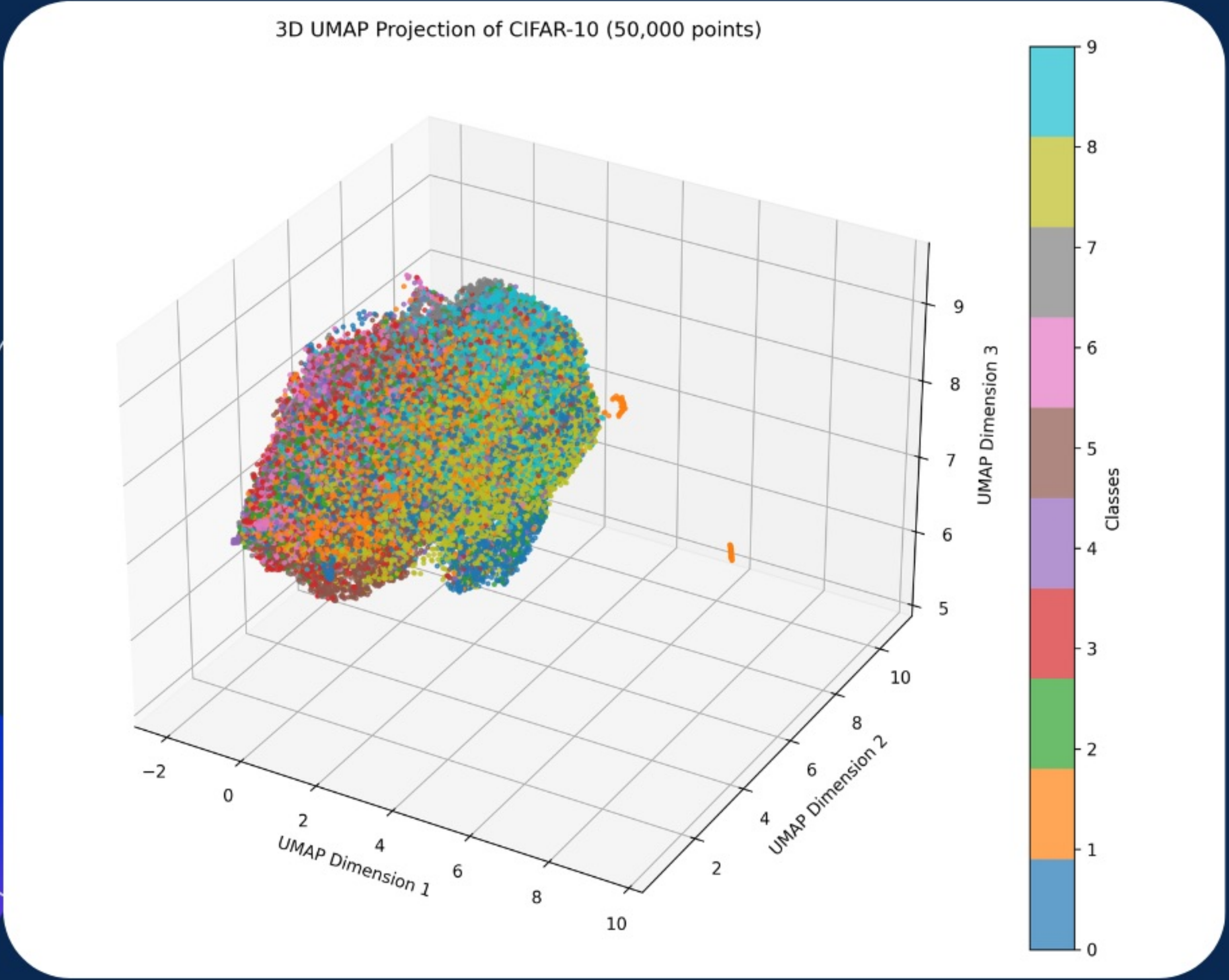


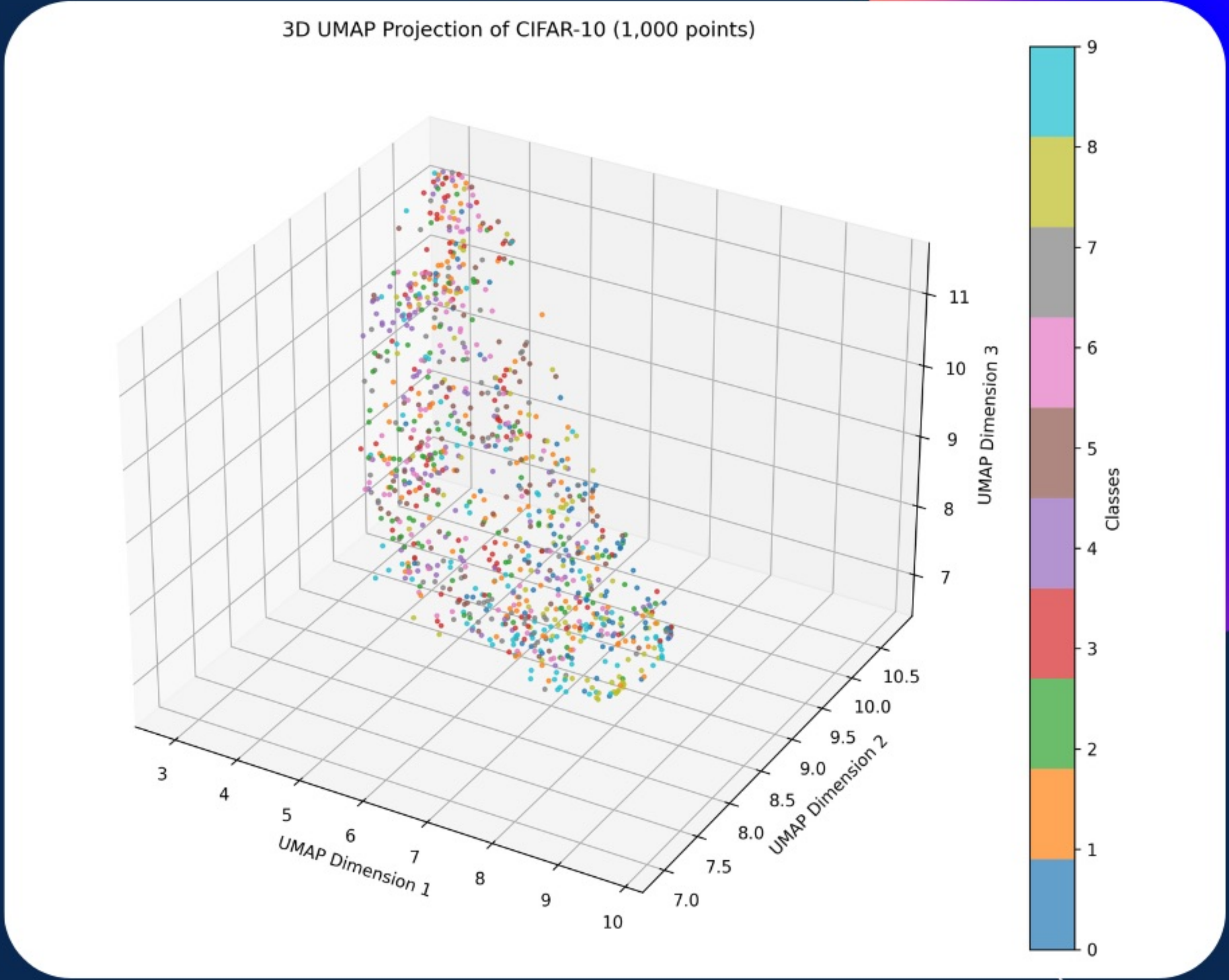
RESNET-NN MODEL

- Trained off of the CIFAR-10 Dataset
  - Trained to be 70% accurate
- Computed on the Hopper Compute Cluster



LIMITATIONS





ALL IN ALL

Quick Recap

1. Training a NN ↔ WHAT IS *REALLY* HAPPENING?
2. Simplicial Complexes
3. Persistence Diagrams
4. Topological Complexity
5. Activation Landscapes
6. Experiments & Results

CITATIONS

1. Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim. Topology of deep neural networks. The Journal of Machine Learning Research, 21(1):7503–7542, 2020.
2. M. Wheeler, J. Bouza and P. Bubenik, "Activation Landscapes as a Topological Summary of Neural Network Performance," in 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021 pp. 3865-3870.
3. Giusti, C., & Lee, D. (2023b). Signatures, Lipschitz-free spaces, and paths of persistence diagrams. SIAM Journal on Applied Algebra and Geometry, 7(4), 828–866. https://doi.org/10.1137/22m1528471

GITHUB

README

Topology of Neural Networks

GitHub respository for the Topology of Neural Networks team at the Mason Experimental Geometry Laboratory.

https://megl.science.gmu.edu/

Abstract

A neural network may be geometrically interpreted as nonlinear function that stretches and pulls apart data between vector spaces. If a dataset has interesting geometric or topological structure, one might ask how the structure of the data will change when passed through a neural network. This is achieved by explicitly viewing the dataset as a manifold and observing how the topological complexity (i.e., the sum of the Betti numbers) of the manifold changes as it passes through the activation layers of a neural network. The goal of this project is to study how the topological complexity of the data changes by tuning the hyper-parameters of the network. This enables us to possibly understand the relationship between the structural mechanics of the network and its performance.