

Homological Inference of Embedding Dimensions in Neural Networks

Luciano Melodia

Friedrich-Alexander University

Chair of Computer Science 6
Martensstraße 3, 91058 Erlangen
luciano.melodia@fau.de

January 11, 2021

1 Motivation

- The manifold of data
- The manifold of a neural network

2 Simplicial structures

- Simplicial complexes
- Persistent homology

3 Counting betti numbers

- Experiments with connected abelian Lie group assumption
- Outlook

Manifolds and Lie groups

Smooth manifold

Let X be a topological space. A pair (X, \mathcal{A}) , consisting of a second countable Hausdorff space X and a differentiable structure on X given by $\mathcal{A} = (U_i, \varphi_i)_{i \in I}$, the family of pair-wise compatible coordinate charts such that $X = \bigcup_{i \in I} U_i$, is said to be a differentiable manifold. If $\varphi(U) \subseteq \mathbb{R}^n$ for all $(U, \varphi) \in \mathcal{A}$, then we say $\dim X = n$.

Lie group

A Lie group is a smooth manifold G equipped with a group structure so that the maps $\mu : (x, y) \mapsto xy$, $G \times G \rightarrow G$ and $\iota : x \mapsto x^{-1}$ are smooth.

The manifold of data

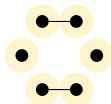
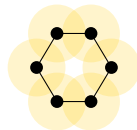
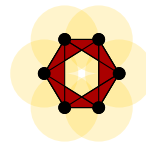
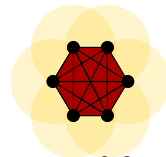
Manifold assumption

A set of points in Euclidean space \mathbb{R}^d is underlain by a manifold, which can be embedded in \mathbb{R}^d . Euclidean space and on which the points lie.

Decomposition of connected abelian Lie groups

Every connected commutative (abelian) Lie group G is isomorphic to a product space $\mathbb{R}^p \times \mathbb{T}^q \cong G$ with $p + q = \dim G$.

The manifold of data


 $r = 0$

 $r = 0.2$

 $r = 0.4$

 $r = 0.6$

 $r = 0.8$

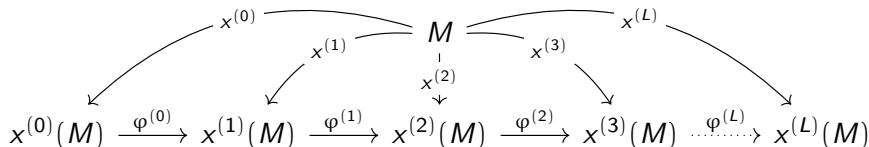
Idea:

- 1 Estimate a suitable structure.
- 2 Estimate one of its invariants.
- 3 Relate the invariant to the dimension.

How many dimensions do I need to represent the manifold that I suspect underlies the above set of points?

How many and which neurons are needed to represent this manifold?

The manifold of a neural network



Coordinate systems $x^{(l)} := \varphi^{(l-1)} \circ \dots \circ \varphi^{(1)} \circ \varphi^{(0)} \circ x^{(0)}$ of a deep neural network induced by change of coordinate charts $\varphi^{(l)} : x^{(l)}(M) \mapsto (\varphi^{(l)} \circ x^{(l)})(M)$ learned by the neural network acting on the data.

Following Michael Hauser and Asok Ray: Principles of Riemannian Geometry in Neural Networks.

Realise a good representation

Question: How can we adjust the neuromanifold to fit the data manifold?

- 1 Take suitable assumptions on the dataset and its manifold.
- 2 Measure invariants on the filtration of a dataset to get a descriptor.
- 3 Infer possible manifolds from these measured invariants.
- 4 Relate the invariants to the dimension of the manifold.
- 5 Seek for approximate solutions if assumptions may not hold.

Building blocks: simplices

Consider the set of points $X := \{v_0, v_1, \dots, v_n\} \subset \mathbb{R}^d$. They are said to be **affinely independent** if the points $\{v_0 - v_n, v_1 - v_n, \dots, v_{n-1} - v_n\}$ are linearly independent. The **convex closure** of these points is a **simplex** and written as

$$[v_0, v_1, \dots, v_n] = \left\{ \sum_{i=0}^{n-1} \lambda_i (v_i - v_n) \mid \sum_{i=0}^{n-1} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \right\}. \quad (1)$$

The dimension of the simplex is n .

The i -th face of a simplex $[v_0, v_1, \dots, v_n]$ (with an omitted element \hat{v}_i) is defined by

$$d_i[v_0, v_1, \dots, v_n] = [v_0, v_1, \dots, \hat{v}_i, \dots, v_n]. \quad (2)$$

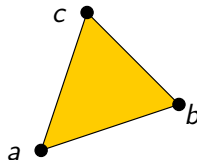
Examples



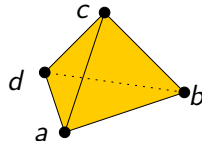
0-simplex
(point)



1-simplex
(line segment)



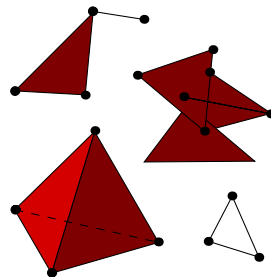
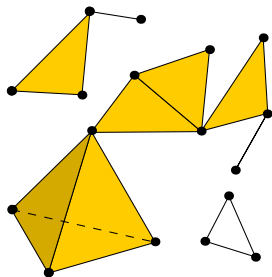
2-simplex
(triangle)



3-simplex
(tetrahedron)

The coefficients λ_i are chosen from the **field** $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$, such that we can **neglect the orientation** of the simplices. This is used for highly efficient computations.

Definition of simplicial complexes



A **simplicial complex** K is a finite union of simplices satisfying that every face of a simplex in K is in K and that the non-empty intersection of two simplices in K is a face of each.

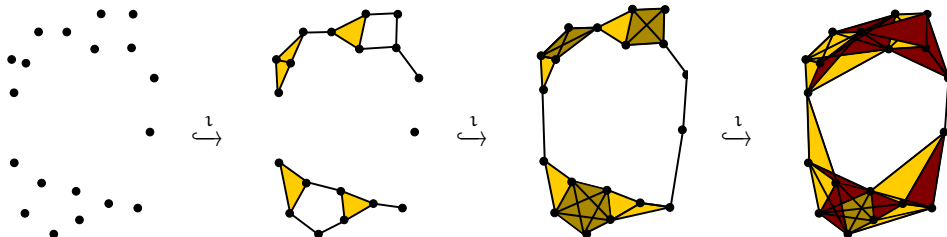
Filtered simplicial complexes

A **filtration** is a nested sequence of complexes K_i , which induce an ordering of the sublevel complexes. These complexes, together with the inclusion $K_i \hookrightarrow K_j$ for $0 \leq i \leq j \leq n$ are called a filtration and denoted by \mathbb{K} :

$$\mathbb{K}: \quad \emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K. \quad (3)$$

The inclusion on the filtration induces a homomorphism of groups $f_k^{ij} : H_k(K_i) \rightarrow H_k(K_j)$, in this case H_k are the k -th homology groups.

Filtered simplicial complexes



Vastly used complexes to create a filtration on a point set X are:

Čech complex: $(v_0, v_1, \dots, v_n) \in \check{\text{Cech}}_r(X) \iff \bigcap_{i=0}^n B_r(x_i) \neq \emptyset.$

Vietoris-Rips complex: $(v_0, v_1, \dots, v_n) \in \text{Rips}_r(X) \iff \|v_i - v_j\| \leq r.$

Isomorphism of homology theories

Isomorphic homology theories

For a trianguliable smooth manifold X and a simplicial complex K , forming its triangulation, the following holds for the homology groups from a field of coefficients \mathbb{F} :

$$H_k(K; \mathbb{F}) \cong H_k(X; \mathbb{F}) \cong H_k^\infty(X; \mathbb{F}) \cong H_{\text{deRham}}^k(X; \mathbb{F}). \quad (4)$$

Proof of $H_k(K; \mathbb{F}) \cong H_k(X; \mathbb{F})$: Allen Hatcher: Algebraic Topology.

Proof of $H_k(X; \mathbb{F}) \cong H_k^\infty(X; \mathbb{F})$: John Lee: Introduction to smooth manifolds.

Proof of $H_k^\infty(X; \mathbb{F}) \cong H_{\text{deRham}}^k(X; \mathbb{F})$: John Lee: Introduction to smooth manifolds.

Persistent homology

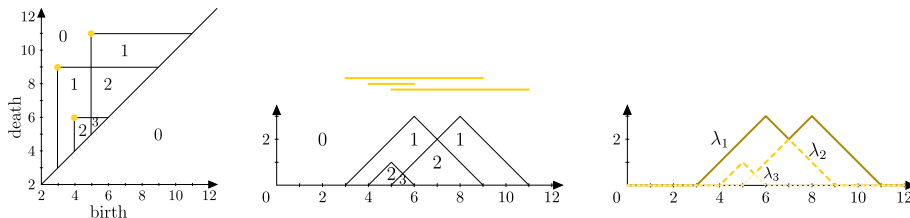
A persistence module is a family of \mathbb{F} -vectorspaces V_i for every real number i together with \mathbb{F} -linear maps $f^{ij} : V_i \rightarrow V_j$. These are called structure maps. For each pair $i \leq j$, satisfying that $k \leq i \leq j$, then $f^{kj} = f^{ij} \circ f^{ki}$.

Let $\{K_i\}_{i \in \mathbb{R}}$ be a set of ordered simplicial complexes together with simplicial maps $f^{ij} : K_i \rightarrow K_j$ for each pair $i \leq j$, such that $k \leq i \leq j$ implies $f^{kj} = f^{ij} \circ f^{ki}$. An example is the aforementioned filtered simplicial complex. Then the persistence module with coefficients in \mathbb{F} is given by

$$H_\star(K_i; \mathbb{F}), \quad H_\star(f^{ij}) : H_\star(K_i; \mathbb{F}) \rightarrow H_\star(K_j; \mathbb{F}). \quad (5)$$

Persistent landscapes

Functional representation of persistence diagrams as a sequence of functions $\lambda_k : \mathbb{R} \rightarrow [-\infty, \infty]$ with $\lambda_k(x)$ being the k th largest value of $\min(x - b_i, d_i - x)$. It is stable with respect to Bottleneck distance and lies in a Banach space.



Taken from Peter Bubenik: Statistical Topological Data Analysis using Persistence Landscapes.

Commutative abelian Lie groups

Main assumption: $H_k(G) \cong H_k(\mathbb{R}^p \times S_1^1 \times \cdots \times S_q^1)$ with $p + q = \dim G$.

Goal: Sufficiently good representation of the topological structure of input space.

Künneth's Theorem

$$H_k(X \times Y) \cong \bigoplus_{i+j=k} H_i(X) \otimes H_j(Y), \text{ thus}$$

$$H_k(\mathbb{R}^p \times S_1^1 \times \cdots \times S_q^1) \cong \bigoplus_{i_1 + \cdots + i_r = k} H_{i_1}(\mathbb{R}^p) \otimes H_{i_2}(S_1^1) \otimes \cdots \otimes H_{i_r}(S_q^1).$$

Computing dimensions

We get

$$H_0(S^1) = H_1(S^1) = \mathbb{Z}, \quad (6)$$

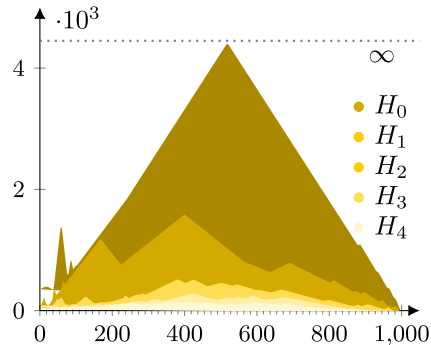
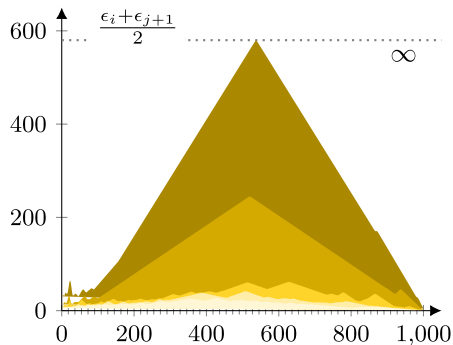
$$H_i(S^1) = 0, \text{ for all } i \geq 2. \quad (7)$$

Applying Künneth's formula only indices for $i_j \in \{0, 1\}$ remain, thus we get

$$H_0(\mathbb{R}^p) \cong \mathbb{Z}, \quad (8)$$

$$H_k(\mathbb{T}^q) \cong H_k(S_1^1 \times \cdots \times S_q^1) \cong \mathbb{Z}^{\binom{q}{k}}. \quad (9)$$

Experimental results on cifar10 & cifar100



Results for the Betti numbers

	Homology groups					\approx embedding dimension					
	H_0	H_1	H_2	H_3	H_4	p	$q H_1$	$q H_2$	$q H_3$	$q H_4$	$\dim U$
cifar10	12	16	40	59	50	12	16	9 ± 4	8 ± 3	7 ± 15	92 ± 44
cifar100	13	18	34	46	48	13	18	9 ± 2	8 ± 10	7 ± 13	97 ± 50

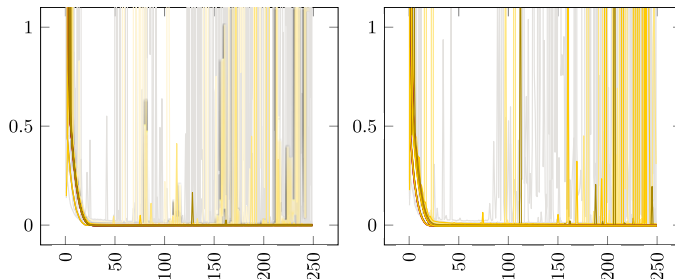
Problem: It becomes apparent, that this can't be a torus. Further, it might be, that the components are not connected. Our theory does give us results for *connected* abelian Lie groups.

Results for the Betti numbers

	Homology groups					\approx embedding dimension					
	H_0	H_1	H_2	H_3	H_4	p	$q H_1$	$q H_2$	$q H_3$	$q H_4$	$\dim U$
cifar10	12	16	40	59	50	12	16	9 ± 4	8 ± 3	7 ± 15	92 ± 44
cifar100	13	18	34	46	48	13	18	9 ± 2	8 ± 10	7 ± 13	97 ± 50

Workaround: Choose a dimension being capable of embedding n -tori according to the rank of *every single* persistent homology group.

Losses on cifar10 & cifar100







cifar10: $\bullet \in [2, 148]$, $\bullet \in [150, 198]$, $\bullet \in [200, 270]$ and $\bullet \in [272, 784]$, (10)

cifar100: $\bullet \in [2, 148]$, $\bullet \in [150, 198]$, $\bullet \in [200, 292]$ and $\bullet \in [294, 784]$. (11)

Outlook



- 1 **Sliding window embeddings** – investigated by Jose Perea et al. [PersHS2016, SliWi2015] – embed time series as a curve on or a curve dense on a torus. For these embeddings our method is an accurate estimate.
- 2 For arbitrary datasets we do not yield an exact solution for the binomial coefficient. The assumption of a **connected structure** fails. How could one generalize this?
- 3 We use vanilla neural networks. If we know the homology groups we want to represent, we can assign a commutative Lie group structure to the neural network itself, so that the **neuromanifold** has the same invariants as the **data manifold**.

References

-  Michael Hauser and Asok Ray (2017)
Principles of Riemannian Geometry in Neural Networks
Advances in Neural Information Processing Systems 30.
-  Jose Perea (2016)
Persistent Homology of Toroidal Sliding Window Embeddings
IEEE International Conference on Acoustics, Speech and Signal Processing.
-  Jose Perea, John Harer (2015)
Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis
Foundations of Computational Mathematics.
-  Peter Bubenik (2015)
Statistical Topological Data Analysis using Persistence Landscapes
Journal of Machine Learning Research.

Thank you.

Got interested?

Drop a line to `luciano.melodia@fau.de` 
or follow karhunenloeve on GitHub .