Topological Data Analysis and Persistent Homology

Josh Ibrahim

Mentored by Miguel Lopez

Introduction

Topological Data Analysis (TDA)

- Data analysis method designed to mathematically describe the shape of data

Persistent Homology

- The main tool used in TDA
- The shape can be quantified using persistence diagrams (barcodes)

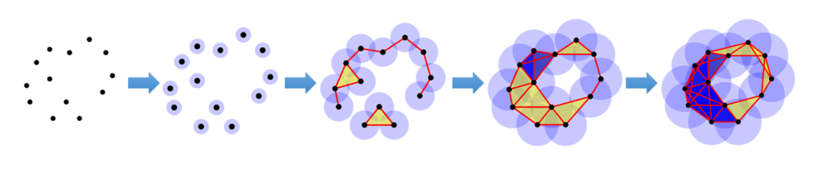

Figure 1: The illustration of a filtration process using persistent homology [1].

Simplicial Complexes

Due to simultaneous relationships between points within a point cloud, **simplicies** are used to give the point cloud a topology.

Definition

An n-**simplex** is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^n$.

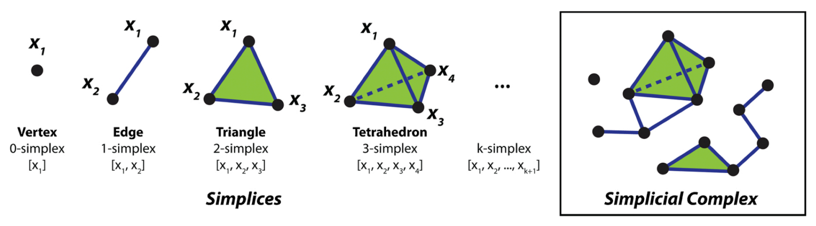

Figure 2: From simplicies to simplicial complexes [2].

Filtrations

Construction of a simplicial complex is done by a **filtration**

Definition

Let S be a simplicial complex; a **filtration** of S (of length n) is a nested sequence of subcomplexes of the form

$$\mathrm{F_1}S \subset \mathrm{F_2}S \subset \ldots \subset \mathrm{F_{n-1}}S \subset \mathrm{F_n}S = S$$

with inclusion maps $g_i : \mathrm{F_i}S \to \mathrm{F_{i+1}}S$

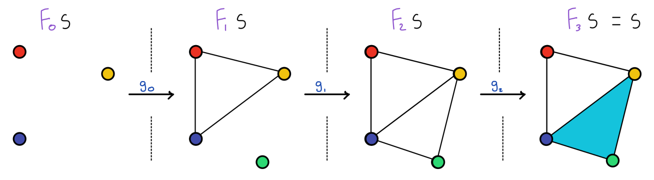

Figure 3: Illustration of a filtration and inclusion maps.

Vietoris-Rips Filtration

Definition

Let (M, d) be a finite metric space. The **Vietoris-Rips filtration** of M is an increasing family of simplicial complexes $\mathsf{VR}_\epsilon(\mathsf{M})$ such that a subset $x_0, x_1, \ldots, x_k \subset M$ forms a k-dimensional simplex in $\mathsf{VR}_\epsilon(\mathsf{M})$ if and only if the pairwise distances satisfy $d(x_i, x_j) \leq \epsilon$ for all i, j.

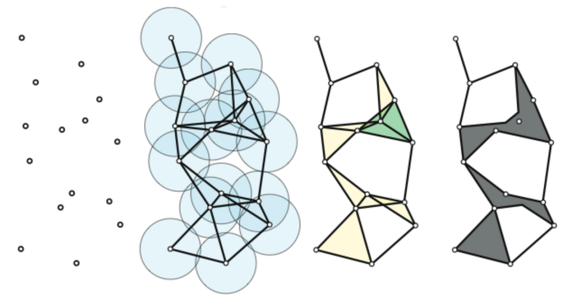

Figure 4: Simplicial complex built from a Vietoris-Rips filtration [3].

Chains and Boundary Maps

A filtration is only as good as the best chosen ϵ. Instead of searching for the optimal ϵ, we instead look at a continuum of VR_{ϵ_i} for all $(\epsilon_i)_1^N$.

Definition

For each dimension $k \geq 0$, the k-th **chain group** of S is the vector space $\mathbf{C_k}$ over $\mathbb{F}$ generated by treating the k-simplicies of S as a basis.

There are maps $\partial_k : C_k \to C_{k-1}$ called **boundary maps**.

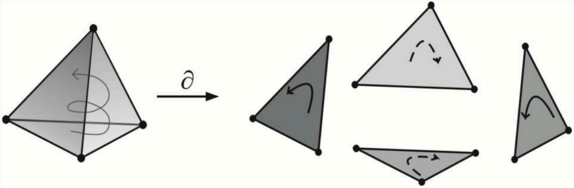

Figure 5: ∂ maps a 3-simplex to four 2-simplices [4].

Chain Complexes and Cycles

Definition

The boundary maps give us a **chain complex** defined as

$$\cdots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} C_{k-2} \xrightarrow{\partial_{k-2}} \cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

with $\partial_{k-1} \circ \partial_k = 0$

The elements of $Im(\partial_{k+1})$ in C_k, are called boundaries ($\mathsf{B_k}$). Elements of C_k that are mapped to zero in C_{k-1} by ∂_k are called **cycles**.

Definition

Let $\mathbf{Z_k}$ be the cycles of C_k, then

$$\mathsf{Z_k} = Ker(\partial_k)$$

Homology and Betti Numbers

Definition

Let S be a simplicial complex. Then, the k-th **homology group** of S is

$$\mathsf{H_k}S = Ker(\partial_k)/Im(\partial_{k+1}) = \mathsf{Z_k}/\mathsf{B_k}$$

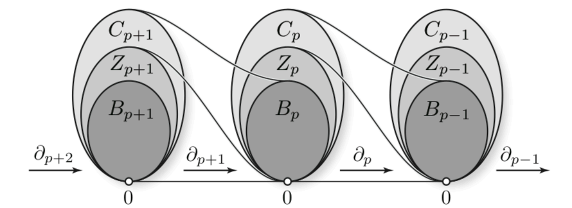

Figure 6: Relationship between boundaries, cycles, and homology groups [5].

The dimension of $\mathsf{H_k}S$ is the k-th **Betti Number** β_k.

$$dim(\mathsf{H_k}S) = \beta_k = dim(\mathsf{Z_k}) - dim(\mathsf{B_k})$$

Persistent Homology

Notation

Let S^l be a filtration of a simplicial complex S at l, and $\mathsf{Z}_\mathsf{k}^\mathsf{l} = Z_k(S^l)$ and $\mathsf{B}_\mathsf{k}^\mathsf{l} = B_k(S^l)$ be the k-th cycle and boundary group of S^l, respectively. The k-th homology group of S^l is $\mathsf{H}_\mathsf{k}^\mathsf{l} = \mathsf{Z}_\mathsf{k}^\mathsf{l}/\mathsf{B}_\mathsf{k}^\mathsf{l}$. The k-th Betti number $\boldsymbol{\beta}_k^l$ of S^l is $dim(\mathsf{H}_\mathsf{k}^\mathsf{l})$.

Definition

The p-**persistent** k-th **homology group** of the l-th complex S^l in filtration is

$$\mathsf{H}_\mathsf{k}^{\mathsf{l,p}} = \mathsf{Z}_\mathsf{k}^\mathsf{l}/(\mathsf{B}_\mathsf{k}^{\mathsf{l+p}} \cap \mathsf{Z}_\mathsf{k}^\mathsf{l})$$

The p-**persistent** k-th **Betti number** $\boldsymbol{\beta}_k^{l+p}$ is

$$\boldsymbol{\beta}_k^{l+p} = dim(\mathsf{H}_\mathsf{k}^{\mathsf{l,p}})$$

Thus, $\boldsymbol{\beta}_k^{l+p}$ counts the homological classes in the complex S^p that were created during filtration in the complex S^l or earlier.

Everything Together

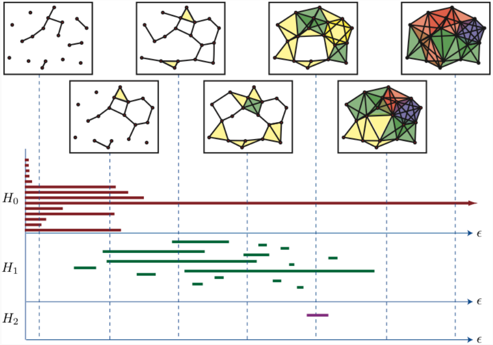

Figure 7: An example of barcodes [6].

Barcodes represent the ”life span” of connected components as ϵ increases. The Betti numbers of a certain degree (0, 1, or 2 in this example) at a certain value of ϵ is the number of barcodes at that degree.

Examples

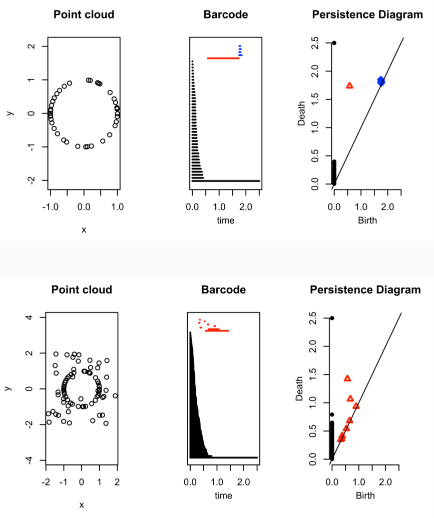

Figure 8: Barcode and persistence diagram of a point cloud of a sample of a circle and noisy circle [7].

References i

[1] Pun, Chi Seng, et al. "Persistent-Homology-Based Machine Learning and Its Applications – A Survey." *ArXiv.org*, 1 Nov. 2018, `https://arxiv.org/abs/1811.00252`.

[2] Zhang, Mengsen, et al. "Topological Portraits of Multiscale Coordination Dynamics." *Journal of Neuroscience Methods*, Elsevier, 6 Mar. 2020, `https://doi.org/10.1016/j.jneumeth.2020.108672`.

[3] Chambers, Erin W., et al. "Rips Complexes of Planar Point Sets." *ArXiv.org*, 3 Dec. 2007, `https://arxiv.org/abs/0712.0395`.

[4] Maletic, Slobodan, et al. "Persistent Topological Features of Dynamical Systems." *ArXiv.org*, 22 Oct. 2015, `https://arxiv.org/abs/1510.06933`.

[5] S. Balakrishnan et al. "Minimax Rates for Homology Inference." *ArXiv.org*, 23 Dec. 2011, `https://arxiv.org/abs/1112.5627`.

[6] Ghrist, Robert. "Barcodes: The Persistent Topology of Data." *American Mathematical Society*, 26 Oct. 2007, `http://www.ams.org/journal-getitem?pii=S0273-0979-07-01191-3`.

References ii

[7] Kraft, Rami. ”Illustrations of Data Analysis Using the Mapper Algorithm and Persistent Homology.” 2016.

[8] Sizemore, Ann E., et al. “The Importance of the Whole: Topological Data Analysis for the Network Neuroscientist.” *ArXiv.org*, 13 June 2018, `https://arxiv.org/abs/1806.05167`.

[9] Nanda, Vidit. “Computational Algebraic Topology.” `https://people.maths.ox.ac.uk/nanda/cat/`.

[10] Ghrist, Robert. ”Elementary Applied Topology”. *Createspace*, 2014, `https://www2.math.upenn.edu/~ghrist/notes.html`.

Questions?