Introduction to Persistent Homology

Sunay Doğan

METU

04.01.2023

Outline

Introduction

Persistent homology, as a tool in topological data analysis (TDA), studies topological features of a sample data set at different scales to understand the topological characteristic of the underlying space.

The key idea is this: Topological aspects of the sample data set that persist as the scale grows can give us an understanding about the topology of population data set.

Simplicial Complexes

Definition

Given $n+1$ affinely independent points in $\mathbb{R}^{n+1}$,
$\Delta^n = \{(t_0, t_1, ..., t_n) \in \mathbb{R}^{n+1} : \sum_{i=0}^{n} t_i = 1, t_i \geq 0\}$ is called a standard n-simplex. Simplices are generalizations of triangles in any dimension.

Definition

Let V be a vertex set such that

- $\forall S \subseteq V$, for any element $\sigma \in S$, all subsets of σ are in S and
- If for any $\sigma, \tau \in S$, then $\sigma \cap \tau \in S$.

Then V is called an abstract simplicial complex. Simplicial complexes are triangluable spaces.

Vietoris-Rips Complexes

Definition

Let X be a metric space and $S \subseteq X$ is finite. The Vietoris-Rips complex, Rips(S, r) is an abstract simplicial complex such that

- S is the vertex set
- $\sigma \subseteq S$ is a simplex if and only if $Diam(\sigma) \leq r$.

Čech Complexes

Definition

Let X be a metric space and $S \subseteq X$ is finite. The Čech complex, $\text{Čech}(S, r)$ is an abstract simplicial complex such that

- S is the vertex set
- $\sigma \subseteq S$ is a simplex if and only if $\bigcap_{x \in \sigma} B(x, r) \neq \emptyset$.

Differences Between Rips and Čech Complexes

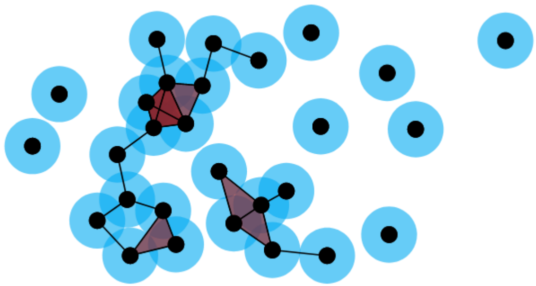


Figure 1: An example of Rips complex [1]

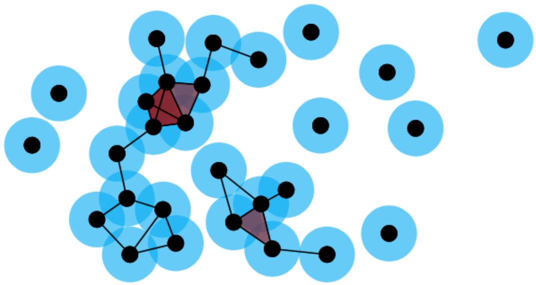


Figure 2: An example of Čech complex [1]

Differences Between Rips and Čech Complexes

- Rips complexes are easier to compute than Čech complexes that is why they are widely used. Yet, geometric interpretation of Čech complexes are well understood.
- If $r_1 \leq r_2$, then
  - $Rips(S, r_1) \subseteq Rips(S, r_2)$
  - $\check{C}ech(S, r_1) \subseteq \check{C}ech(S, r_2)$
  - $\check{C}ech(S, r) \subseteq Rips(S, 2r)$
  - $Rips(S, r) \subseteq \check{C}ech(S, r)$
  - $Rips(S, r\sqrt{2}) \subseteq \check{C}ech(S, r)$ in Euclidean spaces

The Nerve

Definition

Let $\mathcal{U} = \{U_1, U_2, ..., U_k\}$ where $k \in \mathbb{Z}$ be a collection of sets. The nerve of $\mathcal{U}$ ($\mathcal{N}(\mathcal{U})$) is an abstract simplicial complex such that

- $\mathcal{U}$ is the vertex set
- $\sigma \subseteq \mathcal{U}$ is a simplex if and only if $\bigcap_{i \in \sigma} U_i \neq \emptyset$.

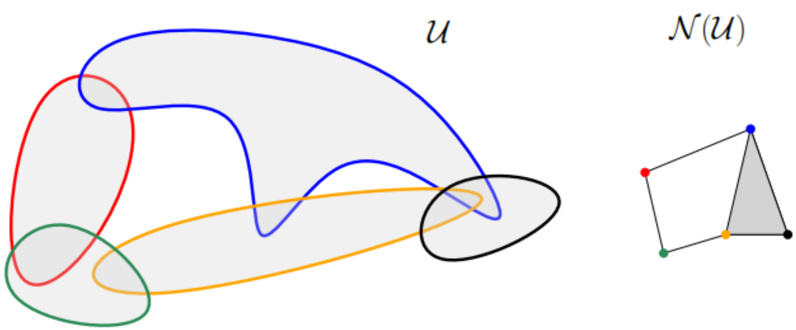


Figure 3: An example of a nerve[3]

The Nerve Theorem

Notice: $\check{C}ech(S,r) = \mathcal{N}(\{B(s,r)\}_{s\in S})$.

A paracompact space X is a space such that every open cover of X has an open subcover which is locally finite.

Theorem (The Nerve Theorem)

If $\mathcal{U}$ is an open cover of a paracompact space X such that every nonempty intersection of finitely many sets in $\mathcal{U}$ is contractible, then X is homotopy equivalent to the nerve $\mathcal{N}(\mathcal{U})$.[2]

Corollary

$$\bigcup_{i=1}^{k} U_i \simeq \mathcal{N}(\mathcal{U})$$

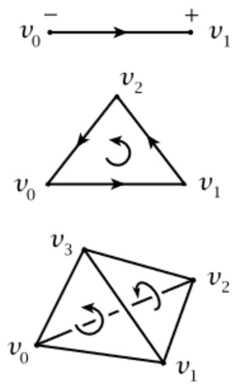
Simplicial k-Chains

Definition

Let V be an abstract simplicial complex and V_k denote the set of all k-simplices of V. Then, $\sum_{i=1}^{N} r_i \sigma_i$ for $\sigma \in V_k$ and $r_i \in \mathbb{Z}$ is called a simplicial k-chain.

The map $\partial_n : V_n \to V_{n-1}$ such that $\partial_n(\sigma) = \sum_i (-1)^i (v_1, ..., \hat{v_i}, ..., v_n)$ where $\sigma = (v_1, ..., v_n)$ is boundary operator. Boundary operators satisfy $\partial_{k-1} \circ \partial_k = 0 \ \forall k$.

Chain Complexes



$$\partial[v_0, v_1] = [v_1] - [v_0]$$

$$\partial[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$$

$$\partial[v_0, v_1, v_2, v_3] = [v_1, v_2, v_3] - [v_0, v_2, v_3] + [v_0, v_1, v_3] - [v_0, v_1, v_2]$$

Figure 4: An example for simplicial k-chains and boundary maps [2]

Definition

The sequence $...V_{k+1} \xrightarrow{\partial_{k+1}} V_k \xrightarrow{\partial_k} V_{k-1}...$ is a chain complex.

Homology

Notice: $Im(\partial_{k+1}) \subseteq Ker(\partial_k)$.

Definition

k^{th} homology of V is defined as $H_k = Z_k/B_{k+1}$ where $Z_k := Ker(\partial_k)$ and $B_{k+1} := Im(\partial_{k+1})$.

If α is a k-chain such that $\partial_k(\alpha) = \emptyset$ then $\alpha \in Z_k$ is called a k-cycle. If there exists a $(k+1)$-chain β such that $\partial_{k+1}(\beta) = \alpha$ then $\alpha \in B_{k+1}$ is called a k-boundary.

Intuitively, k^{th} homology measures k-cycles of space which are not boundary of a region. The k^{th} Betti number $\beta_k = rank(H_k)$ is the number of k dimensional holes of simplicial complex V.

Persistent Homology

Definition

Let X be a finite set of sample data. Our purpose is to construct simplicial complexes, K^δ using elements of X, for any given $\delta \in \mathbb{R}$ such that if $m \leq n$ then $K^m \subseteq K^n$. This ordered set of simplicial complexes is called filtration.

Let S be a finite sample set. Then the Rips filtration on S is the collection of abstract simplicial complexes $\{Rips(S, r)\}_{r \geq 0}$ with inclusion.

Similarly, if S is a finite sample set the Čech filtration on S is the collection of abstract simplicial complexes $\{\check{C}ech(S, r)\}_{r \geq 0}$ with inclusion.

Persistent Homology

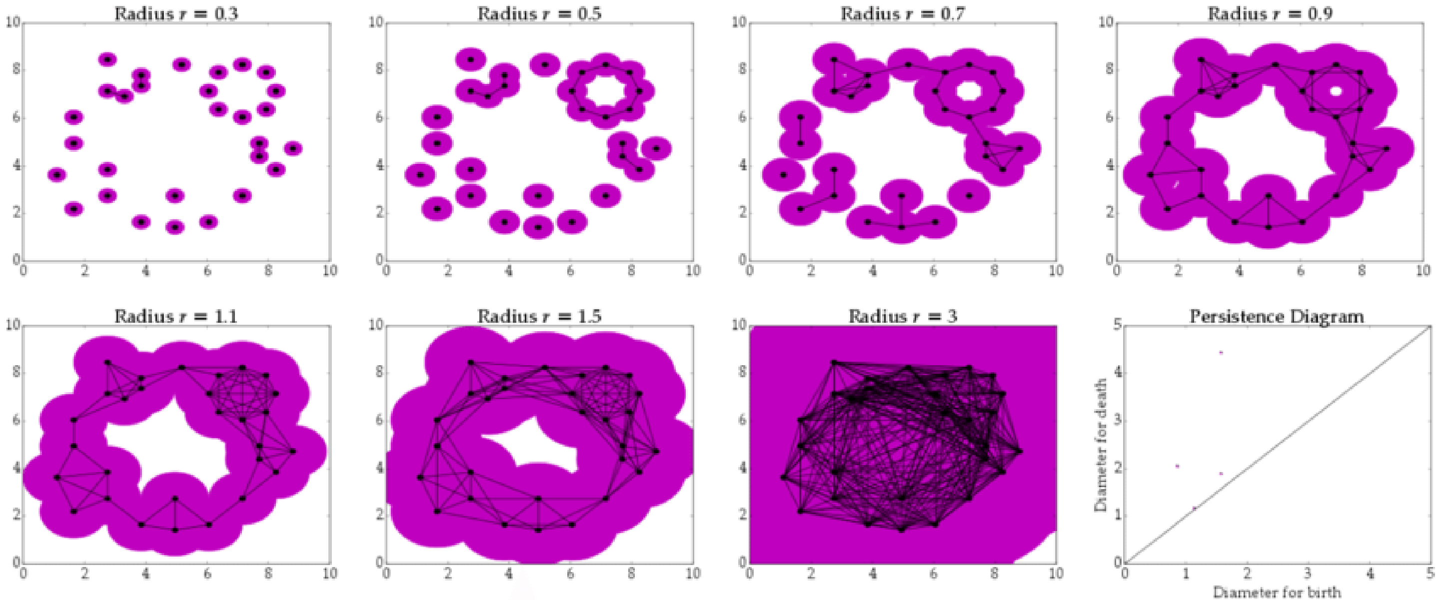


Figure 5: An example of filtration

Source: Munch, Elizabeth. (2017). A User's Guide to Topological Data Analysis. Journal of Learning Analytics. 4. 47-61. 10.18608/jla.2017.42.6.

Persistent Homology

As simplicial complexes transform from one another, some simplicial complexes may appear (birth) and some others disappear (death). Persistence homology examines lifespan (persistence barcode) of these.

Definition

$H_k^{\delta,p} = Z_k^{\delta}/(B_k^{\delta+p} \cap Z_k^{\delta})$ is called p-persistent k^{th} homology of K^{δ}.

Then, the corresponding Betti number is $\beta_k^{\delta,p} = rank(H_k^{\delta,p})$. This is the number of k-dimensional holes that persist (neither birth nor death) during the interval $(\delta, \delta + p)$.

Persistence Diagrams

For a simplicial complex, let $[s, t)$ represents its persistence barcode. Here $\delta = s$ is the birth and $\delta = t$ is the death. Then, each point (s, t) has a multiplicity $\mu_n^{s,t}$ which is the number of simplicial comlexes that are born exactly at s and die exactly at t. Then,

$$\mu_n^{s,t} = (\beta_n^{s,t-1} - \beta_n^{s,t}) - (\beta_n^{s-1,t-1} - \beta_n^{s-1,t}).$$

Persistence diagram is the set of points (s,t), with multiplicities $\mu_n^{s,t}$ such that $0 \leq s < t \leq m+1$ where m is the largest value of δ.

Persistence Diagrams (Example)

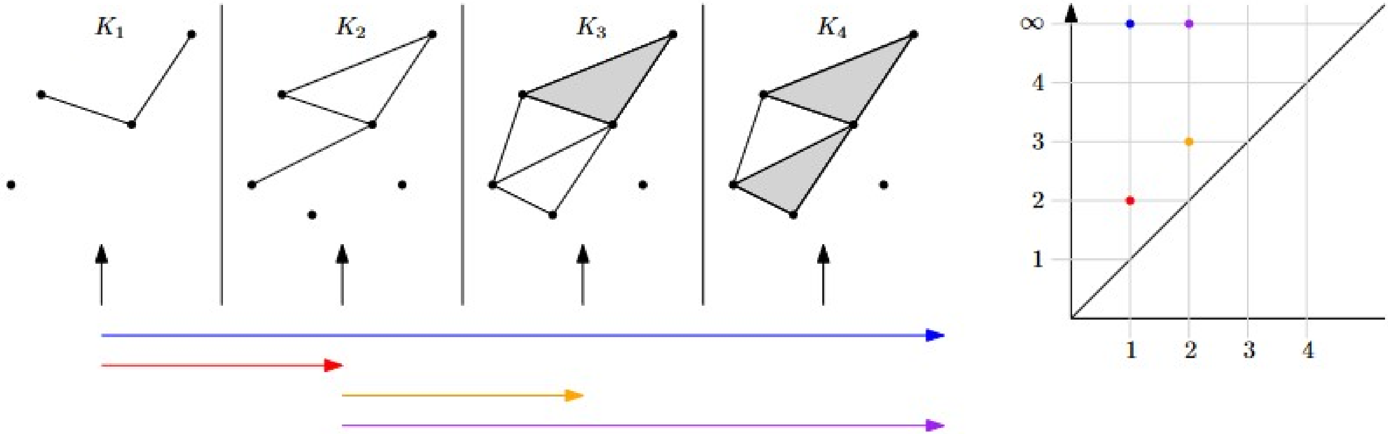


Figure 6: Persistence barcodes and persistence diagram

Source: Ž. Virk. Introduction to Persistent Homology. Založba UL FRI, University of Ljubljana, 2022, doi: 10.51939/0002.

Fundamental Lemma of Persistent Homology

Lemma (Fundamental Lemma of Persistence Homology)

Let $K^0 \subseteq K^1 \subseteq ... \subseteq K^m$ *be a filtration. For every pair of indices* k, l *such that* $0 \leq k \leq l \leq m$,

$$\beta_n^{k,l} = \sum_{\substack{l<j, \\ 0 \leq i \leq k}} \mu_n^{i,j}$$

This equation implies that the information given by persistent Betti numbers are the same as the information given by barcodes or persistence diagrams.[3]

References

[1] Henry Adams and Jan Segert. *Simplicial complex filtration demonstrations in Mathematica*. 2011. URL: https://www.math.colostate.edu/~adams/research/.

[2] Allen Hatcher. *Algebraic topology*. Cambridge: Cambridge Univ. Press, 2000.

[3] Žiga Virk. *Introduction to Persistence Homology*. Založba UL FRI (University of Ljubljana, Faculty of computer science and informatics), 2022.