Topological Data Analysis

JinHwa Jeong

Department of Mathematics
Sungkyunkwan University

November 15,2022

Introduction

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

Introduction

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

- Linear Regression
- Decision Tree
- Machine Learning

Introduction

Topology is concerned with the properties of a geometric object that are preserved under continuous deformations

Introduction

- Qualitative information is needed

Introduction

- Qualitative information is needed
- Metrics are not theoretically justified

Introduction

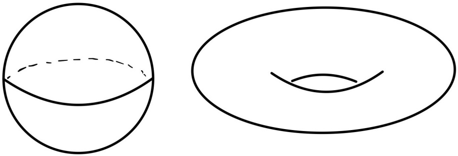
- Qualitative information is needed
- Metrics are not theoretically justified
- Coordinates are not natural

Introduction

- Qualitative information is needed
- Metrics are not theoretically justified
- Coordinates are not natural
- Summaries are more valuable than individual parameter choices

Homology

Figure 1: S_2 & T_2



Using Homology, we can distinguish two spaces by examining their holes.

Simplicial Complex

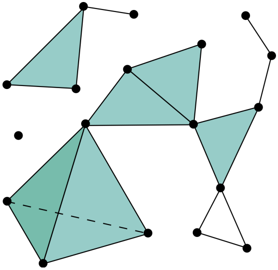
Definition

A simplicial complex K is a collection of simplices such that

- Every face of a simplex of K is in K
- Every pair of distinct simplices of K have disjoint interiors

Simplicial Complex

Figure 2: Simplicial Complex



Intuitively, a simplicial complex structure on a space is an expression of the space as a union of points, intervals, triangles, tetrahedrons and higher dimensional analogues

Homology

Definition

A p-chain on K, $C_p(K)$ is the of formal linear combinations of (oriented) p-simplices

Homology

Definition

A p-chain on K, $C_p(K)$ is the of formal linear combinations of (oriented) p-simplices

Definition

We now define a homomorphism

$$\partial_p : C_p(K) \to C_{p_{-1}}(K)$$

$$\partial_p[v_0, \ldots, v_p] = \sum_{i=0}^{p} (-1)^i \, [v_0, \ldots, \hat{v}_i, \ldots, v_p]$$

is called the boundary operator

Homology

Lemma

$\partial_{p-1} \circ \partial_p = 0$

Homology

Lemma

$\partial_{p-1} \circ \partial_p = 0$

Corollary

$\mathrm{Image}(\partial_{p+1}) \subseteq \mathrm{Kernel}(\partial_p)$

Homology

Lemma

$\partial_{p-1} \circ \partial_p = 0$

Corollary

$\text{Image}(\partial_{p+1}) \subseteq \text{Kernel}(\partial_p)$

Definition

The p^{th} homology group of K is defined by
$H_p(K) = \text{Kernel}(\partial_p)/\text{Image}(\partial_{p+1})$

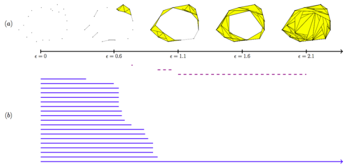
Homology

Definition

The p^{th} betti number $\beta_p(K)$ is the rank of p^{th} homology group $H_p(K)$

Intuitively, the p^{th} Betti number refers to the number of k-dimensional holes on a topological surface.

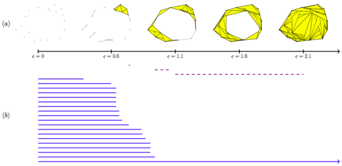
Figure 3: WorkFlow



- Point Cloud Data

WorkFlow

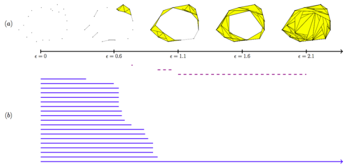
Figure 3: WorkFlow



- Point Cloud Data
- Filtered Complex

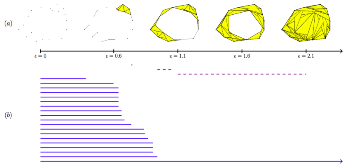
WorkFlow

Figure 3: WorkFlow



- Point Cloud Data
- Filtered Complex
- Persistent Homology

Figure 3: WorkFlow



- Point Cloud Data
- Filtered Complex
- Persistent Homology
- Barcode/Diagram

Filtered Complex

Definition

The nerve of $\mathcal{U}$, denoted by $N\mathcal{U}$, will be the abstract simplicial complex with vertex set A, and where a family $\{\alpha_0, \ldots, \alpha_k\}$ spans a k-simplex if and only if $U_{\alpha_0} \cap \cdots \cap U_{\alpha_k} \neq \emptyset$

Filtered Complex

Definition

The nerve of $\mathcal{U}$, denoted by $N\mathcal{U}$, will be the abstract simplicial complex with vertex set A, and where a family $\{\alpha_0, \ldots, \alpha_k\}$ spans a k-simplex if and only if $U_{\alpha_0} \cap \cdots \cap U_{\alpha_k} \neq \emptyset$

Theorem

Suppose that X and U are as above, and suppose that the covering consists of open sets and is numerable. Suppose further that for all $\emptyset \neq S \subseteq$ A, we have that $\bigcap_{s \in S} U_s$ is either contractible or empty. Then $N\mathcal{U}$ is homotopy equivalent to X

Filtered Complex

Definition

For any subset $V \subseteq X$ or which $X = \bigcup_{v \in V} B_\epsilon(v)$, one can construct the nerve of the covering $B_\epsilon(v)$. We will denote this construction by $\check{C}(V, \epsilon)$, and is called Čech complex attached to V and ϵ

Filtered Complex

Definition

Let X denote a metric space, with metric d. Then the Vietoris-Rips complex for X, attached to the parameter ϵ, denoted by $\mathrm{VR}(X, \epsilon)$, will be the simplicial complex whose vertex set is X, and where $\{x_0, \ldots, x_k\}$ spans a k-simplex if and only if $d(x_i, x_j) \leq \epsilon$ for all $0 \leq i, j \leq k$

Filtered Complex

Theorem

If $\epsilon \leq \epsilon'$. then we have an inclusion of complexes $\check{C}(X, \epsilon) \subseteq \check{C}(X, \epsilon')$

Filtered Complex

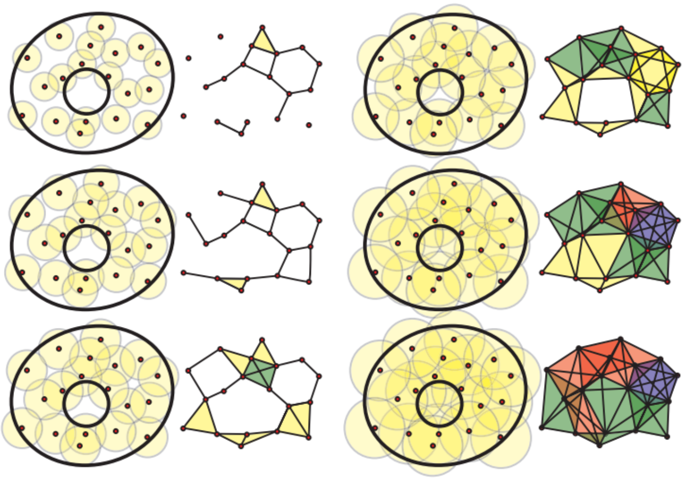
Theorem

If $\epsilon \leq \epsilon'$. then we have an inclusion of complexes $\check{C}(X, \epsilon) \subseteq \check{C}(X, \epsilon')$

Definition

Let K be a finite simplicial complex, and let $K_1 \subset K_2 \subset \cdots \subset K_l = K$ be a finite sequence of nested subcomplexes of K

Figure 4: Filtered Complex



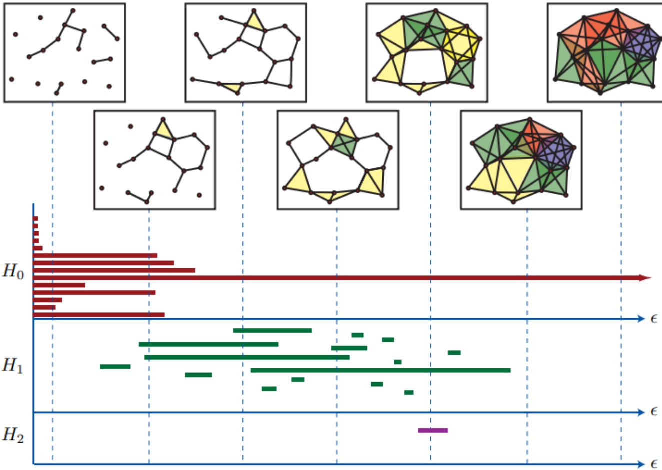
Persistent Homology

Definition

The inclusion map $K_i \to K_j$ induce linear maps $f_{i,j} : H_p(K_i) \to H_p(K_j)$
The *pth* persistent homology of K is the pair $(\{H_p(K_i)\}_{1 \leq i \leq l}, \{f_{i,j}\}_{1 \leq i \leq j \leq l})$

Barcode

Figure 5: Barcode



Thank You

Reference

Ghrist, R. (2008).
Barcodes: the persistent topology of data.
Bulletin of the American Mathematical Society, 45(1):61–75.

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., and Harrington, H. A. (2017).
A roadmap for the computation of persistent homology.
EPJ Data Science, 6(1).