TOPOLOGICAL DATA ANALYSIS

what I have learned so far

Res Proj
Rim FAROUDY
28/03/2025

PLAN

Motivation

Simplices and simplical complexes

Nerves

Homology

Persistent homology

I will skip

Basic definitions of

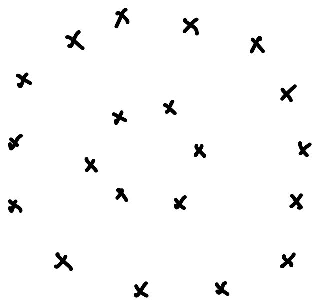
- distance
- metric space
- open / closed set
- cover
- convex set
- connected set

not new (saw them in prépa)

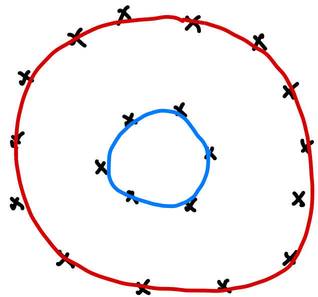
MOTIVATION

MOTIVATION

higher order structure



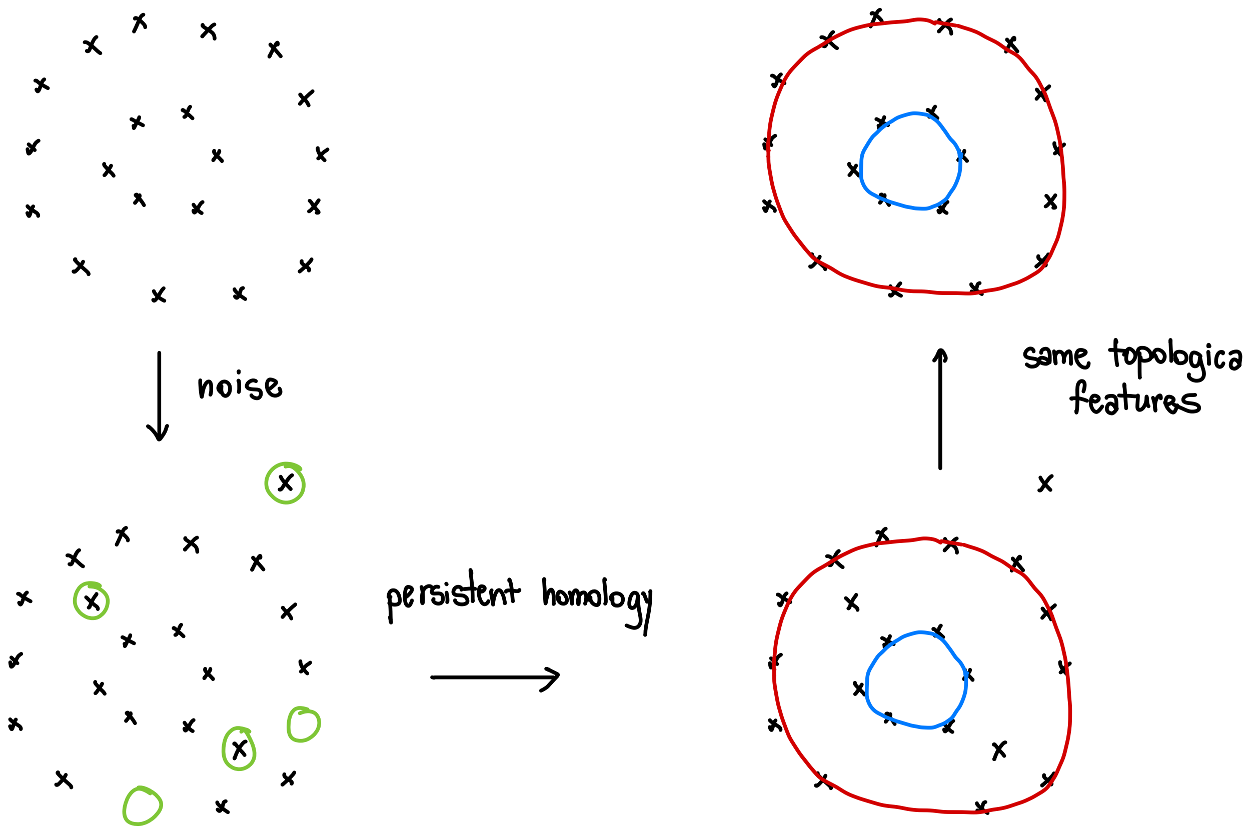
the data



what we want

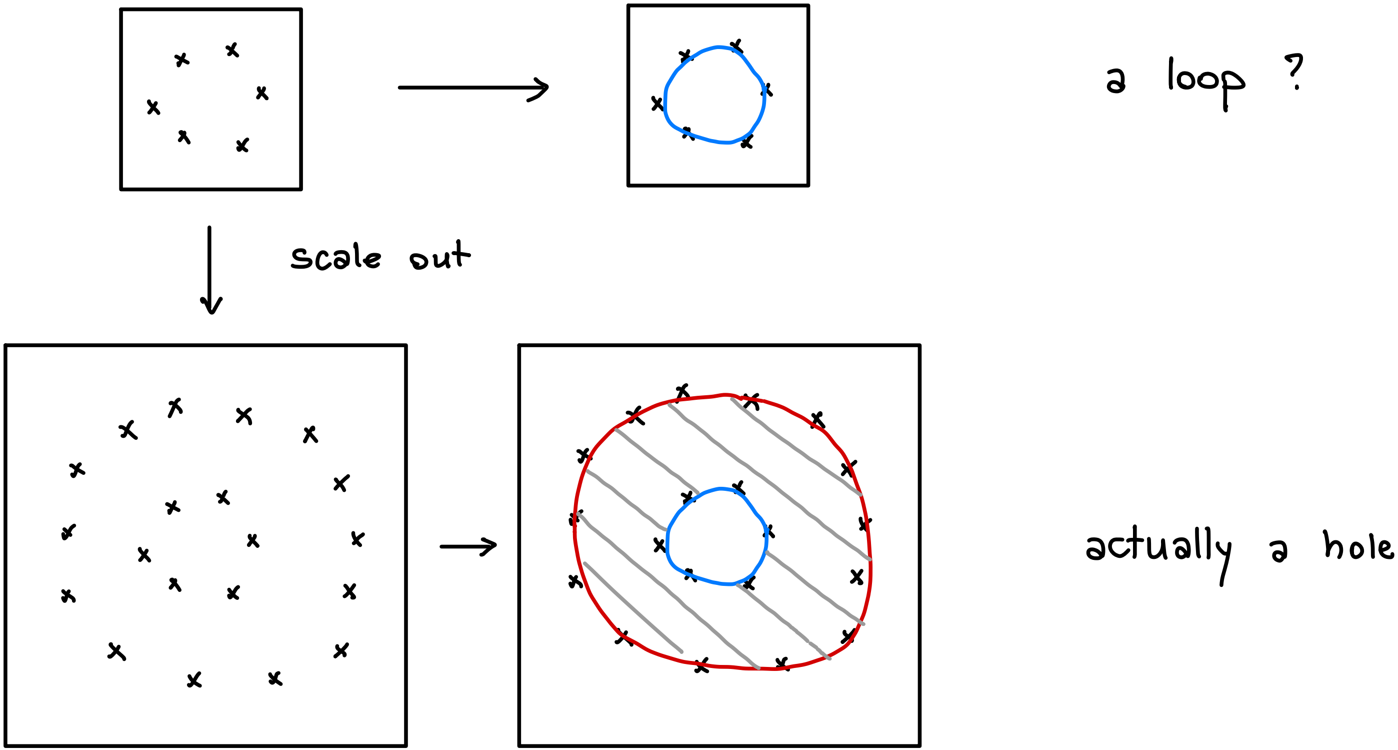
MOTIVATION

robustness to noise



MOTIVATION

multi-scale analysis



SIMPLICES AND SIMPLICAL COMPLEXES

SIMPLICES AND SIMPLICAL COMPLEXES definitions

geometric

- $X = \{ \underbrace{x_0, \dots, x_k}_{\text{affinely independent vertices}} \} \subset \mathbb{R}^d$

- $\sigma = [x_1, \dots, x_k]$: k-dimensional simplex spanned by X

- faces of σ : simplices spanned by subsets of X

- K is a geometric simplical complex if :
  i) every face of a simplex of K is a simplex of K
  ii) the non-empty intersection of any two simplices is a face of both

abstract

- $\tilde{K}$ is an abstract simplicial complex with vertex set V if it is a collection of finite subsets of V such that :
  i) $\forall v \in V$, $v \in \tilde{K}$
  ii) $\forall \sigma \in \tilde{K}$, $\forall \tau \subset \sigma$, $\tau \in \tilde{K}$

SIMPLICES AND SIMPLICAL COMPLEXES properties

geometric

- K can be seen as a topological space through its underlying space
  $$|K| = \bigcup_{\sigma \in S} \sigma$$
  and inherits from the topology of $\mathbb{R}^d$

- if the vertices are known, K is defined combinatorially by a collection of simplices

abstract

- we can associate a geometric realization

topological

intuitive / visualizable

well-suited for algorithms

combinatorial

SIMPLICES AND SIMPLICAL COMPLEXES examples

Given a metric space (M, ρ), $\sigma \in \mathbb{R}_+$, $X \subset M$

VIETORIS – RIPS

$$\mathrm{Rips}_\alpha(X) = \left\{ [x_0, \dots, x_k] \,/\, d_X(x_i, x_j) \leq \alpha \ , \ \forall i, j \right\}$$

⚠ not always embedded in $\mathbb{R}^d$

used with noisy data

CECH

$$\mathrm{Cech}_\alpha(X) = \left\{ [x_0, \dots, x_k] \,/\, \bigcap_{i=0}^{k} B(x_i, \alpha) \neq \phi \right\}$$

used for precision

$$\mathrm{Rips}_\alpha(X) \subset \mathrm{Cech}_\alpha(X) \subset \mathrm{Rips}_{2\alpha}(X)$$

SIMPLICES AND SIMPLICAL COMPLEXES to look into

- DELAUNEY complex and VORONOI diagram
- Alpha complex
- more ?

NERVES

NERVES definitions

- The maps $f_0, f_1 \in \mathcal{C}(X,Y)$ are homotopic $(f_0 \sim_h f_1)$ if

$$\exists H \in \mathcal{C}(X \times [0,1], Y) \;/\; \forall x \in X,\ H(x,0) = f_0(x) \text{ and } H(x,1) = f_1(x)$$

- The topological spaces X and Y are :

homeomorphic if $\exists f \in \mathcal{C}(X,Y),\ g \in \mathcal{C}(Y,X)$ bijective and $\begin{cases} f \circ g = id_Y \\ g \circ f = id_X \end{cases}$

$\uparrow$ weaker than

homotopy equivalent if $\exists f \in \mathcal{C}(X,Y),\ g \in \mathcal{C}(Y,X)$ and $\begin{cases} f \circ g \sim_h id_Y \\ g \circ f \sim_h id_X \end{cases}$

- A space is **contractible** if it is homotopy equivalent to a point

NERVES definition and nerve theorem

Let X be a topological space and $\mathcal{U} = (U_i)_{i \in I}$ a cover of X.
The nerve of $\mathcal{U}$ is the abstract simplicial complex with vertex set $\mathcal{U}$ and such that

$$\sigma = [U_{i_0}, \dots, U_{i_k}] \in C(\mathcal{U}) \Leftrightarrow \bigcap_{j=0}^{k} U_{i_j} \neq \emptyset$$

Nerve theorem

Let X be a topological space and $\mathcal{U} = (U_i)_{i \in I}$ a cover of X such that any subcollection $(U_i)_{i \in J \subset I}$ is either empty or contractible.

Then, X and $C(\mathcal{U})$ are homotopy equivalent.

Why is it useful? It allows us to translate a complex topological problem into a combinatorial one that is simpler to compute.

NERVES the mapper algorithm

dataset X

metric or dissimilarity measure

$f : X \to \mathbb{R}^d$

cover $\mathcal{U}$ of X

input

for each U :

decompose $f^{-1}(U)$ into clusters $C_{U,1}, \dots, C_{U,k_U}$

compute the nerve of X defined by $C_{U,1}, \dots, C_{U,k_U}$

algorithm

$C(\mathcal{U})$

vertices $v_{U,i}$

edges $(v_{U,i}, v_{U',j})$
iff $C_{U,i} \cap C_{U',j} \neq \emptyset$

output

HOMOLOGY

HOMOLOGY introduction and definition

- Homology is a way to associate algebraic structures (groups) to topological spaces to study their features.

- The **space of k-chains** on K $C_k(K)$ is the set of formal sums of k-simplices of K. i.e. if $\sigma_1, \dots, \sigma_p$ are the k-simplices of K then :

$$C(K) = \left\{ \sum_{i=1}^{p} \varepsilon_i \sigma_i \;/\; \varepsilon_i \in \mathbb{Z}_2 \right\} \leftarrow \text{or any field}$$

- The **boundary** of a k-simplex $\sigma = [v_0, \dots, v_k]$ is the $(k-1)$-chain

$$\partial_k(\sigma) = \sum_{i=0}^{k} (-1)^i \, [v_0, \dots, \hat{v}_i, \dots, v_k]$$

Property : $\partial_{k-1} \circ \partial_k \equiv 0 \;, \; \forall k \geqslant 1$

HOMOLOGY introduction and definition [cont.]

- The k^{th} homology group of K is the quotient vector space

$$H_k(K) = \underset{\substack{\uparrow \\ \mathrm{Ker}(\partial_k)}}{Z_k(K)} \, / \, \underset{\substack{\uparrow \\ \mathrm{Im}(\partial_{k+1})}}{B_k(K)}$$

$\partial_k \circ \partial_{k+1} \equiv 0 \Rightarrow \mathrm{Im}(\partial_{k+1}) \subset \mathrm{Ker}(\partial_k)$

- The k^{th} BETTI number of K is $\beta_k(K) = \dim(H_k(K))$

HOMOLOGY to look into

- statistical aspects of homology inference
- reconstruction
- more ?

PERSISTENT HOMOLOGY

PERSISTENT HOMOLOGY introduction and definitions

- The power of persistent homology lies in its ability to distinguish the real structure of data from noise (it allows to track homology groups)
- A **filtration** of a simplical complex K (*resp.* a topological space M) is is a nested family of subcomplexes $(K_r)_{r \in T}$ (subspaces $(M_r)_{r \in T}$) where $T \subset \mathbb{R}$ such that : $(\forall r, r' \in T,\ r \leq r' \Rightarrow K_r \subset K_{r'})$ and $K = \bigcup_{r \in T} K_r$

$$\left(\qquad M_r \subset M_{r'} \qquad M = \bigcup_{r \in T} M_r \right)$$

e.g. the families $\left(\mathrm{Rips}_r(X)\right)_{r \in \mathbb{R}}$ and $\left(\mathrm{Cech}_r(X)\right)_{r \in \mathbb{R}}$

resolution/granularity

PERSISTENT HOMOLOGY persistent modules and persistence diagrams

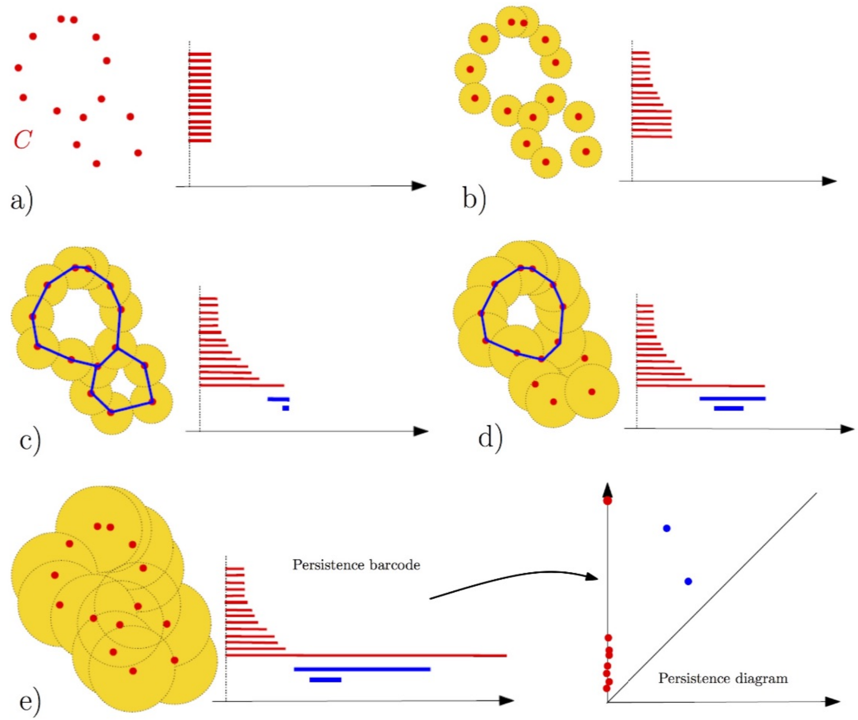
- With $(F_r)_{r \in T \subseteq \mathbb{R}}$ a filtration, the sequence of the vector spaces F_r and the linear maps connecting them is a **persistence module** (obtained by considering $H_k(F_r)$ and $F_r \subset F_{r'}$, $r \leq r'$)

- **Persistence diagrams** provide a geometric representation of the information contained in persistence modules.

  → stable wrt to perturbation in the data

PERSISTENT HOMOLOGY

persistent modules and persistence diagrams



PERSISTENT HOMOLOGY to look into

- statistical aspects of persistent homology
- a lot more

MAIN SOURCE

An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists

Frédéric CHAZAL and Bertrand MICHEL

& ChatGPT ♡

THANK YOU