Persistent homology
A new tool for data analysis

LAMINE Zakaria

25 Novembre 2017

What do we mean by data ?

FIGURE – A point cloud

What do we mean by data

FIGURE – A Data set with a loop

What do we mean by data

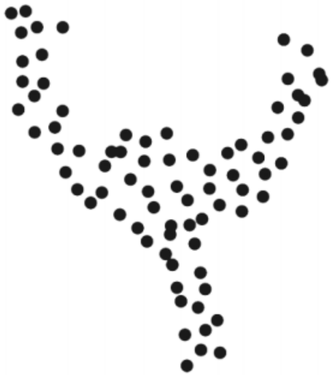


FIGURE – A data set with three tendrils

What do we mean by data

FIGURE – A data set with loop corrupted by noises

Table des matières

Definition :p-simplex

A p-dimensional simplex (or p-simplex) $\sigma^p = [e_0, e_1, ..., e_p]$ is the smallest convex set in a Euclidean space R^m containing the $p+1$ points $e_0, ..., e_p$

Example :

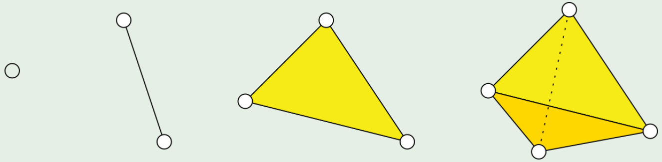


FIGURE – 0,1,2,3 simplex

Definition : Simplicial complex

A simplicial complex is a set composed of points, line segments, triangles, and their p-dimensional counterparts.

Example :

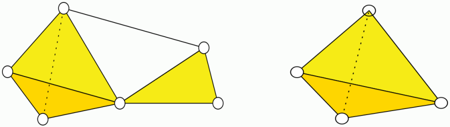


FIGURE – Simplicial complex

Definition : Simplicial complex

A simplicial complex is a set composed of points, line segments, triangles, and their p-dimensional counterparts.

Example :

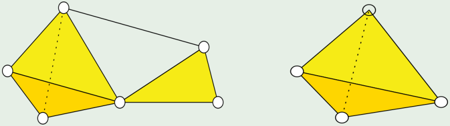


FIGURE – Simplicial complex

Definition : Simplicial complex

A simplicial complex is a finite set of simplices satisfying the following conditions :

1. For all simplices $A \in K$ with α a face of A, we have $\alpha \in K$.
2. $A, B \in K \Rightarrow A, B$ are properly situated.

The dimension of a complex is the maximum dimension of the simplices contained in it.

Example :

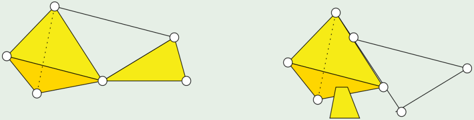


FIGURE – Simplicial complex on the left, not the case on the right

Definition : Nerve

The nerve of a family D of subsets $(c_i)_{i \in I}$, denoted $\mathcal{N}(D)$, is the abstract simplicial complex $\mathcal{K}$ whose elements are all sub-families $(c_i)_{i \in J}$ such that

$$\bigcap_{i \in J} c_i \neq \varnothing$$

.

Definition : Vietoris-Rips complex

Let $(\mathcal{P}, d)$ be a metric space where $\mathcal{P}$ is a point set. Given a real $r > 0$, the Vietoris-Rips complex is the abstract simplicial complex $R^r(\mathcal{P})$ where a simplex $\sigma \in R^r(\mathcal{P})$ if and only if $d(p,q) \leq r$ for every pair of vertices $p, q \in \sigma$.

Definition : Cech complex

The Cech complex $C^r(\mathcal{P})$ is defined to be the nerve of the $\frac{r}{2}$-radius ball $B(p, \frac{r}{2})/p \in \mathcal{P}$,

Theorem (Nerve lemma)

Let F be a finite collection of closed, convex sets in a Euclidean space. Then the nerve of F and the union of the sets in F are homotopic.

Any data have a shape, and any shape have a meaning.

Table des matières

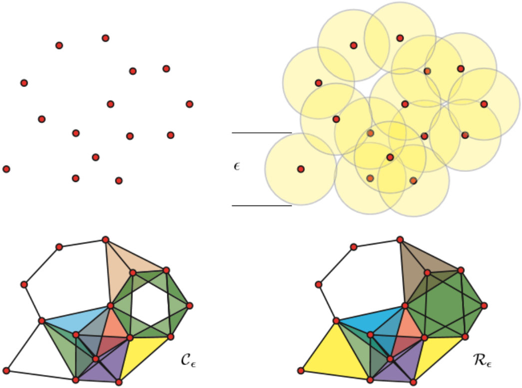


FIGURE – A fixed set of points [upper left] can be completed to a a Cech complex [lower left] or to a Rips complex [lower right] based on a proximity parameter ε [upper right].

Which ε *?.* Converting a point cloud data set into a global complex requires a choice of parameter ε.

Table des matières

Definition

A filtration $\mathcal{F}$ of complex K is any sequence of sub-complexes K_i of K verifying

$$\varnothing \subseteq K_1 \subset \ldots \subset K_n = K.$$

Motivation

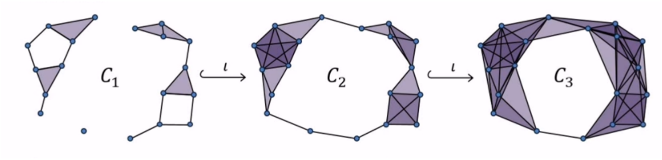


FIGURE – example of filtration

We say that a homology class is "born" in K_i when it does not exist in the image of the application induced by inclusion $K_{i-1} \subset K_i$.

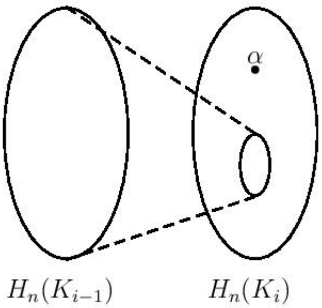


FIGURE – Birth of a homology class

It will be said that it "dies" in K_j when it does not exist in the image of the application induced by inclusion $K_{i-1} \subset K_{j-1}$ but that it exists in the image of the induced application $K_{i-1} \subset K_j$.

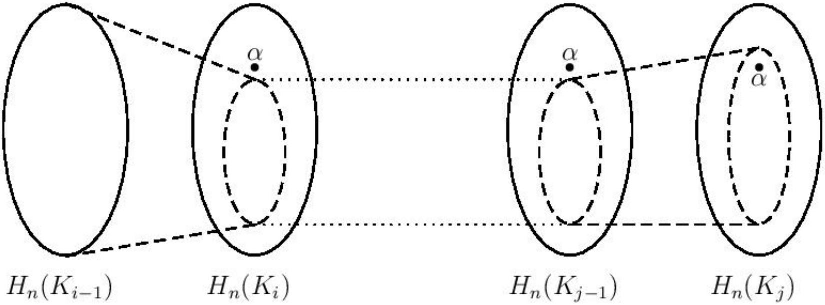
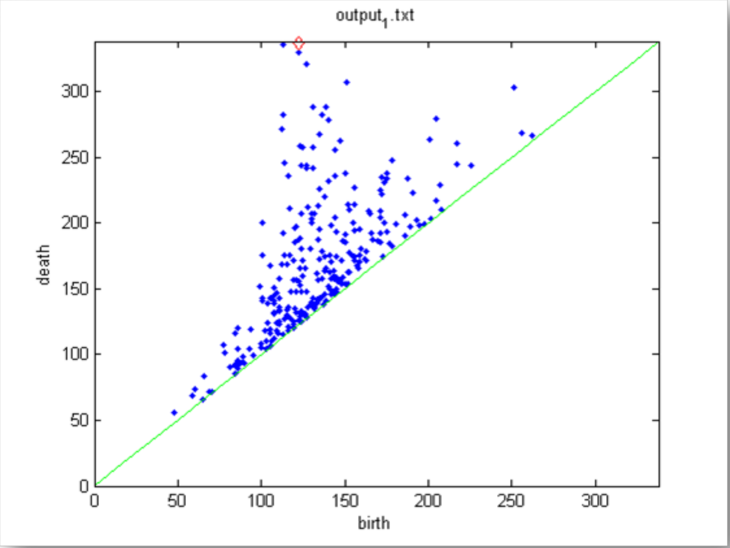


FIGURE – Death of a homology class

Table des matières

Persistent diagrams summarize this information as *two − dimensional* point sets with multiplicities.



given a map $f : X \longrightarrow \mathbb{R}$ and a real t, we define R_t as the preimage of $]-\infty; t]$, ($R_t = f^{-1}]-\infty; t]$) the idea of persistent consists in calculating the homology groups $H_n(R_t)$ as a function of t.

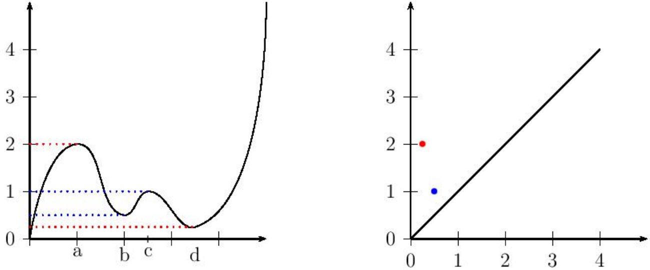


FIGURE – The graph and the associated diagram of f

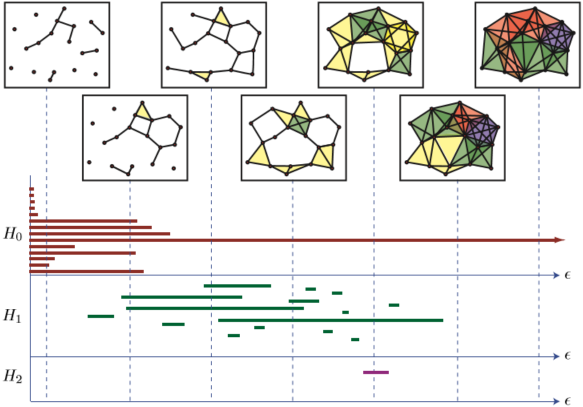


FIGURE – Example of bar codes

Table des matières

Definition : Bottelneck distance

Let X be a topological space with two continuous functions $f; g : X \Rightarrow R$ and let $\mathrm{Dgm}(f), \mathrm{Dgm}(g)$ their respective diagrams. We define their bottleneck distance by taking the infinitum over all supremums

$$d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) := \inf_{\eta} \sup_{x} \|x - \eta(x)\|_{\propto},$$

where $x \in \mathrm{Dgm}(f)$ and $\eta : \mathrm{Dgm}(f) \to \mathrm{Dgm}(g)$ ranges over all bijections, while $\|x - y\|_{\propto} = max|x|, |y|$ is the usual $L_{\propto}$ norm.

Theorem

$$d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leqslant \|f - g\|_\infty.$$

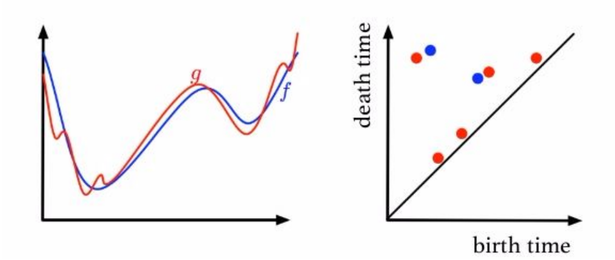


Figure – Left : two functions with small distance. Right : the corresponding two persistence diagrams with small bottleneck distance.

Thank YOU .