Inferring topology through barcodes and cloud points

Marco Praderio

30-10-2019

Table of contents.

Persistent homology

Definition

Given a sequence of chain complexes $C = \left(C_*^i\right)_{i \in I}$ indexed by a **totally ordered** set I together with chain maps $x^{i,j} : C_*^i \to C_*^j$ with $j \geq i$ that behave well (i.e. $x^{j,k} \circ x^{i,j} = x^{i,k}$) we define the (i,j)-persistent homology of C as the image of the induced morphism $x_*^{i,j} : H_*\left(C_*^i, F\right) \to H_*\left(C_*^j, F\right)$, for a fixed **field** F. We denote such an image by $H_*^{i \to j}(C)$.

Example

Take a topological space X embedded in $\mathbb{R}^n$ and define $I = \mathbb{R}$ and $C_*^i = C_*\left(X^i\right)$ for every $i \in \mathbb{R}$

$$X^i = \{(x_1, \ldots, x_n) \in X \mid x_1 < i\}.$$

This example can be then extended to what is known as extended persistent homology by making use of relative homology.

Example

Take a point cloud $\boldsymbol{X} = \{\boldsymbol{x_i}\} \subset \mathbb{R}^n$, set the indexing set to $\boldsymbol{I} = \mathbb{R}^+$. For every point x_i and $\varepsilon > \mathbf{0}$ define B_ε^i as

$$B_\varepsilon^i = \{x \in \mathbb{R}^n \,|\, d = |x - x_i| \leq \varepsilon/2 \text{ and } \min\{|x - x_j|\} = d\}$$

Using this sets we define the ε-**Alpha complex** as the abstract simplicial complex whose k-simplices are determined by unordered $k+1$ tuples of distinct points in X such that the sets B_ε^* containing those points have non empty intersection.

Example

Define the ε-**Cech complex** C_ε as the abstract simplicial complex whose k-simplices are determined by unordered $k+1$ tuples of distinct points in X such that the $\varepsilon/2$ spheres centered on the points of the tuple have a common point.

Theorem

[nerve theorem] *With the previous definition C_ε has the same homotopy type as the union of the the closed balls of radius ε centered in the points of X.*

Example

Define the ε-**Rips complex** C_ε as the abstract simplicial complex whose k-simplices are determined by unordered $k+1$ tuples of distinct points in X such that the distance between any two of those points is less that ε.

Proposition

For any $\varepsilon > 0$ there is a chain of inclusion maps $R_\varepsilon \hookrightarrow C_{\varepsilon\sqrt{2}} \hookrightarrow R_{\varepsilon\sqrt{2}}$.

Life of chains

General definition

Definition

Given $i, j \in I$ such that $i < j$ we say that exactly n k-chains are **alive** in the interval $[i, j]$ if $\dim_F \left(\mathrm{Im}\left(x_k^{i,j}\right)\right) = n$. we denote this number n by $A_k^{i,j}$.

Definition

we say that n k-chains are **born** at time i if

$$\lim_{\varepsilon\to 0}\left(\lim_{j_+\to i^+}\left(A_k^{j_+,i+\varepsilon}\right)-\lim_{j_-\to i^-}\left(A_k^{j_-,i+\varepsilon}\right)\right)=n.$$

Definition

we say that n k-chains **die** at time i if

$$\lim_{\varepsilon\to 0}\left(\lim_{j_-\to i^-}\left(A_k^{i-\varepsilon,j_-}\right)-\lim_{j_+\to i^+}\left(A_k^{i-\varepsilon,,j_+}\right)\right)=n.$$

Life of chains

Special definition

Remark

In the simple case where the set I is at most countable (or we can restrict I to a countable while preserving information about birth and death of all chains) we can give the C a structure of $F[x]$ module by making x act on C_*^i as the map $x_*^{i,i+1}$. This gives $H_*(C, F)$ a structure of $F[x]$ module and if every $H_*(C_i, F)$ has finite dimension then, from the structure theorem of PID we can write

$$H_*(C,F) \cong \bigoplus_i x^{t_i} \cdot F[x] \oplus \left(\bigoplus_j x^{r_j} \cdot (F[x] / x^{s_j} \cdot F[x]) \right)$$

This isomorphism of modules allows an alternative definition of birth and death of chains.

Definition

Using the notation of the previous remark we say for every i that a chain which lives forever is born at t_i and for every j that a chain which dies at $r_j + s_j$ is born at r_j.

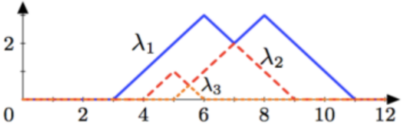
Definition

The n-th persistence landscape is a function $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\lambda(k, t) = \sup\left(m \geq 0 | A_n^{t-m,t+m} \geq k\right).$$

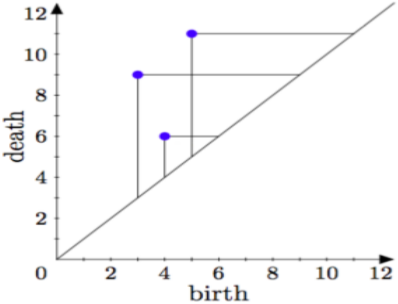
Visualizing persistent homology

persistence landscape



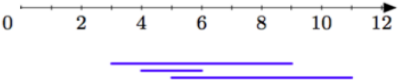
Visualizing persistent homology

persistence diagram



Visualizing persistent homology

persistence barcodes



Analyzing data structure of images

pre-processing

For every image of a dataset of 4167 random outdoor digital images the following steps are performed:

1. 5000 blocks of 3×3 pixels are selected.
2. Those blocks are normalized by mean intensity.
3. The top 20% with the greatest contrast is chosen.

Problem

The dataset obtained with this pre-processing appears at first to be distributed over all the seventh-sphere $S7$.

Solution

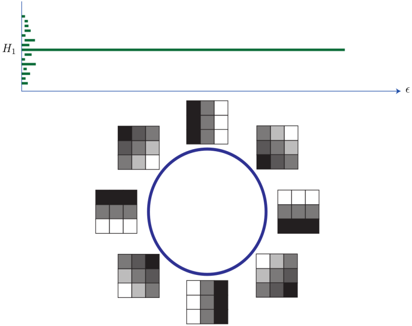
Apply density filtering.

Analyzing data structure of images

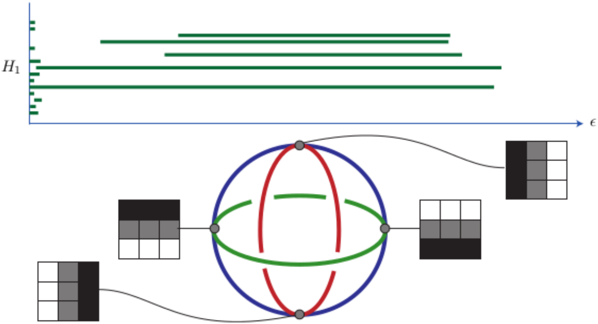
filtering

1. For every point x of the dataset the distance $\delta_k(x)$ to its k-th nearest neighbor is computed.
2. Only the top T percent of the dataset points with the lowest δ_k are kept.

Taking $k = 300$ and $T = 25$ the following result is obtained.



Taking $k = 15$ and $T = 25$ the following result is obtained.



References

Peter Bubenik.
Statistical topological data analysis using persistence landscapes.
Journal of Machine Learning Research, 2015.

Herbert Edelsbrunner and John Harer.
COMPUTATIONAL TOPOLOGY, AN INTRODUCTION.

Robert Ghrist.
Barcodes: The persistent topology of data.
BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY, 2007.