Homología Persistente
Homología Persistente
Introduction
Persistent Homology
Three Applications

Homología Persistente

Dr. Mauricio Toledo-Acosta

Diplomado Ciencia de Datos con Python

Topological Data Analysis

El Análisis Topológico de Datos (TDA) en un enfoque para el análisis de datos usando técnicas de la topología.

Topological Data Analysis

El Análisis Topológico de Datos (TDA) en un enfoque para el análisis de datos usando técnicas de la topología. La extracción de información de conjuntos de datos de dimensión alta y con ruido es difícil. El TDA provee un enfoque general para analizar estos datos de manera que es insensible a la métrica de fondo y es robusta al ruido.

Homología Persistente
Homología Persistente
Introduction
Persistent Homology
Three Applications

Topological Data Analysis

El Análisis Topológico de Datos (TDA) en un enfoque para el análisis de datos usando técnicas de la topología. La extracción de información de conjuntos de datos de dimensión alta y con ruido es difícil. El TDA provee un enfoque general para analizar estos datos de manera que es insensible a la métrica de fondo y es robusta al ruido.

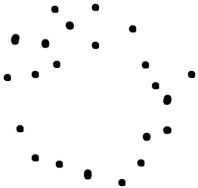
La motivación inicial es estudiar la forma de los datos. El TDA combina la topología algebraica y otras herramientas de las matemáticas para permitir un estudio riguroso de la forma. Una de las herramientas principales es la homología persistente, la cual ha sido aplicada en muchas áreas.

Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.

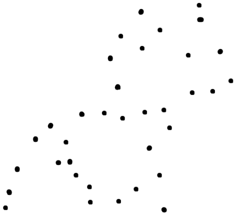
Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.



Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.

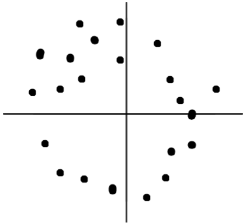


Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas

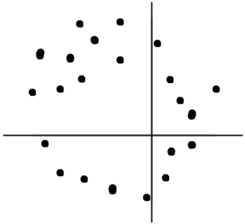
Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas



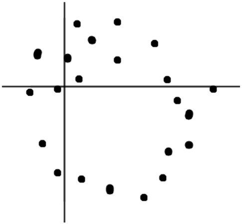
Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas



Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas



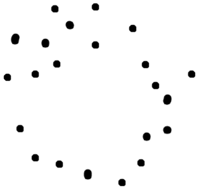
Homología Persistente
Homología Persistente
Introduction
Persistent Homology
Three Applications

Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas, robusto a las perturbaciones de los datos de entrada.

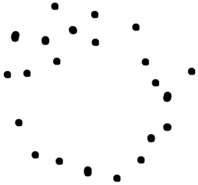
Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas, robusto a las perturbaciones de los datos de entrada.



Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas, robusto a las perturbaciones de los datos de entrada.



Introduction

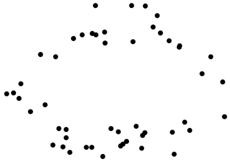
- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas, robusto a las perturbaciones de los datos de entrada.
- La Homología Persistente da cuenta de la geometría local y la topología global de un conjunto de datos.

Introduction

- La homología persistente es un método utilizado en el análisis topológico de datos para estudiar características cualitativas de los datos que persisten a través de múltiples escalas. Se introdujo en 2002.
- Es independiente de las dimensiones y de las coordenadas, robusto a las perturbaciones de los datos de entrada.
- La Homología Persistente da cuenta de la geometría local y la topología global de un conjunto de datos.
- Tiene muchas aplicaciones: Análisis de imágenes, NLP, Flujos, como método de extracción de features para algoritmos de machine learning.

The shape of a cloud of points

Tenemos un conjunto de puntos $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



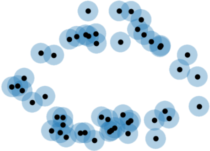
The shape of a cloud of points

Homología Persistente
Homología Persistente
Introduction
Persistent Homology
Three Applications

Tenemos un conjunto de puntos $X = \{x_1, ..., x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.

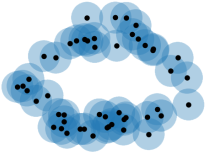
The shape of a cloud of points

Tenemos un conjunto de puntos $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



The shape of a cloud of points

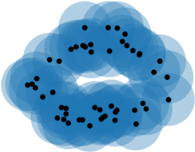
Tenemos un conjunto de puntos $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



The shape of a cloud of points

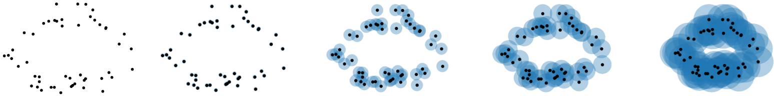
Homología Persistente
Homología Persistente
Introduction
Persistent Homology
Three Applications

Tenemos un conjunto de puntos $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



The shape of a cloud of points

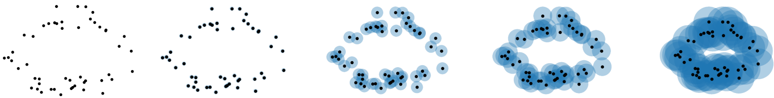
Tenemos un conjunto de puntos $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



Cuál es radio *correcto*?

The shape of a cloud of points

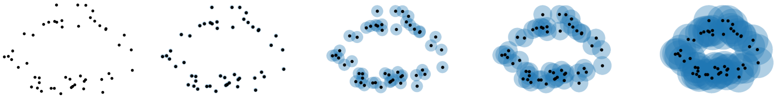
Tenemos un conjunto de puntos $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, queremos averiguar la forma de X.



Cuál es radio *correcto*? Consideramos todos los posibles radios.

Filtrations

- Consideramos la familia de espacios para todos los radios $r > 0$.



Homología Persistente

Homología Persistente

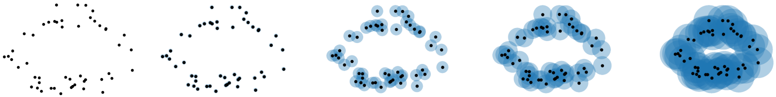
Introduction

Persistent Homology

Three Applications

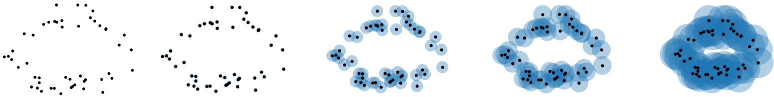
Filtrations

- Consideramos la familia de espacios para todos los radios $r > 0$.
- Cada uno de estos espacios representa la figura determinada por los puntos, en varias escalas.



Filtrations

- Consideramos la familia de espacios para todos los radios $r > 0$.
- Cada uno de estos espacios representa la figura determinada por los puntos, en varias escalas.
- Cada espacio está contenido en el siguiente. Esto es llamado una **filtración**.

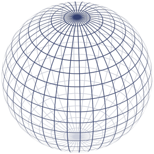


Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.

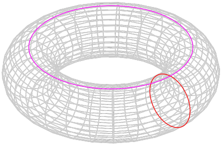
Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.



Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.



Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.

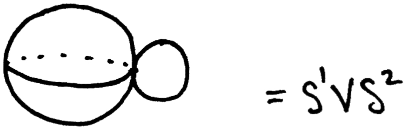
Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.

Esta sucesión de números se llaman los **números de Betti**.

Counting holes

La forma que estamos tratando de determinar está dada por lo número de hoyos k-dimensionales para cada $k < D$.

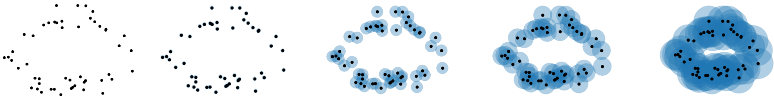


Esta sucesión de números se llaman los **números de Betti**. Tienen la siguiente interpretación:

- b_0: número de componentes conexas.
- b_1: número de hoyos unidimensionales o hoyos "circulares".
- b_2: número of de hoyos 2-dimensionales: "vacíos" o "cavidades".

Tracking holes

El objetivos es llevar el registro de los hoyos que aparecen y desaparecen mientras recorremos la filtración.



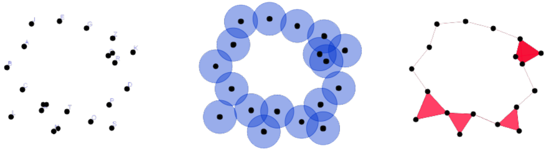
El tiempo que cada hoyo (ciclo) persiste es un indicativo de la presencia o ausencia de dicho ciclo en el conjunto de datos.

Čech Complex and Vietoris-Rips Complex

- Calcular los números de Betti en la filtración previa es difícil.

Čech Complex and Vietoris-Rips Complex

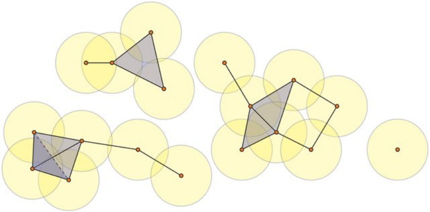
- Calcular los números de Betti en la filtración previa es difícil.
- Construimos un complejo simplicial:



Čech Complex and Vietoris-Rips Complex

Hay dos maneras de construir los complejos simpliciales:

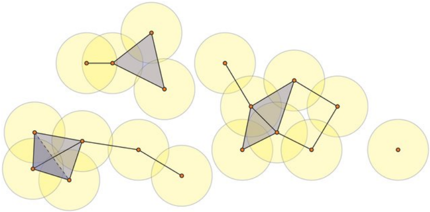
- Vietoris-Rips Complex:



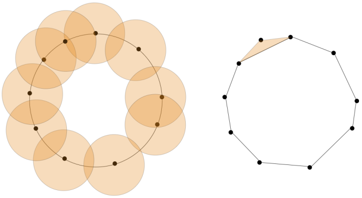
Čech Complex and Vietoris-Rips Complex

Hay dos maneras de construir los complejos simpliciales:

- Vietoris-Rips Complex:



- Čech Complex:



Čech Complex and Vietoris-Rips Complex: Differences

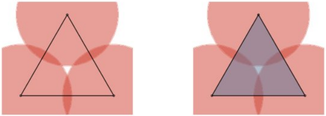


Figure: Čech (left), Vietoris-Rips (right)

Čech complex	**Vietoris-Rips complex**
Smaller	Bigger
Computationally expensive	Less expensive
More accurate	Less accurate

The new filtration

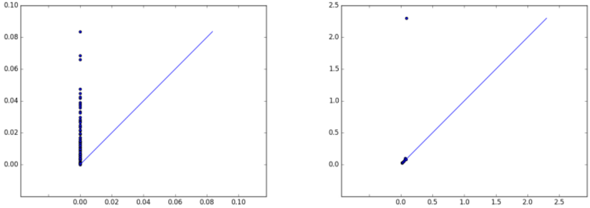
La filtración es infinita, sin embargo, sólo hay un número finito de complejos diferentes.

The new filtration

La filtración es infinita, sin embargo, sólo hay un número finito de complejos diferentes.

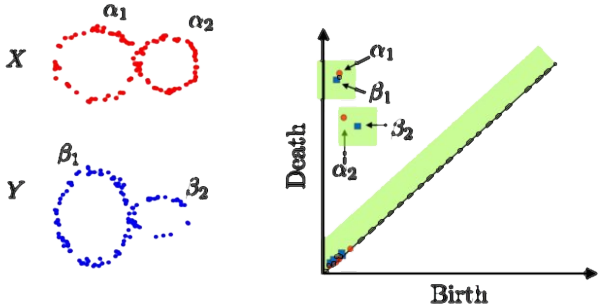
Llevamos el registro de los agujeros que aparecen y desaparecen a medida que pasamos por cada complejo de la filtración. El tiempo que persiste cada ciclo es un indicativo de la presencia de dicho ciclo en la figura.

Representing Results: Persistence Diagrams

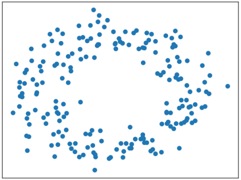


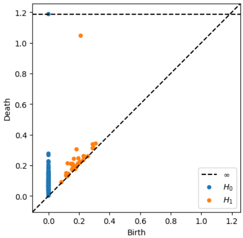
Ciclos que viven poco tiempo se consideran ruido.

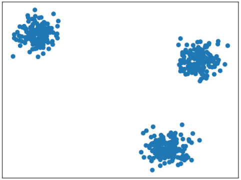
Visualizing Results: Persistence Diagrams

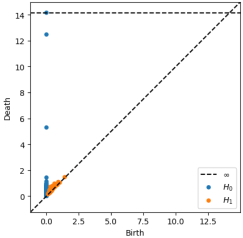


Examples

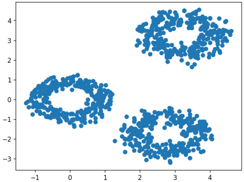


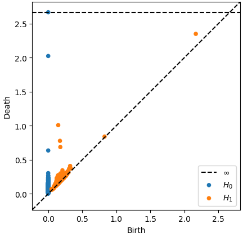

1.2
1.0
0.8
0.6
0.4
0.2
0.0
Death
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Birth
∞
H_0
H_1

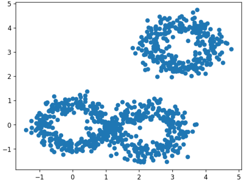


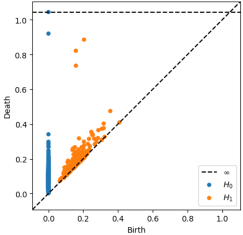

14
12
10
8
6
4
2
0
Death
0.0
2.5
5.0
7.5
10.0
12.5
Birth
∞
H_0
H_1

Examples



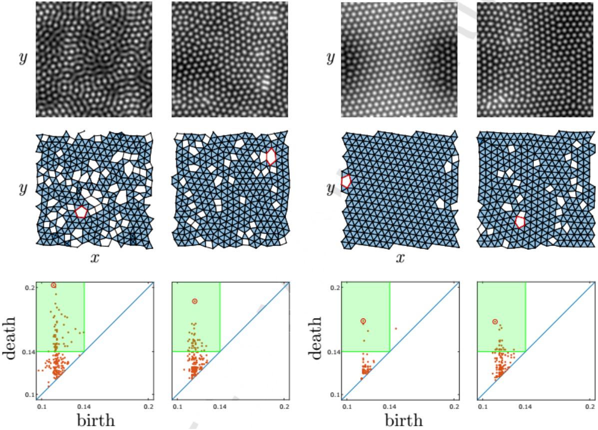






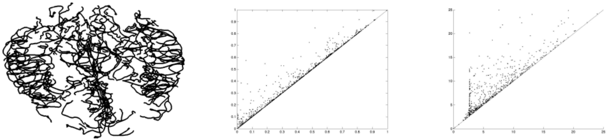
Local Geometry

In this paper, authors compare quantitative measures of order for nearly hexagonal, planar lattices using Persistent Homology.



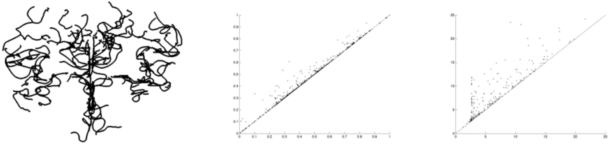
Motta, F. C., Neville, R., Shipman, P. D., Pearson, D. A., Bradley, R. M. (2018). **Measures of order for**

When *noise* is the important feature



(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 13: Persistent homology data objects from a 24-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.

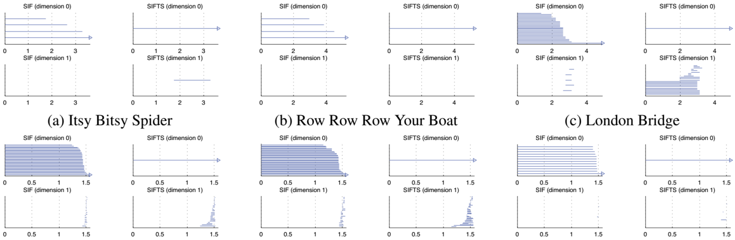


(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 14: Persistent homology data objects from a 68-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.

Bendich, P., Marron, J. S., Miller, E., Pieloch, A., Skwerer, S. (2016). **Persistent homology analysis of brain artery trees**. The annals of applied statistics, 10(1), 198

Identifying Semantic *Tie-Backs*



(a) Itsy Bitsy Spider (b) Row Row Row Your Boat (c) London Bridge
(d) The Emperor's New Clothes (e) Little Red-Cap (f) Alice in Wonderland

Figure 1: Persistent homology on nursery rhymes and other stories

Zhu, X. (2013, August). **Persistent Homology: An Introduction and a New Text Representation for Natural Language Processing**. In IJCAI (pp. 1953-1959).