Proyecto Homología Persistente

Diego Villarreal De La Cerda - 173591
Análisis Topológico de Datos

19 de junio de 2025

1. Introdución

En los últimos años, la creciente disponibilidad de grandes volúmenes de datos ha impulsado la necesidad de herramientas que no solo analicen su contenido cuantitativo, sino que también revelen su estructura subyacente. En este contexto, el análisis de homología persistente ha emergido como una técnica poderosa dentro de la topología computacional, capaz de detectar patrones y formas en los datos más allá de lo que permiten los métodos estadísticos tradicionales.

La homología persistente se enfoca en estudiar las características topológicas de los datos como componentes conexas, ciclos y cavidades a través de distintas escalas. Esto permite identificar estructuras que "persisten.ª lo largo de diferentes niveles de resolución, separando los rasgos relevantes de aquellos que podrían deberse al ruido o a la variabilidad del muestreo. Una de sus grandes fortalezas es precisamente esa capacidad de adaptarse a datos complejos y multidimensionales, extrayendo información robusta y significativa incluso en contextos donde otras herramientas fallan.

Este ensayo tiene como objetivo explorar y aplicar el análisis de homología persistente sobre conjuntos de datos, con el fin de interpretar sus diagramas de persistencia y comprender mejor su estructura topológica. A través de este estudio, se busca no solo ilustrar el potencial de esta técnica, sino también desarrollar una intuición más profunda sobre cómo los datos se conectan y qué nos pueden decir más allá de sus valores numéricos.

2. Implementación del Algoritmo

Si bien ya existen librerías implementadas en python que realizan el análisis de homología, proponemos una implementación propia de uso libre que se puede encontrar en GitHub o ir a la siguiente URL : `https://github.com/DiegoVilla03/Topological-`

`Data-Analysis`

Para realizar el análisis de homología persistente primero hay que definir una filtración adecuada para los datos con base a ella generar el conjunto simplicial, en este caso, dado que es la filtración computacionalmente más eficiente utilizamos la filtración alfa.

La filtración alfa es un tipo de complejo simplicial construido por triangulaciones de Delaunay, La triangulación de Delaunay de un conjunto de puntos en el espacio es una triangulación en la que ninguna de las circunferencias circunscritas a los triángulos contiene en su interior ningún otro punto del conjunto de puntos, excepto quizás sus vértices. En términos más simples, es una forma de triangular un conjunto de puntos de tal manera que los triángulos formados sean óptimos

El valor de la filtración para complejo simplicial es calculada como el cuadrado de los circunradios del simplejo, si la circunsfera es vacía se dice que el complejo simplical es una gráfica de Gabriel , y como mínimo de los valores de filtración de la codimensión 1 cofaces que hacen que no sea de Gabriel de otra manera.

A continuación utilizaremos conjuntos de datos con geometrías de prueba para observar como funciona el algoritmo, los conjuntos de prueba y distintas variaciones puede encontrase en el link de GitHub

2.1. Esfera $\mathbb{S}^2$

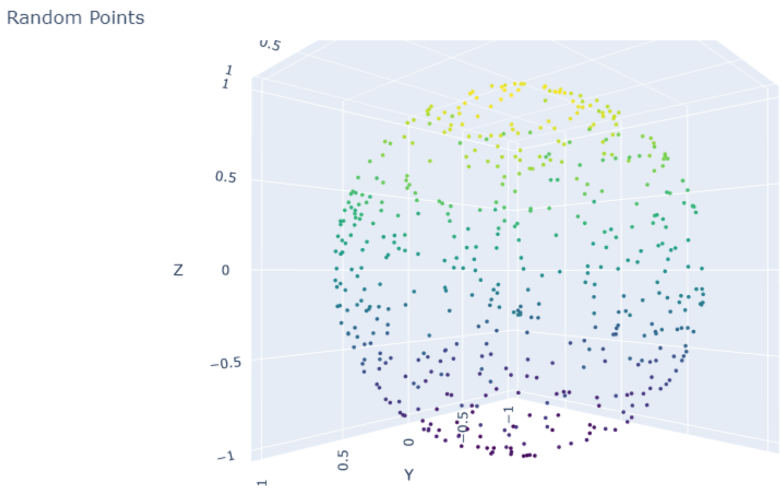

Figura 1: Datos en Esfera uitaria

Se generaron 500 datos uniformemente distribuidos sobre la esfera $\mathbb{S}^2$, la geometría de la esfera es la de una cavidad por lo que esperamos del análisis de homología persistencia de una sola cavidad durante mucho tiempo, es poco probable la formación de ciclos

Aplicando el algoritmo y graficando el diagrama de persistencia

Como se puede esperar hay una cavidad que nace en 0.1 y muere en 1, con el conocimiento de que se trata de una esfera unitaria entonces todos los puntos se unen en 1 sola

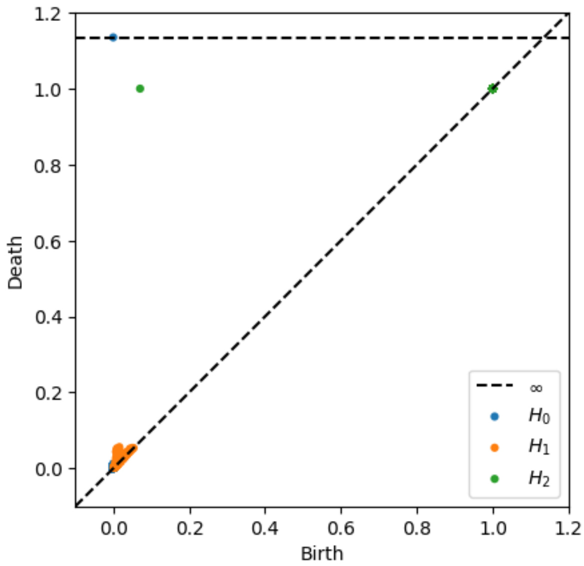

Figura 2: Diagrama de Persistencia para la $\mathbb{S}^2$

componente conexa

Como nota adicional podemos describir la esfera $\mathbb{S}^2$ con los números de Betti

$$\beta_0 = 1 \qquad \beta_1 = 0 \qquad \beta_2 = 1$$

2.2. Cinta de Möbius

La cinta de Möbius es un objeto matemático bastante particular y curioso, ya que se puede recorrer al completo siguiendo una linea que no se corta , en otras palabra que tiene una sola cara

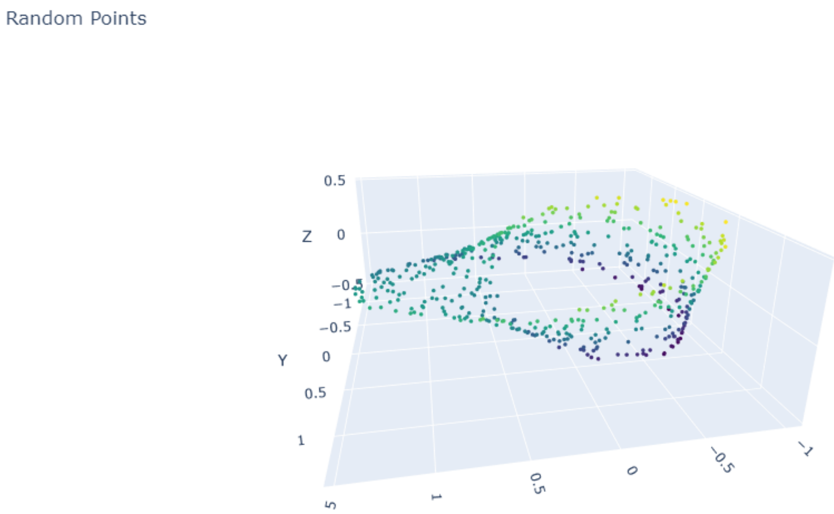

Figura 3: Cinta de Möbius

Esperaríamos encontrar un agujero en medio de la cinta para algun valor de la filtración, que persista hasta que todo de una en una componente conexa

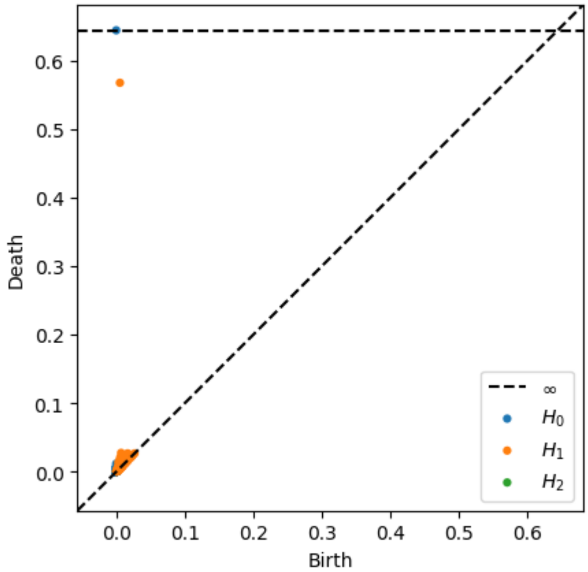

Figura 4: Diagrama de persistencia Cinta de Möbius

Del diagrama de persistencia Extraemos la formación de varios ciclos que no persisten durante mucho tiempo en la filtración, incluso de la formación de cavidades posiblemente generadas por la torsión de la cinta de aquí rescatamos un ciclo que persiste casi desde el 0 hasta 0.6 donde el ciclo desaparece para formar una sola componente

de la figura podemos extraer entonces sus números de betti

$$\beta_0 = 1 \qquad \beta_1 = 1 \qquad \beta_2 = 0$$

2.3. Toroide $\mathbb{T}^2$

El Toroide o particularmente el toro, es una figura geométrica con forma de dona dicho a groso modo, cuya superficie de revolución es generada por una circunferencia que gira alrededor de una recta exterior coplanaria.

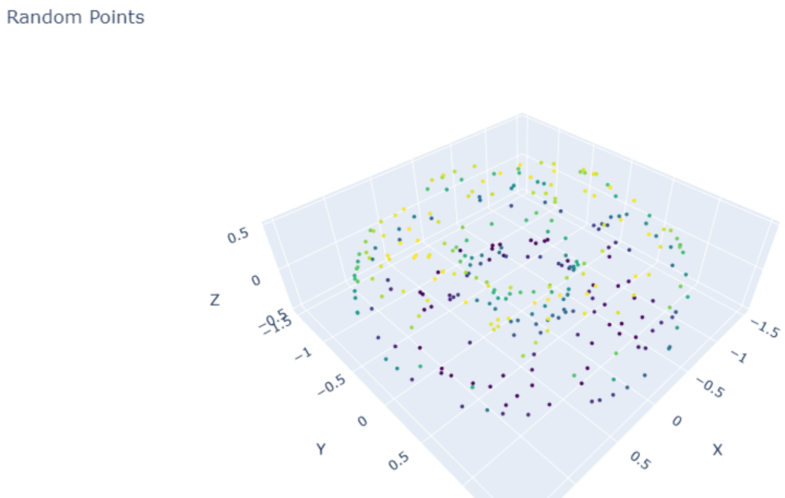

Figura 5: Toro $\mathbb{T}^2$

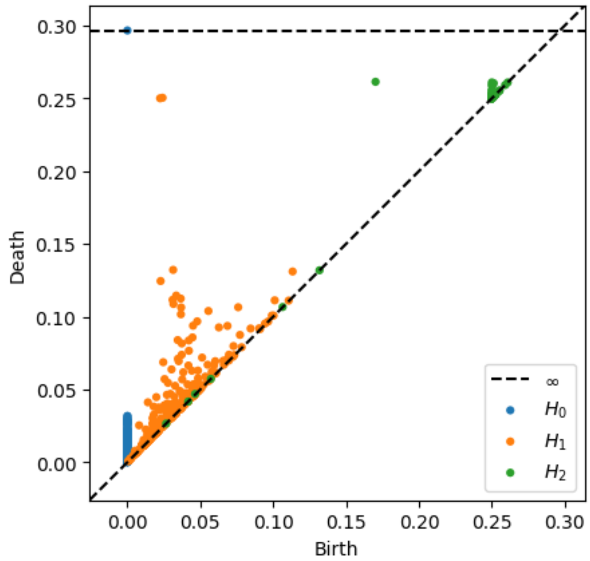

Figura 6: Diagrama de persistencia del Toro $\mathbb{T}^2$

Del diagrama de persistencia extraemos que se forman 2 ciclos que nacen cerca de 0.03 y que mueren en 0.25 ademas de la formacion de una cavidad que nace en 0.15 y persiste igualmente hasta 0.25 por lo que podemos decir de sus números de Betti son

$$\beta_0 = 1 \qquad \beta_1 = 2 \qquad \beta_2 = 1$$

2.4. Series de Tiempo

Antes de entrar a analizar el comportamiento de la persistencia homologica de una serie de tiempo hay que hacer un pequeño preámbulo

Cuando se busca estimar los parámetros de un modelo se series temporales, particularmente cuando este se trata de un SARIMA *(Seasonal Autoregressive Integrated Moving*

Average) es importante poder estimar su estacionalidad el componente S, muchas veces esto puede ser fácil dada la granularidad temporal del fenómeno, la captura de datos o por simple intuición, sin embargo cuando esta estacionalidad no es trivial hay dificultades para poder estudiarlos, si bien hay métodos matemáticos que dan una aproximación bastante buena este problema se puede resolver de una manera muy creativa con el uso de la topología e incrustaciones con retardo temporal.

Una incrustación con retardo temporal o incrustación de Takens, llamada así en honor a Floris Takens, pionero en su uso en el estudio de sistemas dinámicos. Una incrustación con retardo temporal puede concebirse como el deslizamiento de una "ventana"de tamaño fijo sobre una señal, donde cada ventana se representa como un punto en un espacio (posiblemente) de mayor dimensión.

Formalmente dada una serie de tiempo X_{t+k} uno puede extraer los k retardos de la serie de tiempo en la forma

$$f_i = [X_i, X_{i+\tau}, X_{i+2\tau}, \ldots, X_{i+(d-1)\tau}] \in \mathbb{R}^d$$

Donde d es la dimensión de la incrustación y τ es la cantidad de retardos aplicados a la serie de tiempo donde la cantidad $(d-1)\tau$ es el tamaño de ventana y la diferencia entre X_i, X_{i+1} es llamado paso, así podemos construir una función en términos de $f(d,\tau)$

$$TD_{d,\tau}f : \mathbb{R} \to \mathbb{R}^d, \quad t \mapsto \begin{bmatrix} f(t) \\ f(t+\tau) \\ f(t+2\tau) \\ \vdots \\ f(t+(d-1)\tau) \end{bmatrix}$$

Con esto la idea es explorar la topología resultante para varios valores de d, τ en la incrustación para y descubrir si la serie de tiempo X_t tiene patrones o secuencias no triviales.

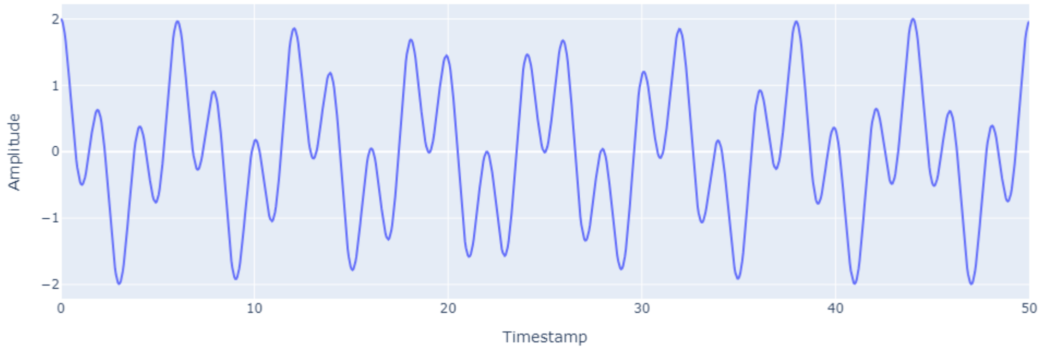

Figura 7: Simulación de una Serie de tiempo

Se simularon puntos aleatorios entre 0 y 50 para la función

$$f(x) = cos(x) + cos(\pi x)$$

Aplicamos la incrustación a los datos , ahora bien la importancia de seleccionar buenos parámetros para la incrustación es que esta dará lugar a una mejor interpretación de los diagramas de persistencia.

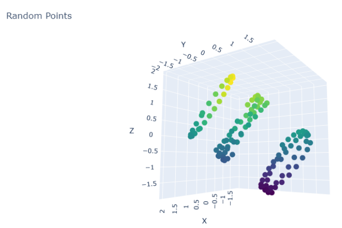

Figura 8: Incrustación d $= 20$ $\tau = 8$

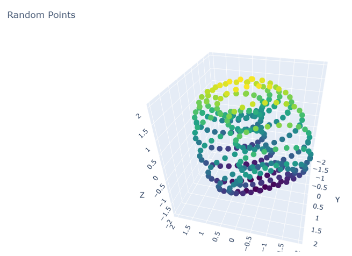

Figura 9: Incrustación d $= 16$ $\tau = 3$

Según la elección de los parámetros la topología de la incrustación puede revelar mas o menos información sobre la serie de tiempo, incluso generando encajes en una bidimensionalidad mayor

Por ejemplo en la Figura 8 se forman 3 componentes bien separadas, mientras que en la Figura 9 queda una esfera con una espiral dentro de ella.

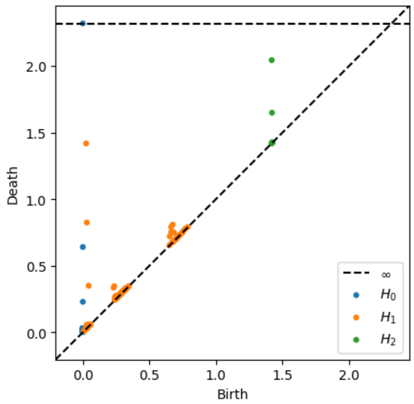

Figura 10: Diagrama de persistencia d = 20 $\tau = 8$

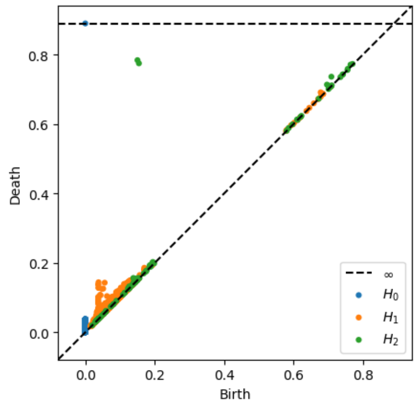

Figura 11: Diagrama de persistencia d = 16 $\tau = 3$

Al analizar los diagramas de persistencia observamos que para la primera incrustación se distinguen ciclos separados persistentes que nos indican lo que hemos visto en la imagen, lo que hace sospechar de un periodo de 8 unidades mientras que para el diagrama de persistencia de la segunda incrustación su homologia es muy parecida a la de una esfera por lo que no podemos deducir mucas conclusiones de esta

3. Resultados y Conclusiones

En este trabajo se ha desarrollado e implementado desde cero un algoritmo de homología persistente basado en filtraciones Alfa, demostrando su eficacia para extraer información topológica relevante de distintos conjuntos de datos sintéticos. A través de la generación de diágramas de persistencia y el cómputo de los números de Betti

Estos resultados ilustran cómo la homología persistente permite separar rasgos geométricos y topológicos "verdaderos" del ruido y la variabilidad del muestreo, proporcionando una visión complementaria a los métodos estadísticos clásicos. Además, la implementación propia abre la puerta a adaptaciones y optimizaciones específicas según el dominio de aplicación.

Comprender la forma de los datos en un sentido topológico abre nuevas puertas a la interpretación, la clasificación y la detección de patrones ocultos. Por ello, el análisis de homología persistente representa una herramienta cada vez más relevante en la ciencia de datos moderna, y su estudio resulta no solo interesante, sino también necesario.