Aprendizaje en Variedades para el Análisis Topológico de Datos

Ximena Fernández

V Encuentro de Jóvenes Topólogos

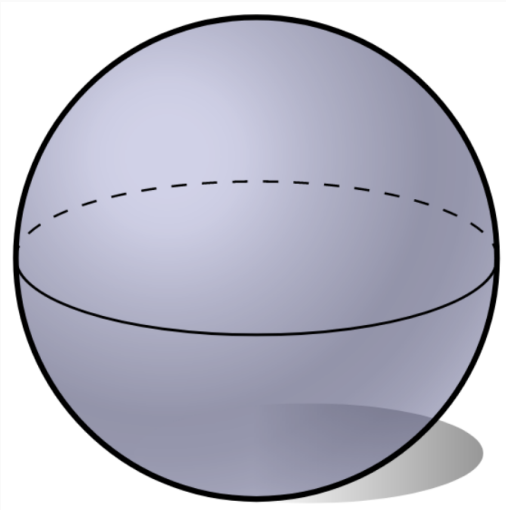
26 febrero 2021

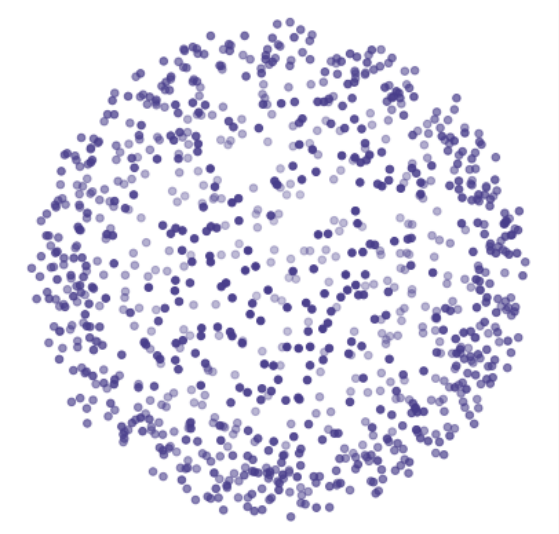
Liverpool-Oxford-Swansea Centre for Topological Data Analysis

Image: Keenan Crane

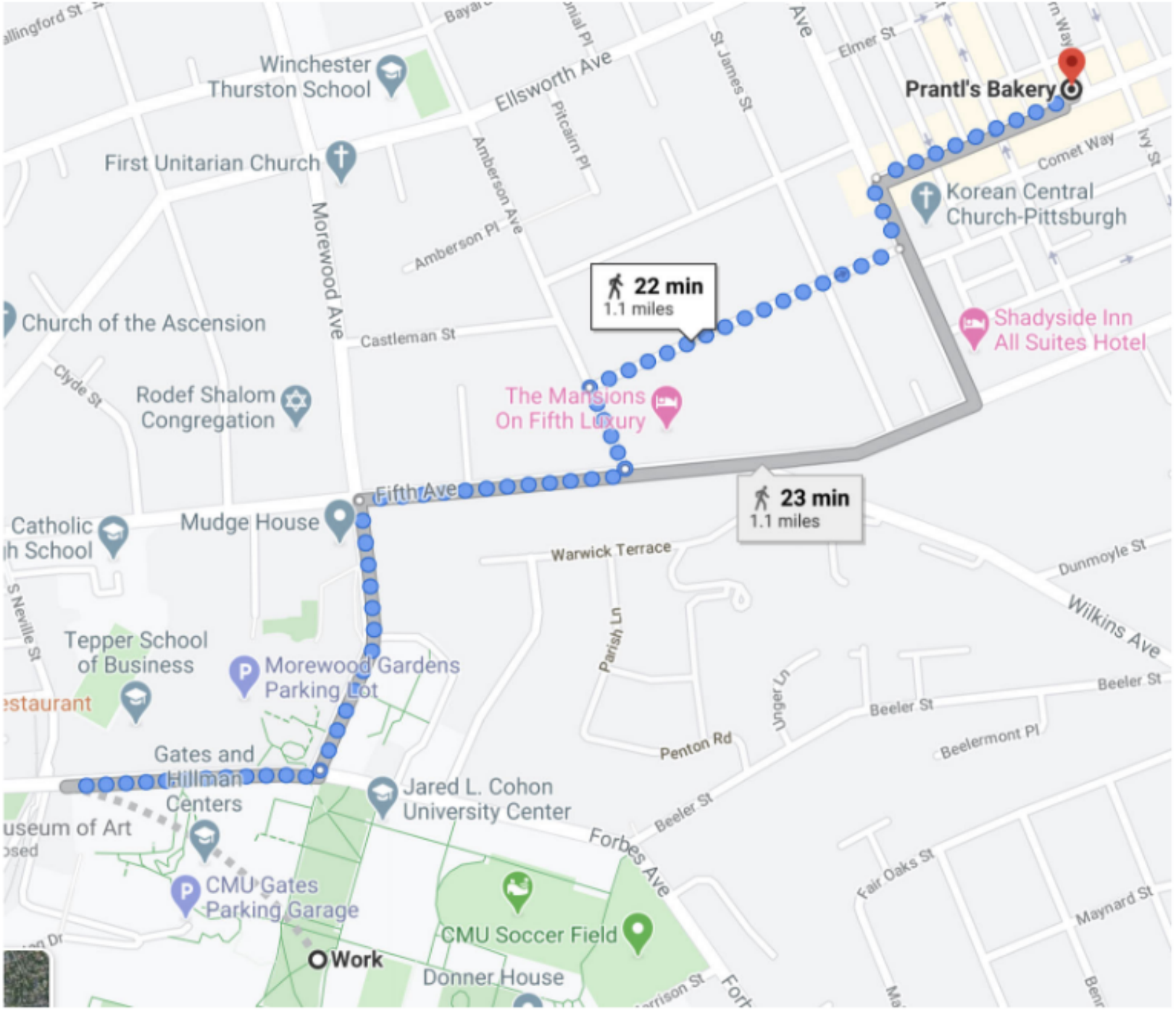
Objetivo

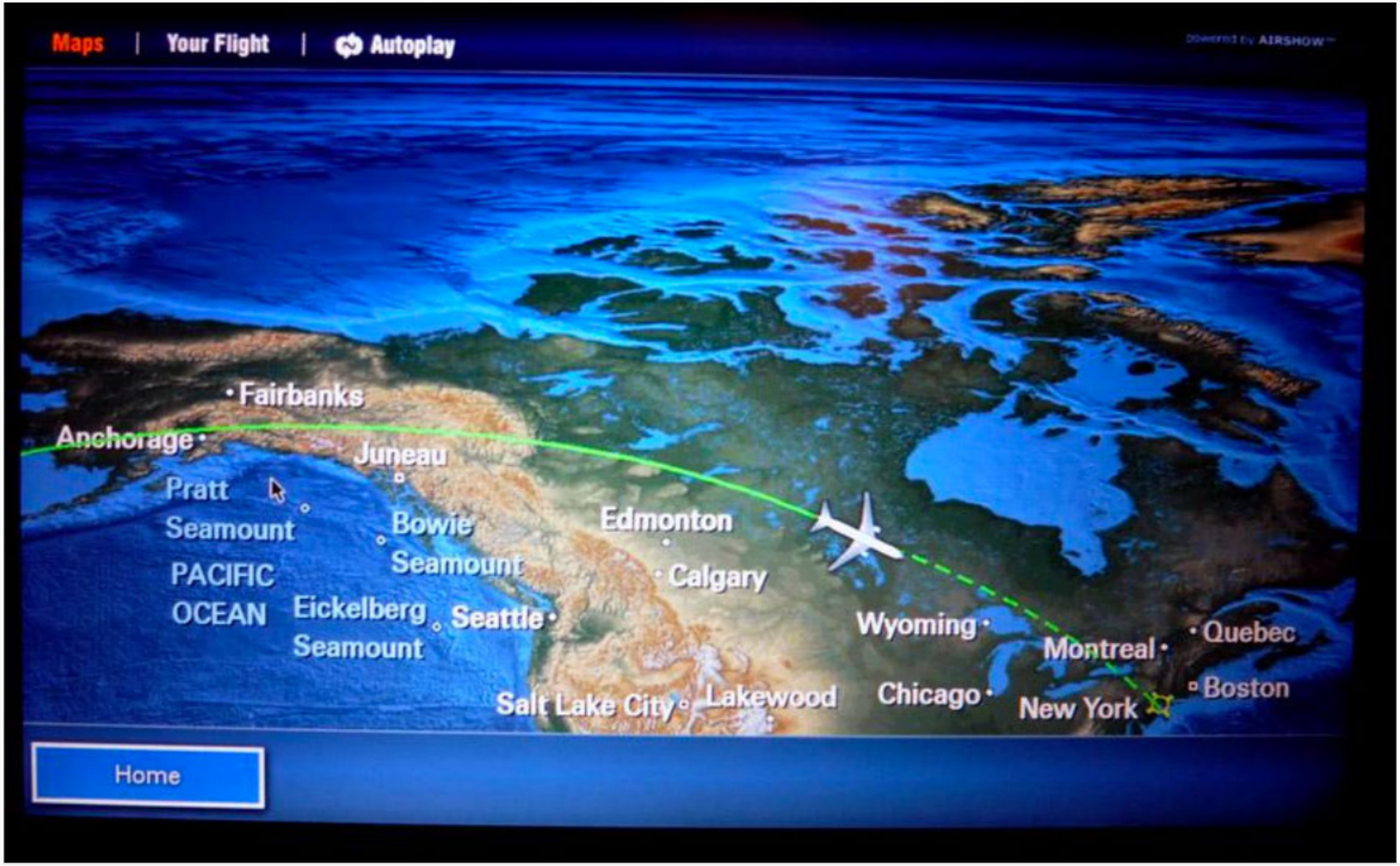
Inferir información de una variedad a partir de una muestra de puntos.

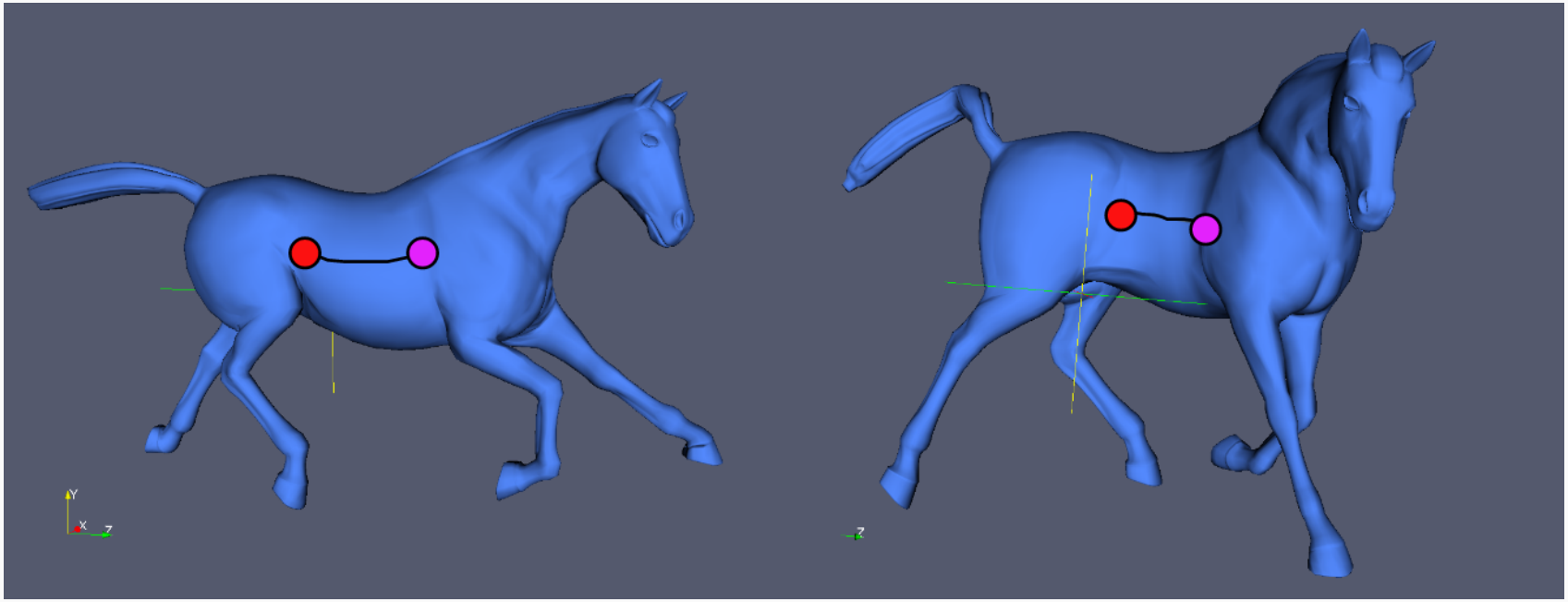




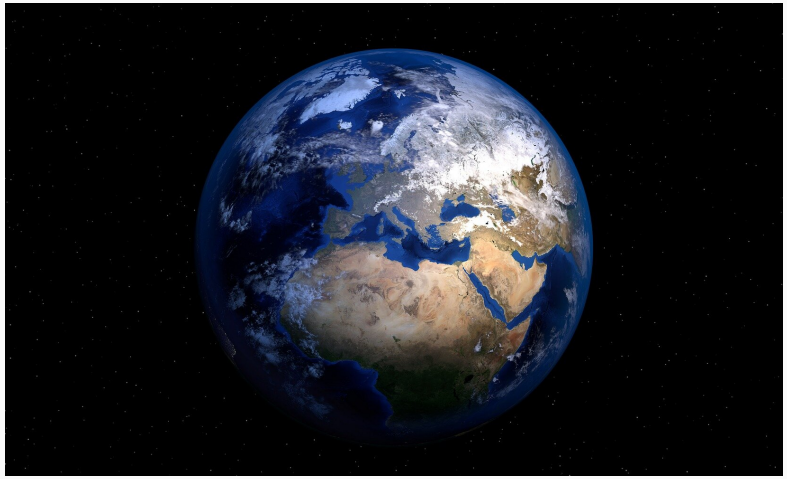
¿Cómo medir distancias?


Prantl's Bakery
22 min
1.1 miles
23 min
1.1 miles
Work
Winchester Thurston School
First Unitarian Church
Church of the Ascension
Rodef Shalom Congregation
Korean Central Church-Pittsburgh
Shadyside Inn All Suites Hotel
The Mansions On Fifth Luxury
Mudge House
Catholic h School
Tepper School of Business
Morewood Gardens Parking Lot
Gates and Hillman Centers
Jared L. Cohon University Center
CMU Gates Parking Garage
CMU Soccer Field
Donner House
Ellsworth Ave
Morewood Ave
Amberson Ave
Amberson Pl
Pitcairn Pl
St James St
Elmer St
Comet Way
Castleman St
Clyde St
Fifth Ave
Warwick Terrace
Parish Ln
Unger Ln
Penton Rd
Beeler St
Beelermont Pl
Dunmoyle St
Wilkins Ave
Forbes Ave
Fair Oaks St
Maynard St
S Neville St


Maps
Your Flight
Autoplay
Fairbanks
Anchorage
Juneau
Pratt
Seamount
PACIFIC
OCEAN
Bowie
Seamount
Eickelberg
Seamount
Seattle
Edmonton
Calgary
Wyoming
Montreal
Quebec
Boston
Chicago
New York
Salt Lake City
Lakewood
Home


Y
X
Z
Z

¿Cómo medir distancias *intrínsecas*?



Bernhard Riemann (1826-1866)

Euclides (325 a.C.-265 a.C.)

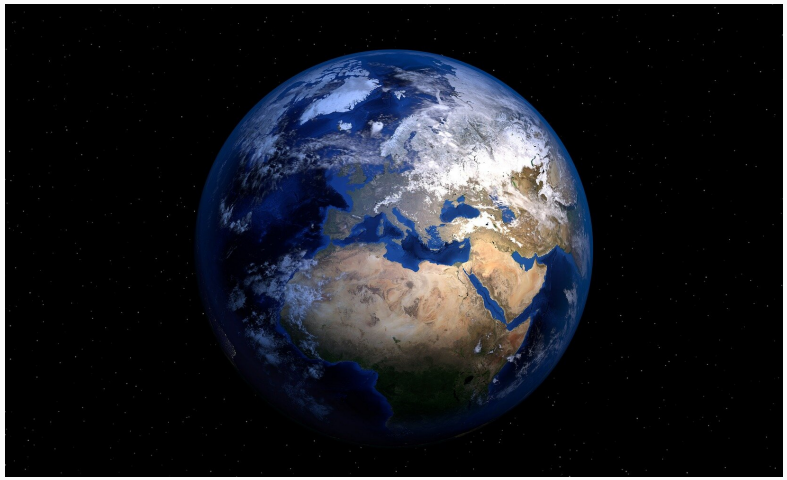
¿Cómo medir distancias intrínsecas?

Dados x, y dos puntos de un *espacio topológico* $\mathcal{M}$.

$\gamma : I \to \mathcal{M}$ tal que $\gamma(0) = x,\ \gamma(1) = y$

$d_{\mathcal{M}}(x, y) = \inf_{\gamma}(\mathrm{len}(\gamma))$

Terraplanistas vs Globistas



Propiedad

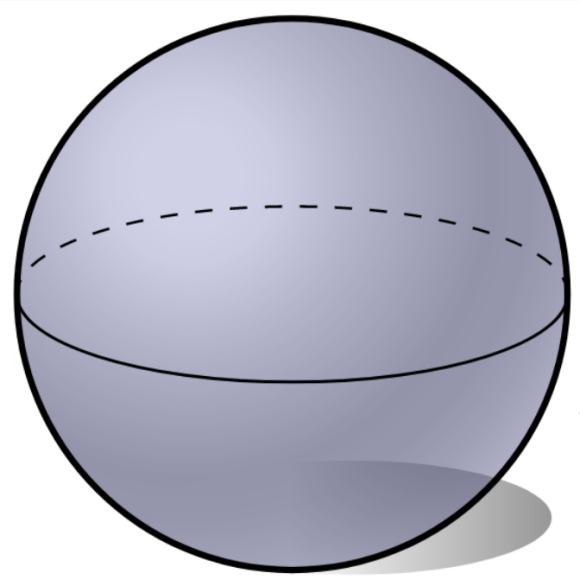
Dado $\delta > 0$, $\exists \epsilon > 0$ tq si $|x - y| < \epsilon$

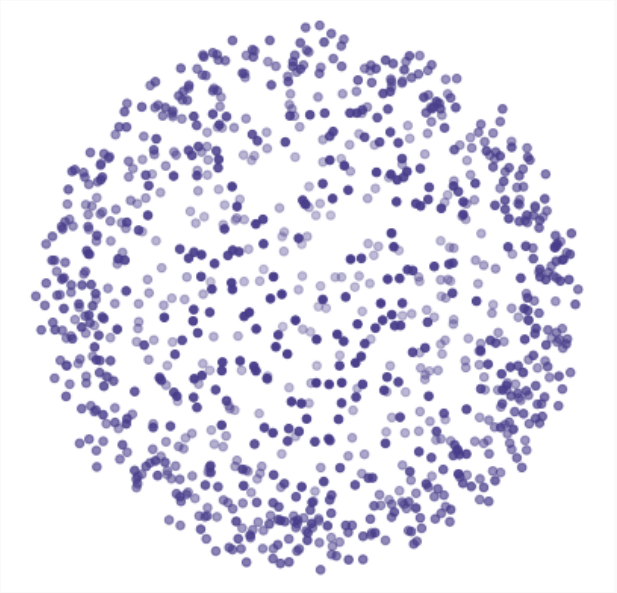
$$|x - y| \leqslant d_{\mathcal{M}}(x, y) \leqslant (1 + \delta)|x - y|$$

Localmente, $d_{\mathcal{M}}$ se puede aproximar por la distancia Euclídea.

¿Cómo *inferir* distancias?

Aprendizaje de distancias





ISOMAP

Dado $\epsilon > 0$, se construye el grafo G_ε con:

- vértices: $V = \mathbb{X}_n$,
- aristas: $E = \{(x_i, x_j) : |x_i - x_j| < \varepsilon\}$.

Se define el **estimador**

$$d_{\mathbb{X}_n,\varepsilon}(x, y) = \inf_{\gamma} \sum_{i=0}^{r} |x_{i+1} - x_i|$$

donde el ínfimo es sobre todos los caminos $\gamma = (x, x_1, \ldots, , x_r, y)$ con $(x_i, x_{i+1}) \in E(G_\varepsilon)$ para todo $0 \leqslant i \leqslant r - 1$.

Distancia de ISOMAP

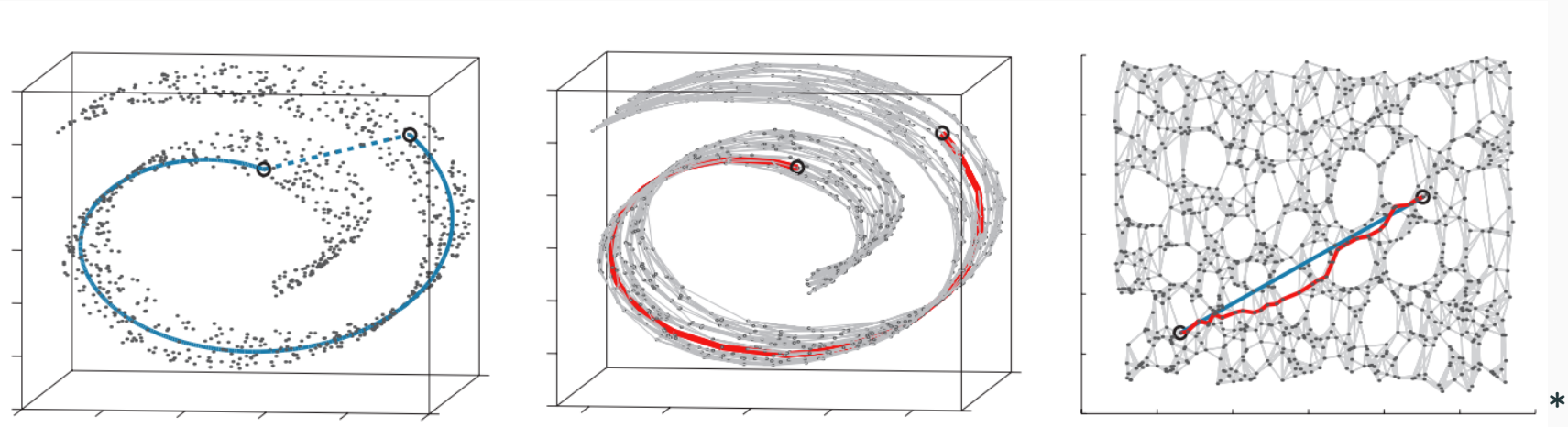


*Imagen: M. Bernstein, V. D. Silva, J. C. Langford, and J. B. Tenenbaum. ***Graph approximations to geodesics on embedded manifolds***, 2000.

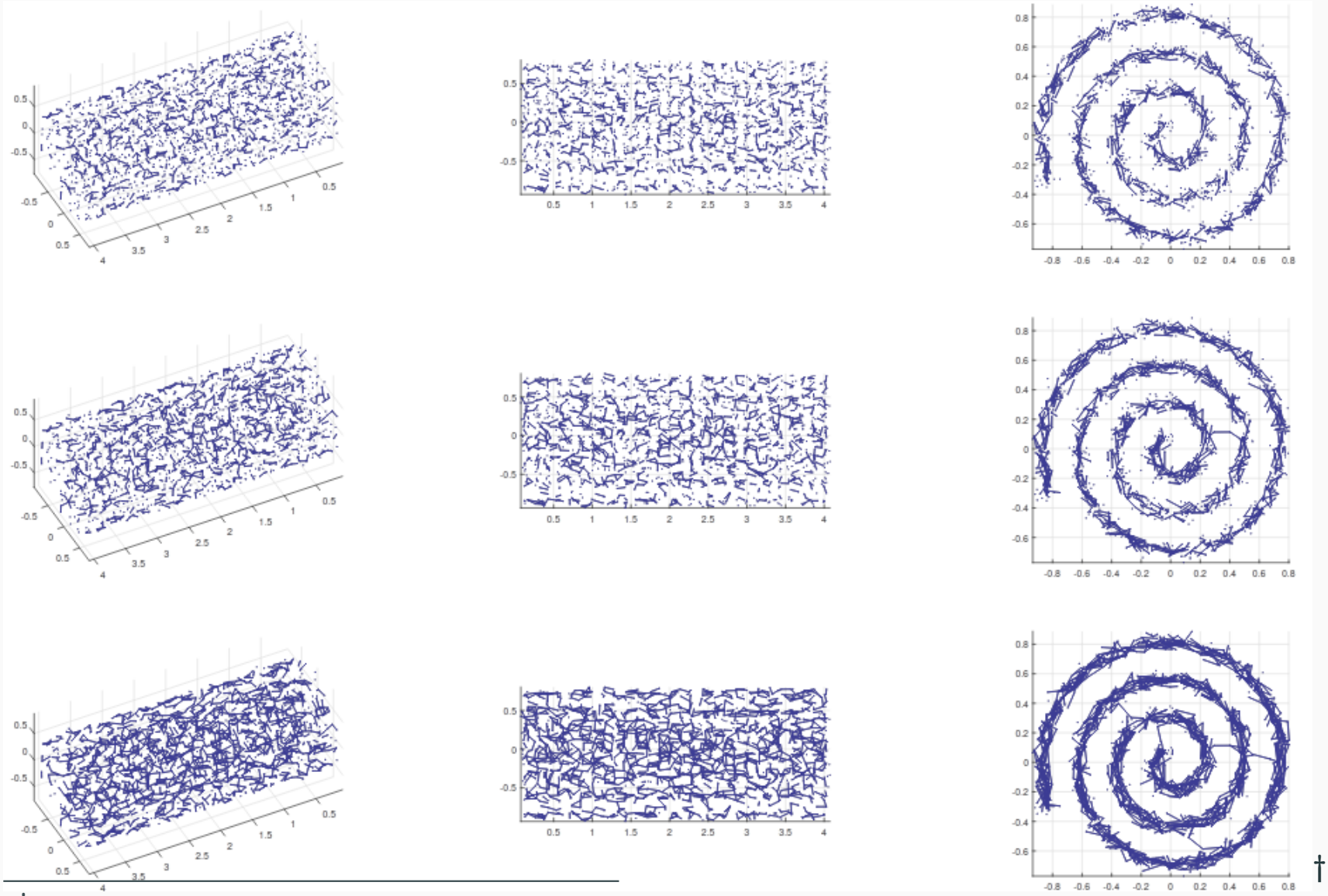
ISOMAP: Algoritmo

1. Calcular el grafo G_ε (o G_k).

2. Calcular la matriz de distancias.

3. Calcular el embedding.

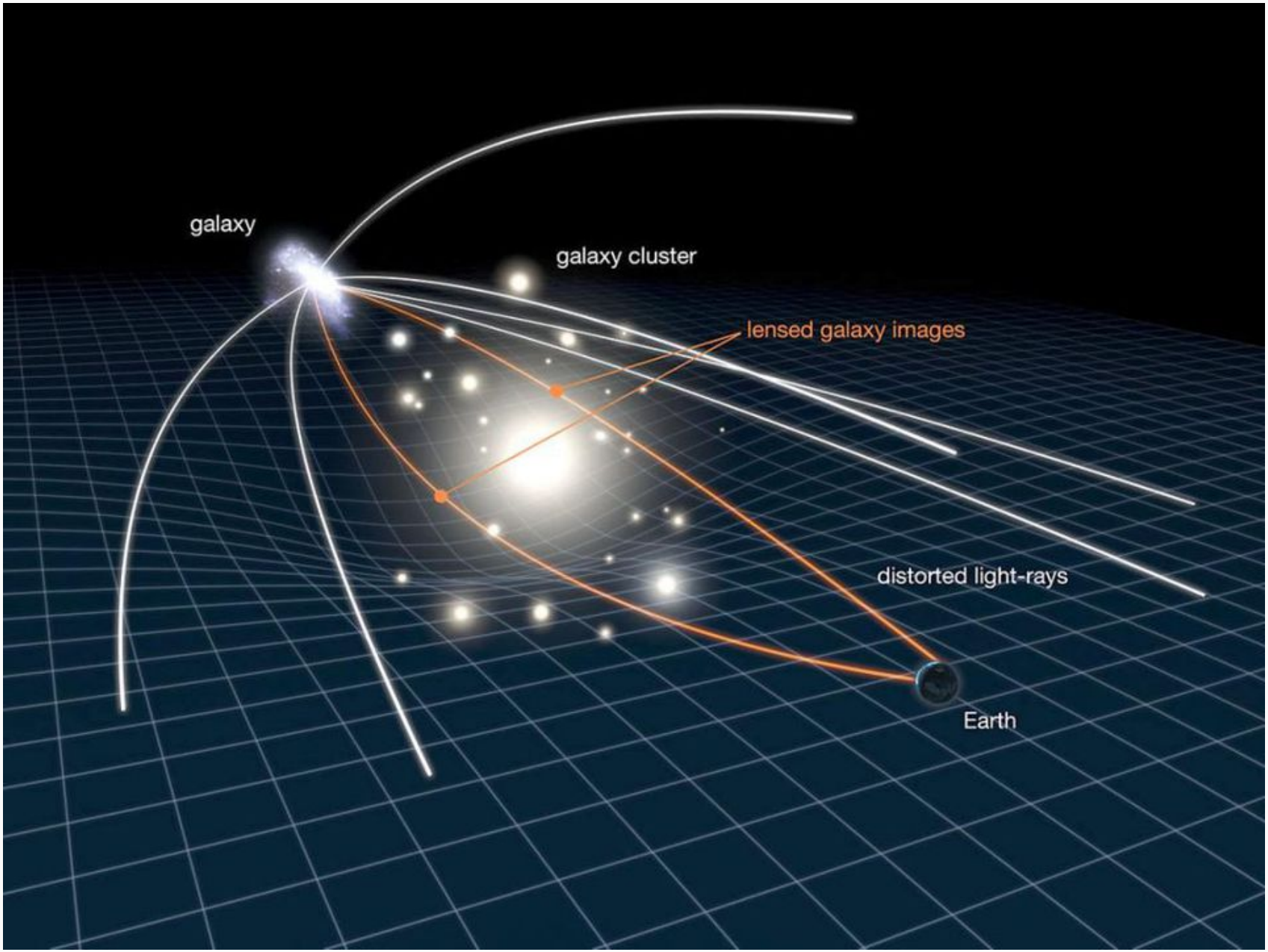
ISOMAP: Debilidades



† Imagen: B. Mahler. *Contagion Dynamics for Manifold Learning*, 2020.

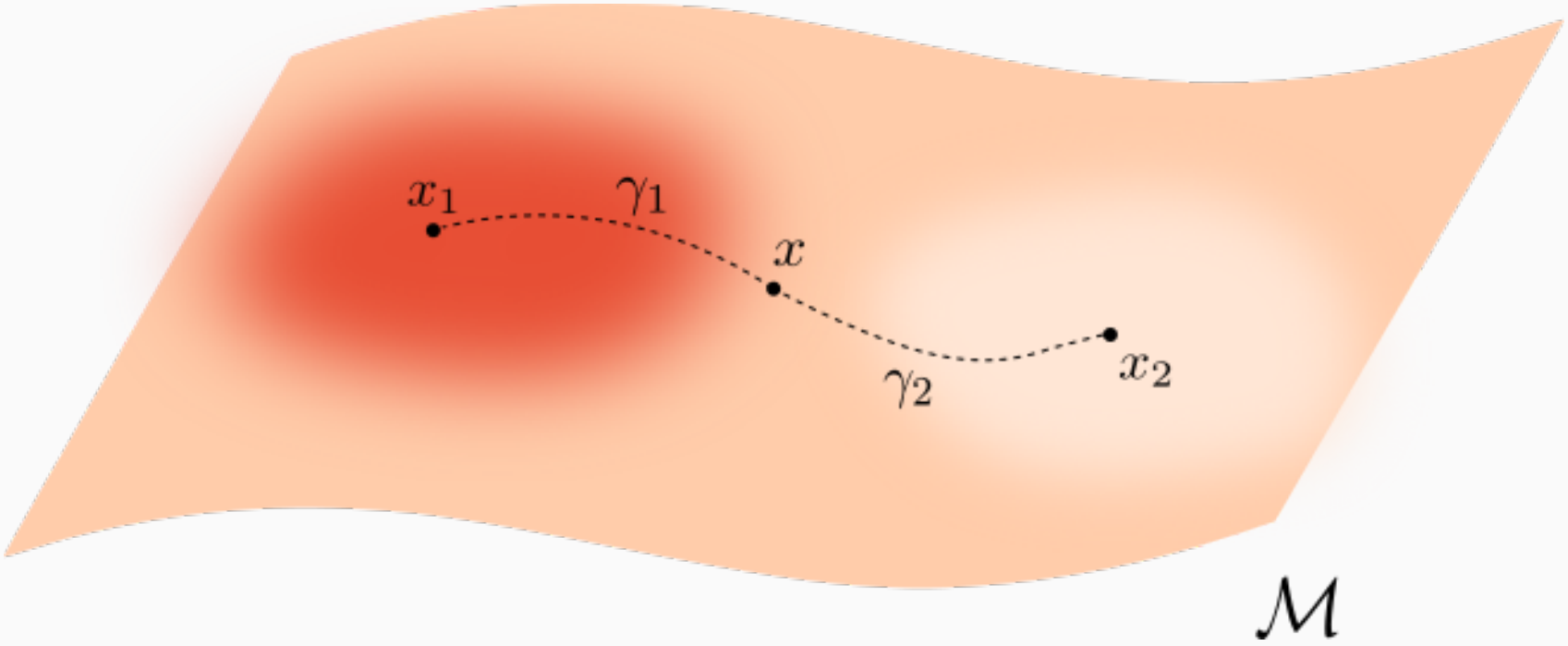
¿Cómo medir distancias basadas en densidad?

Distancias basadas en densidad



Distancias basadas en densidad

Distancias basadas en densidad



Distancia de Fermat

Sea $\mathcal{M}$ una **variedad** suave de dimensión d embebida en $\mathbb{R}^D$.

Sea $f : \mathcal{M} \to \mathbb{R}_{>0}$ una **función de densidad**.

Distancia de Fermat

Sea $\mathcal{M}$ una **variedad** suave de dimensión d embebida en $\mathbb{R}^D$.

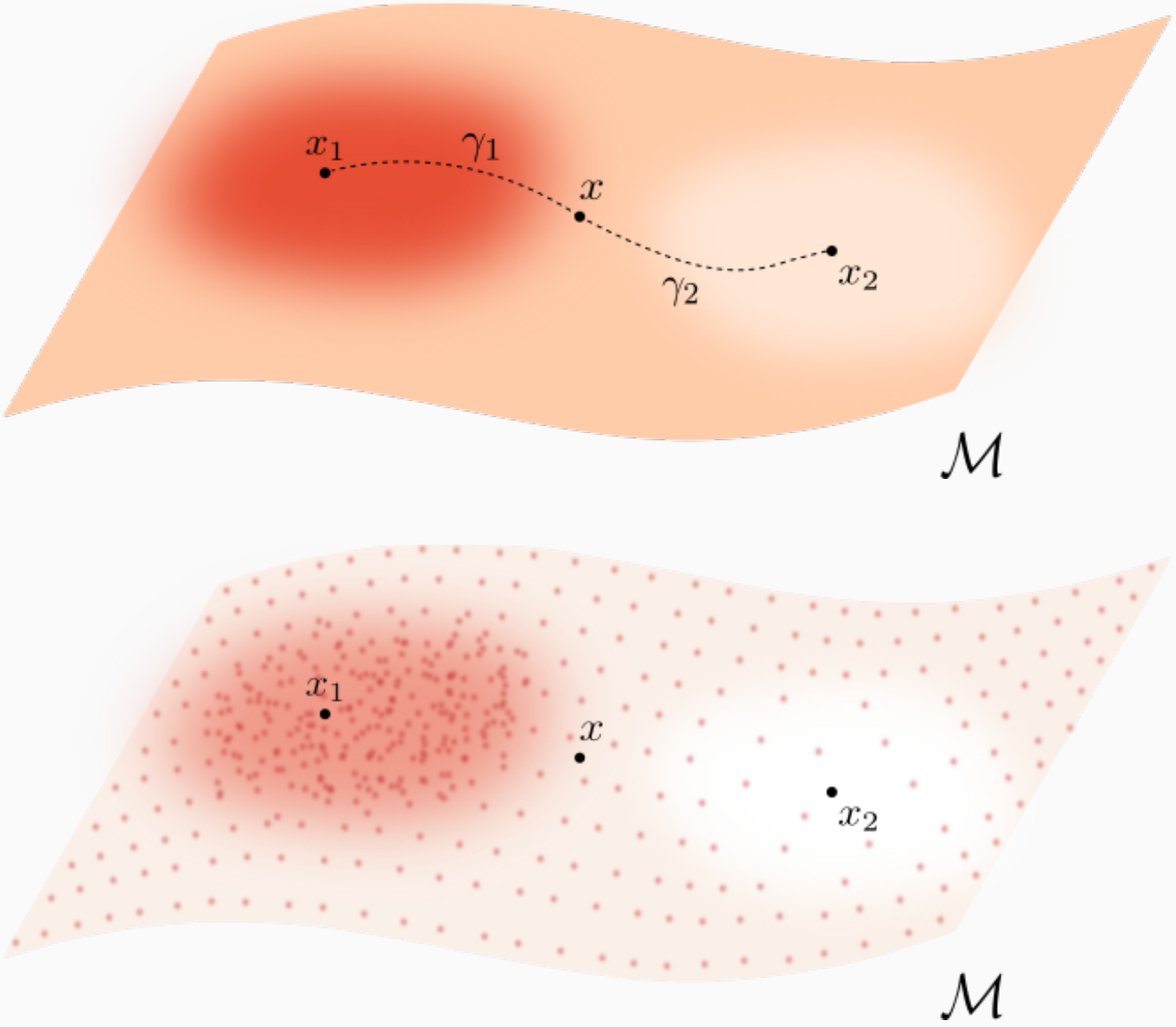
Sea $f : \mathcal{M} \to \mathbb{R}_{>0}$ una **función de densidad**.

Dado $p > 1$, se define la **distancia deformada**

$$d_{f,p}(x, y) = \inf_{\gamma} \int_I \frac{1}{f(\gamma(t))^{(p-1)/d}} ||\gamma'(t)|| dt.$$

$d_{f,p}$ se llama la ***p*-distancia de Fermat** en analogía con el Principio de Fermat en óptica.

Aprendizaje de distancias basadas en densidad

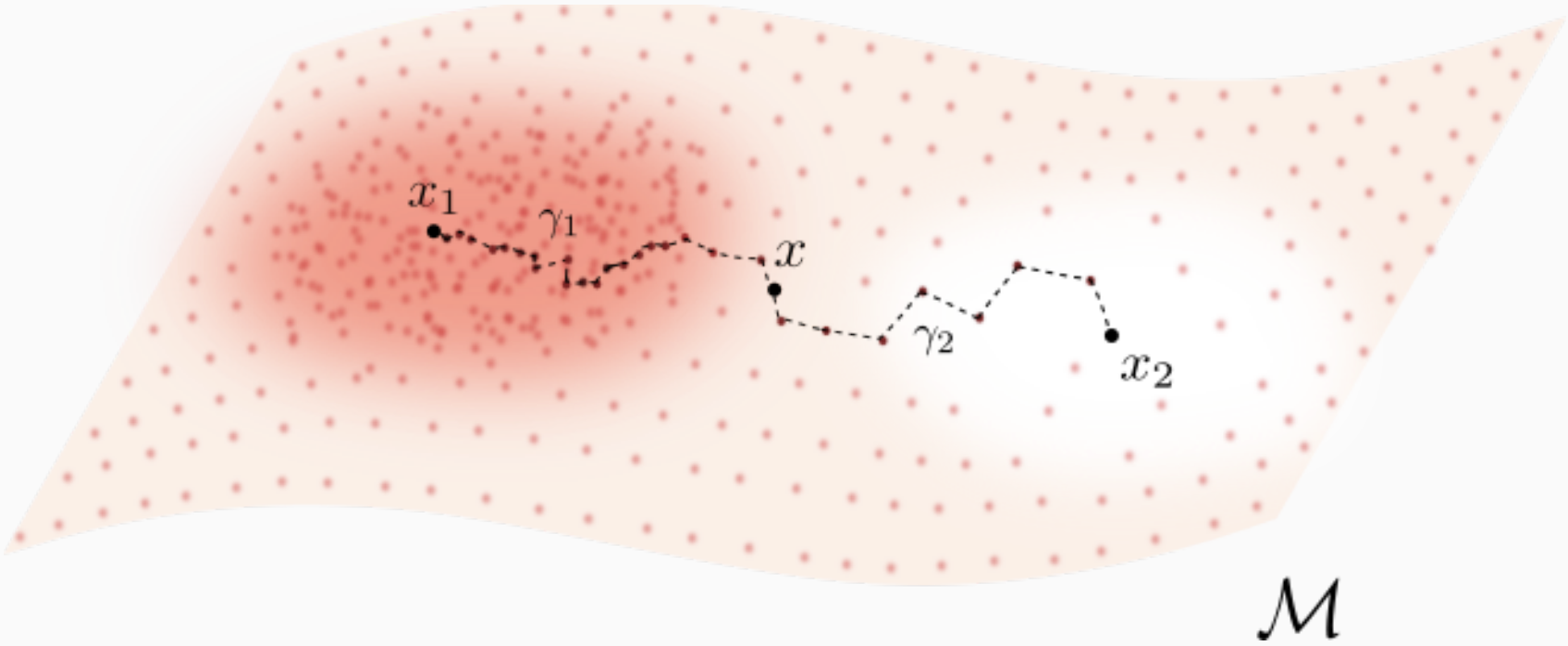


Distancia de Fermat muestral

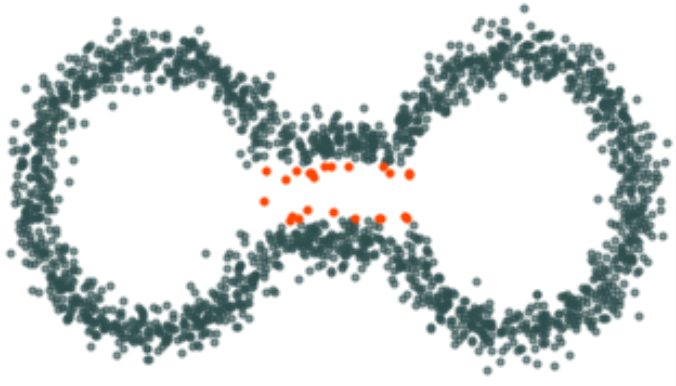
Dado $p > 1$, la **distancia de Fermat muestral** entre x, y se define por

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

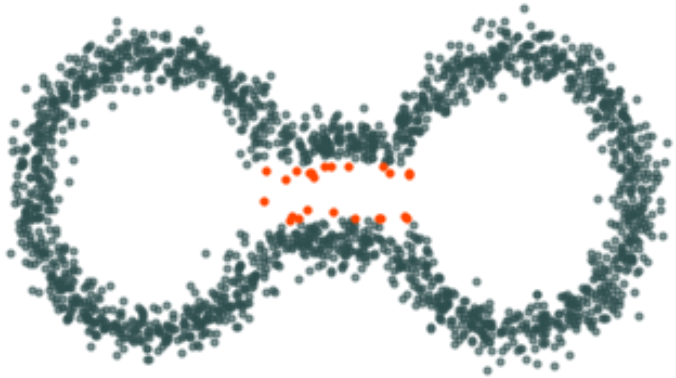
donde el ínfimo es sobre todos los caminos $\gamma = (x_0, \ldots, x_{r+1})$ con $x_0 = x$, $x_{r+1} = y$ y $\{x_1, x_2, \ldots, x_r\} \subseteq \mathbb{X}_n$.

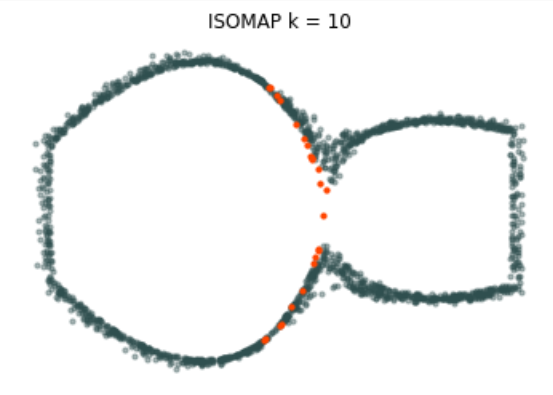


Ejemplo

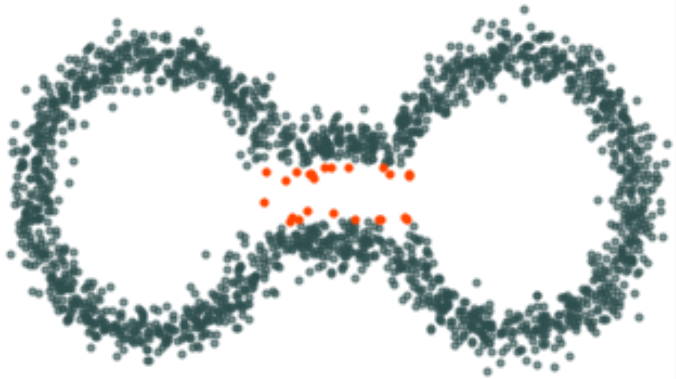


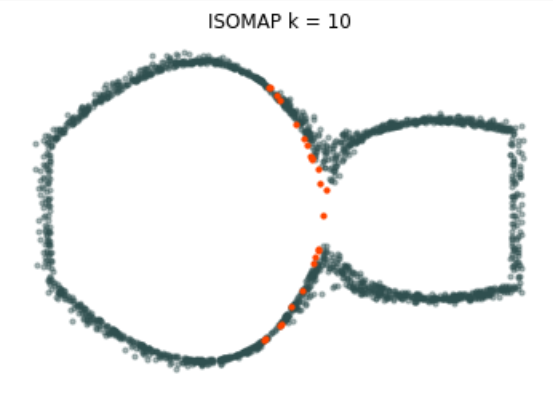
Ejemplo

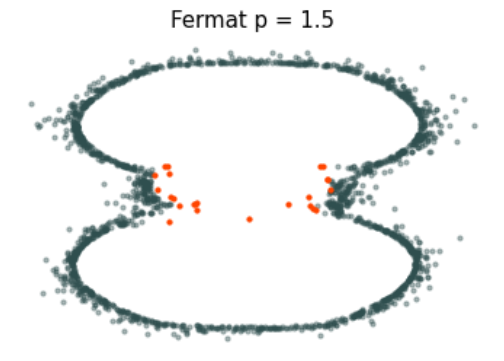




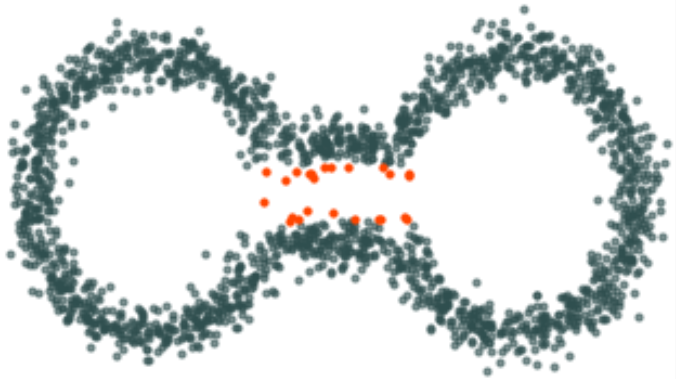
Ejemplo

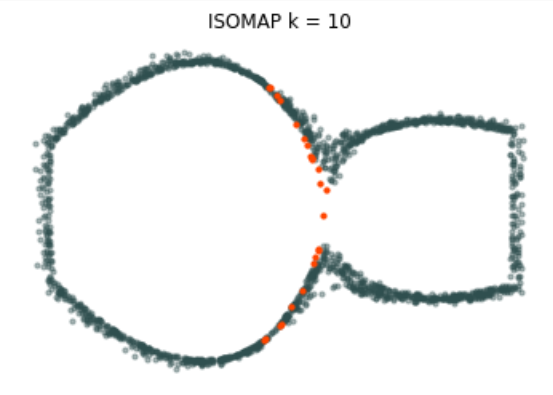


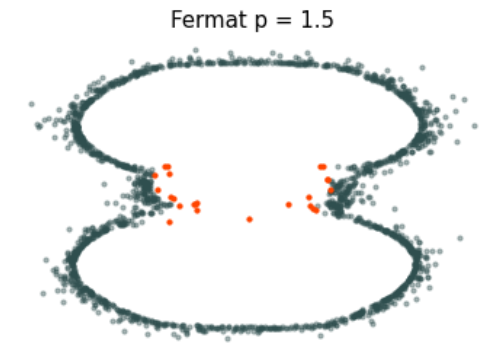


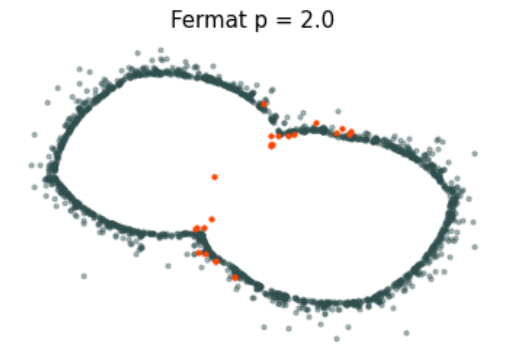


Ejemplo

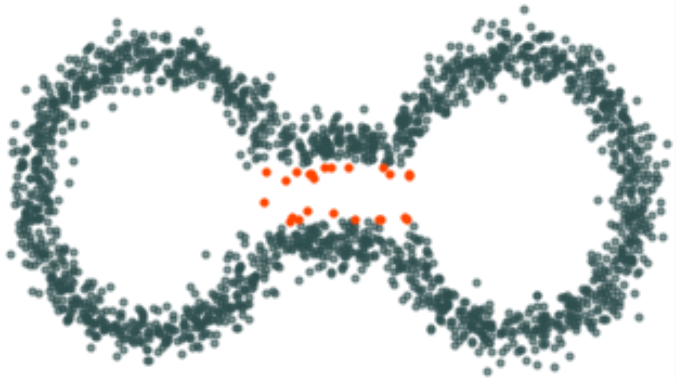


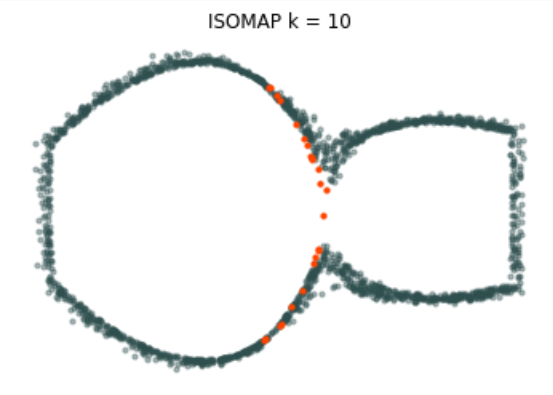


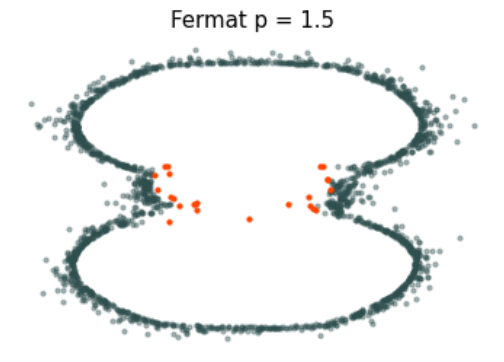


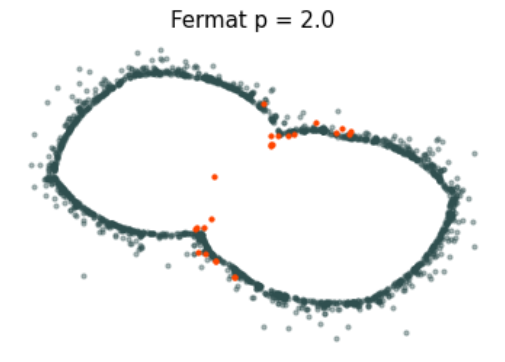


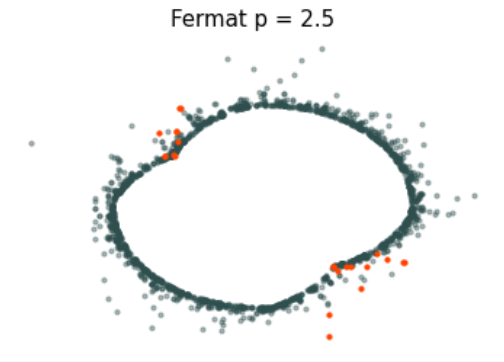
Ejemplo



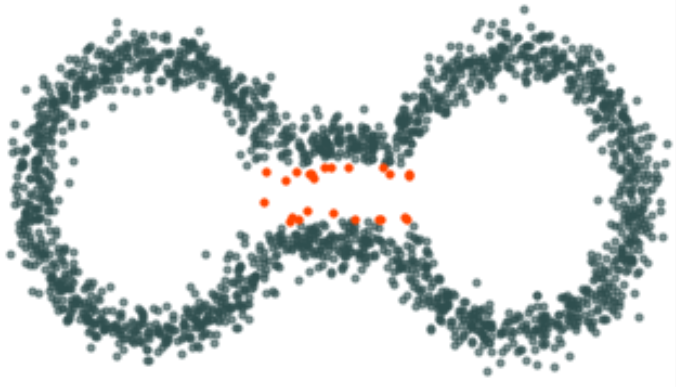


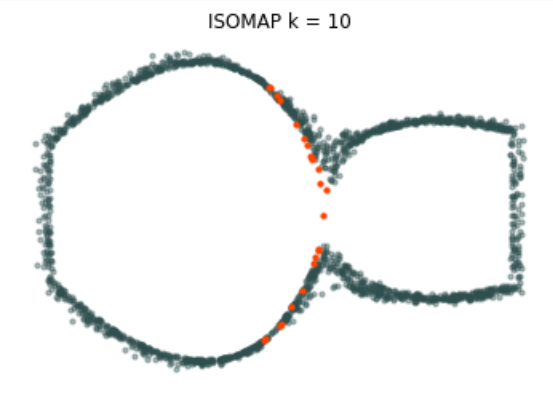


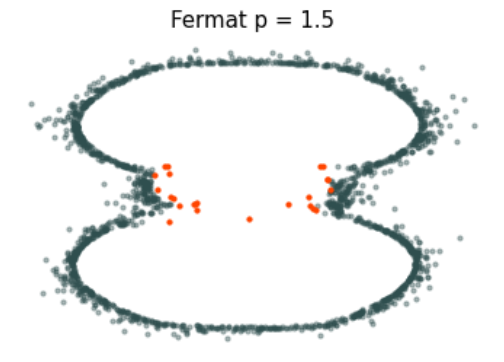


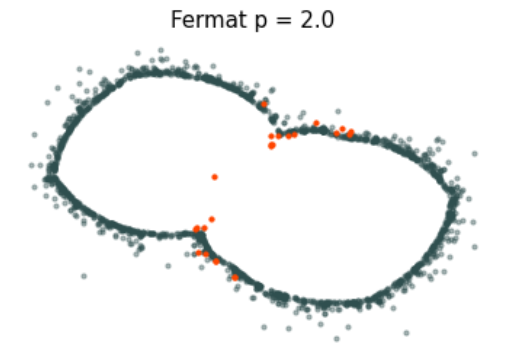


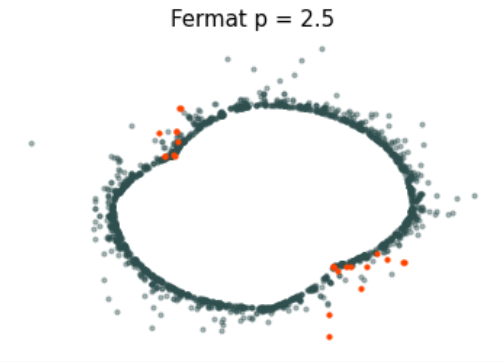
Ejemplo

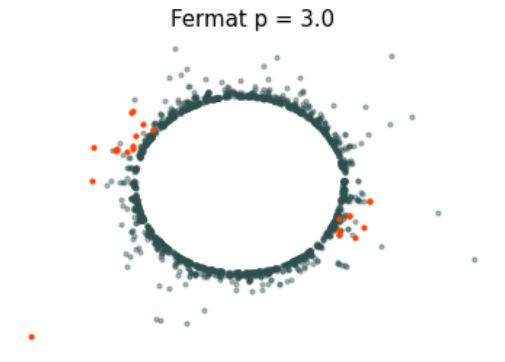








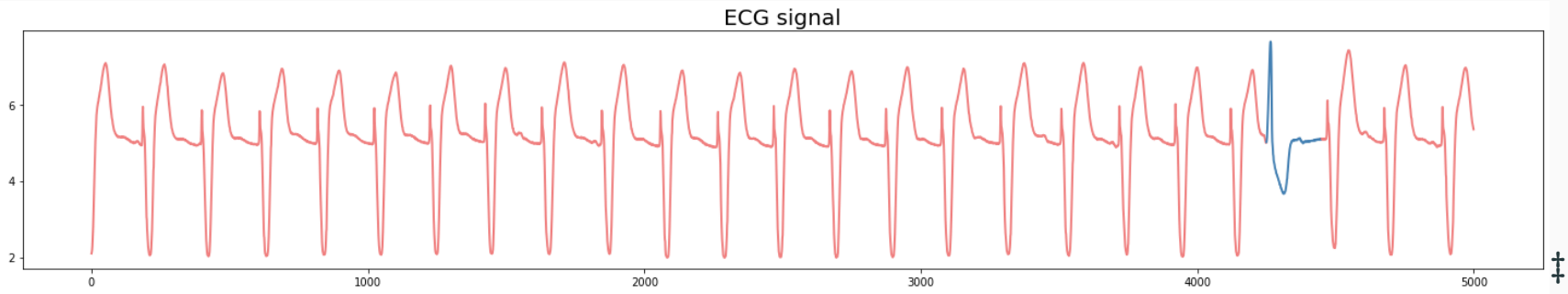




Aplicaciones

Detección de anomalías en series temporales

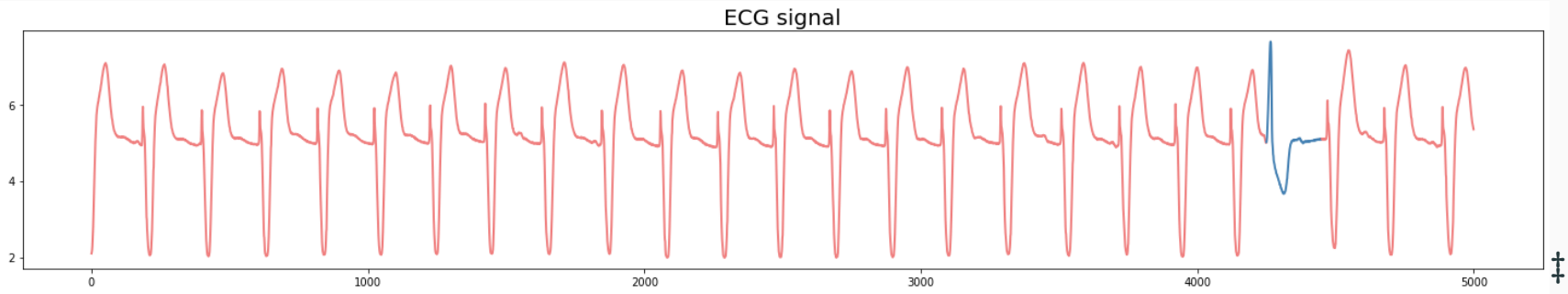
Señal de electrocardiograma con un latido anómalo (arritmia).

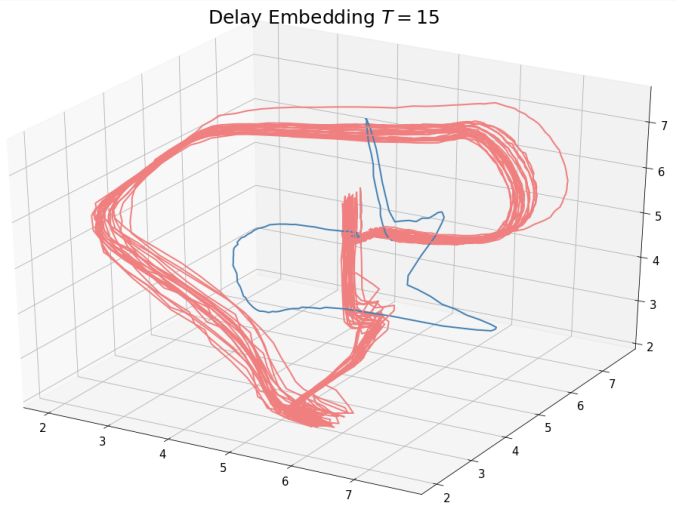


‡Datos de la base de datos Physionet, MIT Laboratorio de Fisiología Computacional.

Detección de anomalías en series temporales

Señal de electrocardiograma con un latido anómalo (arritmia).

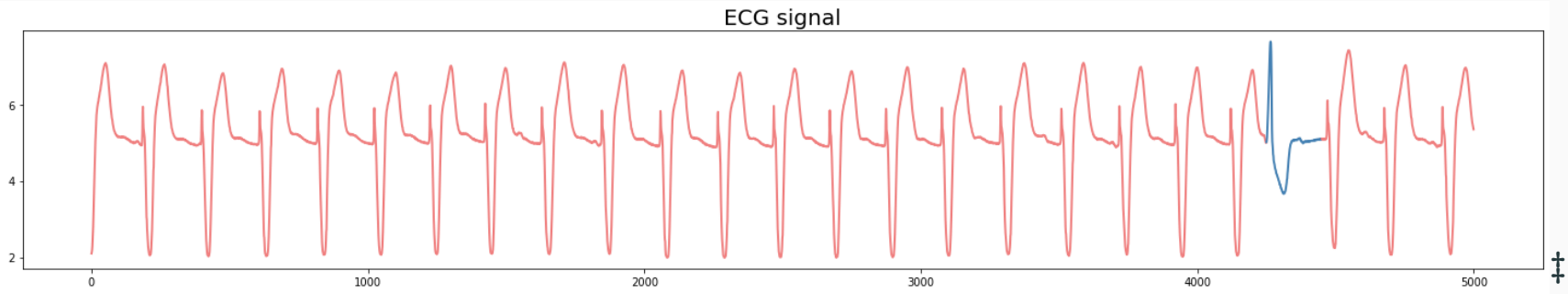




‡Datos de la base de datos Physionet, MIT Laboratorio de Fisiología Computacional.

Detección de anomalías en series temporales

Señal de electrocardiograma con un latido anómalo (arritmia).



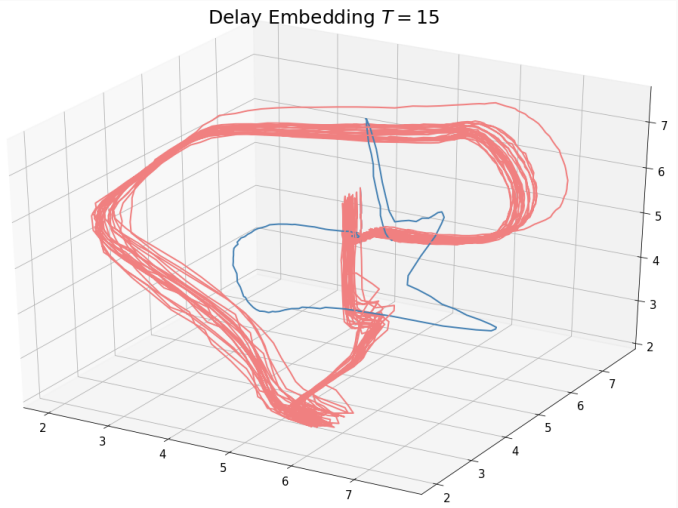
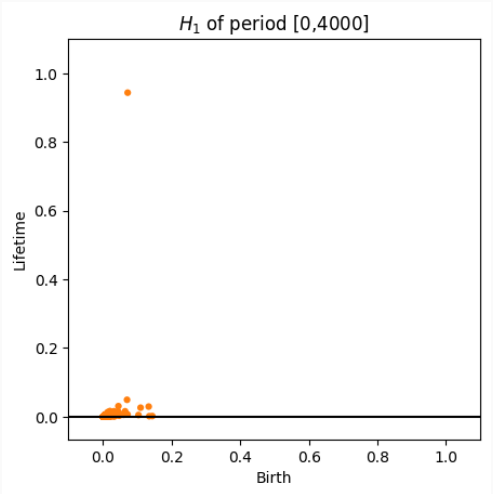


Diagrama de persistencia con distancia de Fermat $p = 2$.



‡Datos de la base de datos Physionet, MIT Laboratorio de Fisiología Computacional.

Detección de anomalías en series temporales

Señal de electrocardiograma con un latido anómalo (arritmia).

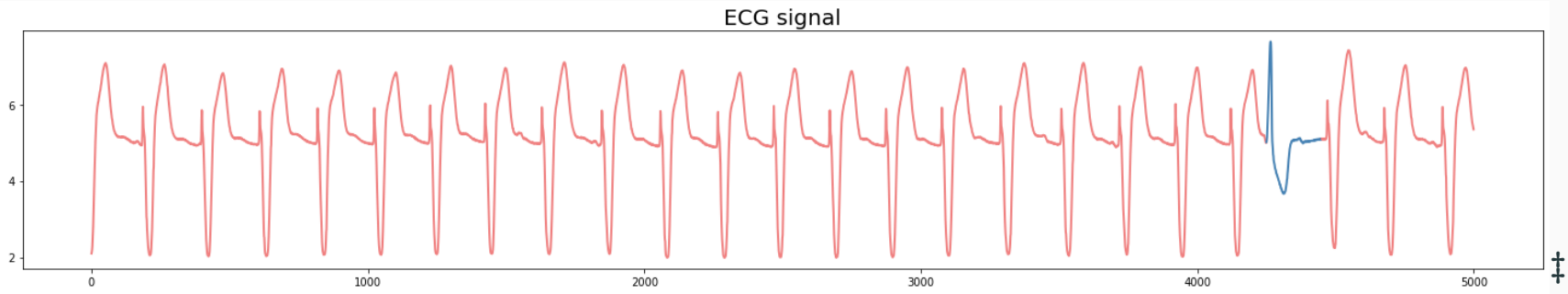
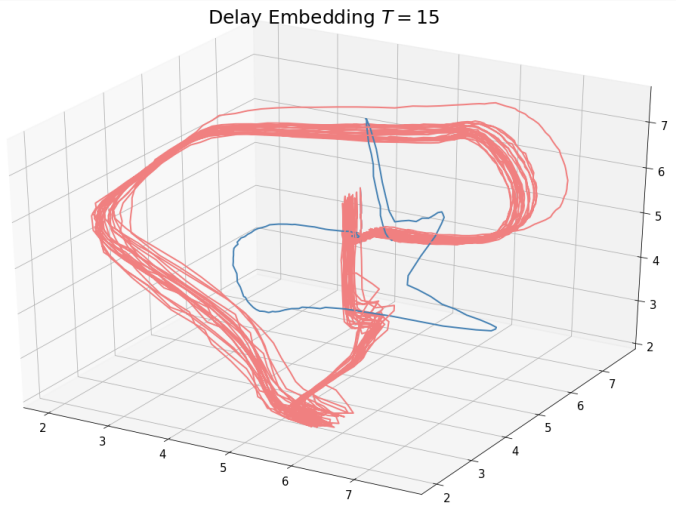
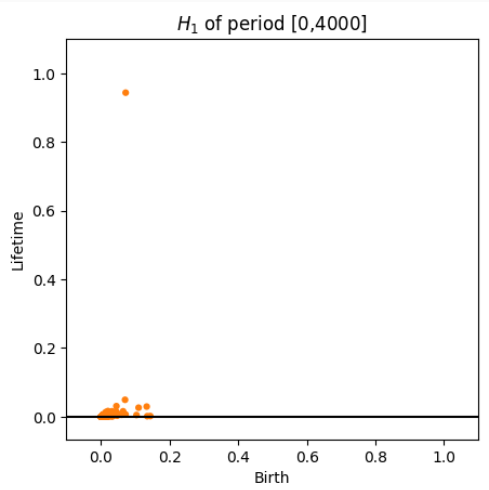
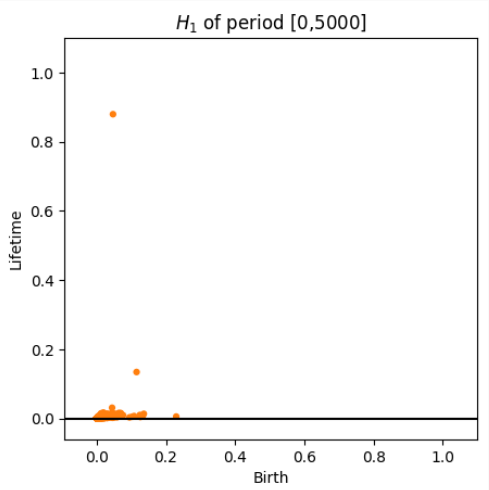


Diagrama de persistencia con distancia de Fermat $p = 2$.







‡Datos de la base de datos Physionet, MIT Laboratorio de Fisiología Computacional.

Referencias

- J. B. Tenenbaum. *A Global Geometric Framework for Non-linear Dimensionality Reduction* (2000).
- M. Bernstein, V. D. Silva, J. C. Langford, J. B. Tenenbaum. *Graph approximations to geodesics on embedded manifolds* (2000).
- E. Borghini, X. Fernández, P. Groisman, G. Mindlin. *Intrinsic persistent homology via density-based distance learning.* arXiv:2012.07621 (2020).

web: `https://ximenafernandez.github.io/`

github: `ximenafernandez`

email: `x.l.fernandez@swansea.ac.uk`

Gracias!