TDA I

Université Lyon 2

2024/2025

Rémi Vaucher

HALIAS & ERIC Lab.

Table des matières

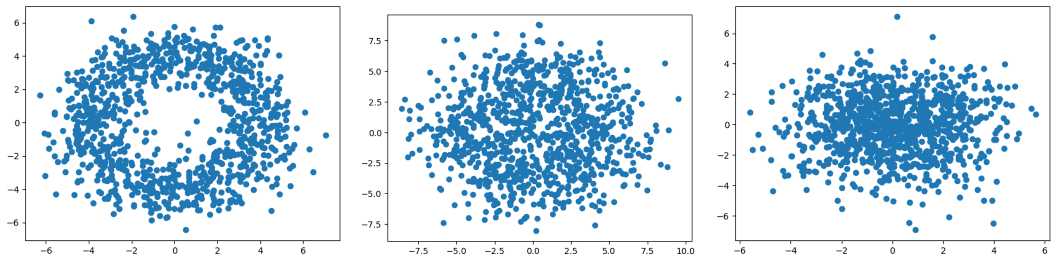
Table of Contents

"Data has shape"

- La forme d'un jeu de données porte des informations cruciales sur le jeu en lui même.
- Il est difficile d'approcher visuellement la forme des données en dimensions élevées.
- Une croyance: les données sont portées par une **variété lisse (ou non)**.

Exemple 1

Comment caractériser la différence entre ces 3 échantillons:



Le manifold learning

N'est ce pas le but du manifold learning?

Le manifold learning

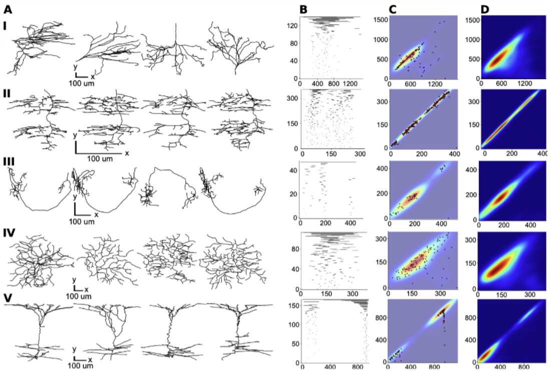
N'est ce pas le but du manifold learning?
Non. Le manifold learning regroupe des techniques de réduction de dimension en se restreignant à une structure de variété. Le but ici est autre:

Quantifier des invariants de forme au travers des dimensions successives.

Deux exemples motivés

Pour trouver des exemples dans lequel la TDA est utile il suffit de trouver des exemples de données ou la forme est importante de manière évidente:

- Classification des morphologies neuronales.



Deux exemples motivés

- Classification des cristaux nanoporeux.

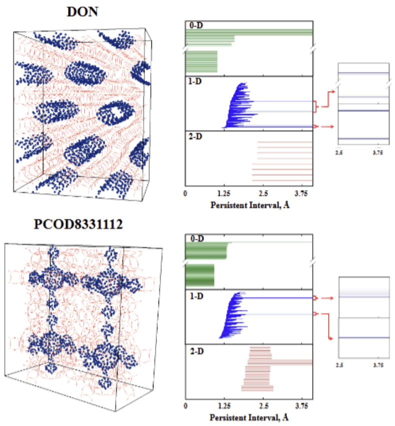


Table of Contents

Définition

Un k-simplexe est l'enveloppe convexe de $k+1$-points.

Figure 1: 0-simplex

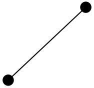


Figure 2: 1-simplex

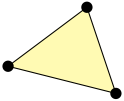


Figure 3: 2-simplex

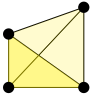


Figure 4: 3-simplex

2

Cette définition n'est valable que dans un **espace vectoriel** : une enveloppe convexe nécessite l'existence d'un principe de segment.
Pour les autres cas, on privilégieras une version ensembliste.

Complexe simplicial

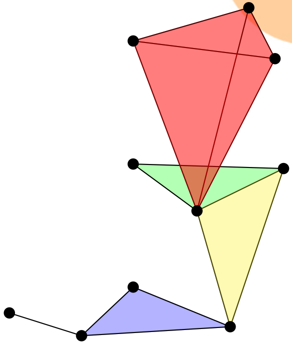
Définition

On considère un ensemble de "sommets" (généralement des données) $V = \{v_1, ..., v_N\}$. Un **complexe simplicial** sur V est une collection de simplexes dont les sommets sont des éléments de V.

De même, l'intersection de deux simplexes doit être un sous-simplexe.

Exemple:

Un complexe simplicial de dimension 3.



Complexe de Vietoris-Rips / Čech

Définition

Soit $V = \{v_1, ..., v_n\}$ un ensemble de sommets. Un complexe de **Vietoris-Rips** à hauteur d'ϵ est un complexe simplicial tel que

$$[v_{i_1} v_{i_2} ... v_{i_k}] \subset \mathcal{C} \iff B(v_{i_j}, \epsilon) \cap B(v_{i_l}, \epsilon) \neq \varnothing \quad \forall 1 \leq j \neq l \leq k$$

Complexe de Vietoris-Rips / Čech

Définition

Soit $V = \{v_1, ..., v_n\}$ un ensemble de sommets. Un complexe de **Čech** à hauteur d'ϵ est un complexe simplicial tel que

$$[v_{i_1} v_{i_2} ... v_{i_k}] \subset \mathcal{C} \iff B(v_{i_1}, \epsilon) \cap B(v_{i_2}, \epsilon) \cap \cdots \cap B(v_{i_k}, \epsilon) \neq \varnothing$$

Exemple Vietoris-Rips

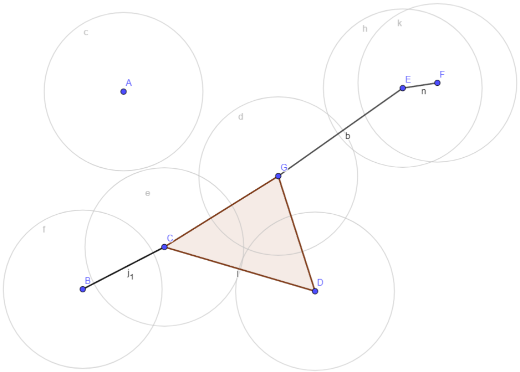


Figure 5: Complexe de Vietoris-Rips au niveau alpha

Exemple de Čech

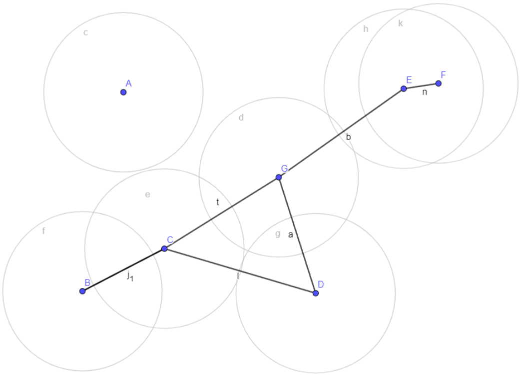


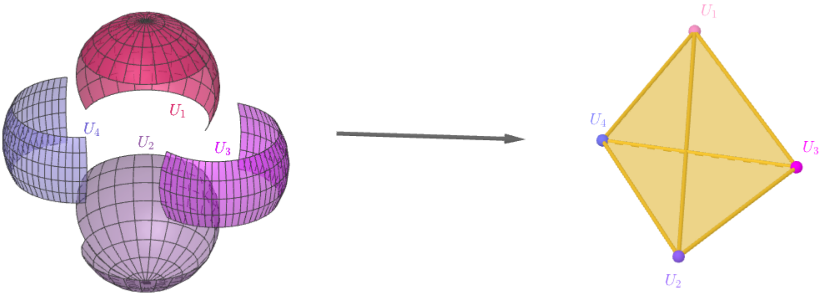
Figure 6: Complexe de Čech au niveau alpha

Nerf d'un recouvrement

Définition

On considère $\mathcal{U} = \{U_i\}$ un recouvrement d'ouvert d'une variété $\mathbb{M}$. Le **nerf de ce recouvrement** est le complexe simplicial tel que:

- Tout sommet v_i du nerf est un ouvert U_i
- Le simplexe $[v_1 ... v_k]$ est dans le complexe ssi $U_{i_1} \cap \cdots \cap U_{i_k} \varnothing$


U_1
U_4
U_2
U_3
U_1
U_4
U_3
U_2