

Introduction to TDA

Mazino

25 March 2024

1 Introduction

The idea of Topological Data Analysis (TDA) is to use topological methods to analyze data. The dataset is often represented as a point cloud in a high-dimensional space. The goal is to extract topological features from the data that are invariant under continuous transformations. These features can then be used to understand the underlying structure of the data and make predictions.

TDA is based on the idea that the shape of the data can reveal important information about its structure. For example, the presence of holes in the data can indicate clusters or other interesting patterns. By analyzing the topology of the data, we can gain insights that are not apparent from traditional statistical methods.

One of the key tools in TDA is the concept of a simplicial complex. A simplicial complex is a set of simplices glued together by their faces. Simplices can be thought of as higher-dimensional generalizations of triangles. By studying the topology of the simplicial complex, we can extract topological features from the data.

In this course, we will introduce the basic concepts of TDA and show how they can be applied to analyze real-world datasets.

2 Introduction to Topology

Topology is the branch of mathematics that studies the properties of geometric objects that are preserved under continuous transformations.

Définition 2.1 (Topological Space): A topological space is a set X together with a collection of subsets of X called open sets, noted \mathcal{T} , that satisfy the following axioms:

1. \emptyset and X are open.
2. The union of any collection of open sets is open.
3. The intersection of any finite collection of open sets is open.

Example 2.1: The real line \mathbb{R} is a topological space with the standard topology. The open sets are intervals of the form (a, b) where $a < b$.

The Euclidean space \mathbb{R}^n is a topological space with the standard topology. The open sets are open balls of the form $B(x, r) = \{y \in \mathbb{R}^n : \|x - y\| < r\}$ where $x \in \mathbb{R}^n$ and $r > 0$.

The discrete topology on a set X is the topology where every subset of X is open.

The trivial topology on a set X is the topology where only \emptyset and X are open.

Définition 2.2 (Closed Set): A set A is closed if its complement $X \setminus A$ is open.

Définition 2.3 (Interior): The interior of a set A , is the union of all open sets contained in A . So the interior of A , noted $\text{Int}(A)$, is the largest open set contained in A .

Définition 2.4 (Closure): The closure of a set A is the intersection of all closed sets containing A . So the closure of A , noted \overline{A} , is the smallest closed set containing A .

Définition 2.5 (Boundary): The boundary of a set A , noted ∂A , is the set of points that are in the closure of A but not in the interior of A .

$$\partial A = \overline{A} \setminus \text{Int}(A)$$

Définition 2.6 (Neighborhood): A neighborhood of a point x in a topological space X is an open set containing x . We can also see a neighborhood of x is a set U such that $x \in U$ and U is open and $x \in \text{Int}(U)$.

The first definition of a neighborhood is more general, as it does not require the set to be open.

Définition 2.7 (Subspace Topology): Let X be a topological space and let $Y \subseteq X$. We define the subspace topology on Y as :

$$\mathcal{T}_Y = \{Y \cap U : U \in \mathcal{T}\}$$

Exemple 2.2: Example for \mathbb{R}^n :

1. the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$.
2. the unit cube $C^{n-1} = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n : \max(|x_i|) = 1\}$.
3. The open balls $B(x, r) = \{y \in \mathbb{R}^n : \|x - y\| < r\}$.
4. The closed balls $\bar{B}(x, r) = \{y \in \mathbb{R}^n : \|x - y\| \leq r\}$.
5. The standard simplex $\Delta^{n-1} = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0, \sum_{i=1}^n x_i = 1\}$.

2.1 Homeomorphisms

To make a parallel with geometry, in geometry we study transformations that preserves the "shape" of an object, through translations, rotations, etc.

In topology, we are more flexible and we allow more general transformations that preserve the "topological" properties of a space, such as stretch, shrink, bend, twist, scale, etc. You are not allow to cut the object or glue it to another object, or make holes / fill holes.

These transformations are called homeomorphisms.

Définition 2.8 (Continuous Map): A map $f : X \rightarrow Y$ between topological spaces is continuous if the preimage of every open set in Y is open in X .

Exemple 2.3: Let $X = Y = \mathbb{R}$ with the standard topology. The map $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = 0$ if $x \leq 0$ and $f(x) = x$ if $x > 0$ is continuous.

The set $\{0\}$ is closed in \mathbb{R} but its preimage $f^{-1}(\{0\}) = (-\infty, 0]$ is not closed in \mathbb{R} . So the map f is not continuous.

Définition 2.9 (Homeomorphism): Let (X, \mathcal{T}_X) and (Y, \mathcal{T}_Y) be topological spaces. A homeomorphism between X and Y is a bijective map $f : X \rightarrow Y$ such that both f and f^{-1} are continuous. If there exists a homeomorphism between X and Y , we say that X and Y are homeomorphic and we write $X \cong Y$.

Example 2.4: Let's consider $\mathbb{S}(0, 1) = \{x \in \mathbb{R}^n : \|x\| = 1\}$ the unit circle and $\mathbb{S}(0, 2) = \{x \in \mathbb{R}^n : \|x\| = 2\}$ the circle of radius 2. We can take the map :

$$\begin{aligned} f : \mathbb{S}(0, 1) &\longrightarrow \mathbb{S}(0, 2) \\ x &\longmapsto 2x \end{aligned}$$

It is continuous and bijective, and its inverse is the map $x \mapsto \frac{x}{2}$ which is also continuous. So the two circles are homeomorphic.

Example 2.5: $\mathbb{S}^2 \not\cong \mathbb{R}^2$. The sphere encloses a 3D space, or void, while the plane does not.

What we mean by "encloses" is that the 2-sphere has "air" inside it, the "inside" is a 3D space (length, depth, width) even if the sphere itself is 2D (longitudinal and latitudinal coordinates).

In contrast, the plane is a 2D space, but it does not enclose or contain anything.

So the sphere is not homeomorphic to the plane. They interact differently with the 3D space around them.

But if we take $\mathbb{R}^2 \cup \{\infty\}$ the plane with a point at infinity, then it is homeomorphic to the sphere.