Examples of Topological Data Analysis and Persistent Homology

Massimiliano Ferrara

1 Introduction

Topological Data Analysis (TDA) is a framework for analyzing the shape of data, often using techniques from algebraic topology. One of the key tools in TDA is persistent homology, which captures the topological features of a dataset across multiple scales.

2 Example 1: Simplicial Complex

Let's consider a set of points sampled from a circle in 2D space. Suppose we have the following points:

$$
\begin{aligned}
P_1 &= (1, 0) \\
P_2 &= (0, 1) \\
P_3 &= (-1, 0) \\
P_4 &= (0, -1) \\
P_5 &= (0.5, 0.5)
\end{aligned}
$$

2.1 Constructing a Simplicial Complex

- Choose a distance threshold $\epsilon = 0.7$.
- Connect points that are closer than this threshold. For example, P_1 and P_5 are connected, and so on.
- This creates a network of edges, forming a simplicial complex.

2.2 Constructing a Rips Complex

- For each ϵ, build a Rips complex R_ϵ, which includes vertices for each point and edges for every pair of points within distance ϵ.

2.3 Calculating Homology Groups

Calculate the homology groups H_0 (connected components) and H_1 (loops) for different values of ϵ.

3 Example 2: Persistence Diagrams

Following the previous example, let's compute the persistence diagram.

3.1 Varying ϵ

- Compute homology groups as ϵ increases (e.g., $\epsilon = 0.1, 0.2, \ldots, 1.0$).
- At $\epsilon = 0.3$, you might find:

$$\begin{aligned} H_0 &= 1 \quad \text{(all points connected)} \\ H_1 &= 1 \quad \text{(the circular loop remains)} \end{aligned}$$

3.2 Birth and Death of Features

Track when features appear (birth) and disappear (death).

- For example, the circle component (in H_1) might be born at $\epsilon = 0.3$ and die (become trivial) at $\epsilon = 0.8$.

3.3 Persistence Diagram

Each feature is represented as a point in the persistence diagram, where:

- The x-coordinate is the birth time.
- The y-coordinate is the death time.

The persistence of a feature can be calculated as:

$$\text{persistence} = \text{death} - \text{birth}$$

4 Example 3: Real Dataset

Consider using TDA on a real dataset, such as digit images from the MNIST dataset.

4.1 Data Preprocessing

- Reduce the dimensionality using techniques like PCA (Principal Component Analysis).

4.2 Computing Persistence

- Create a point cloud from the feature vectors.
- For each image, build a Rips complex and calculate persistence.

4.3 Analysis

- It can be useful identifying significant topological features that can help distinguish between different digits based on their shapes.