Topological Data Analysis

Introduction

Motiejus Valiunas

IM UWr, Summer Semester 2022/23

28 February 2023

Why TDA?

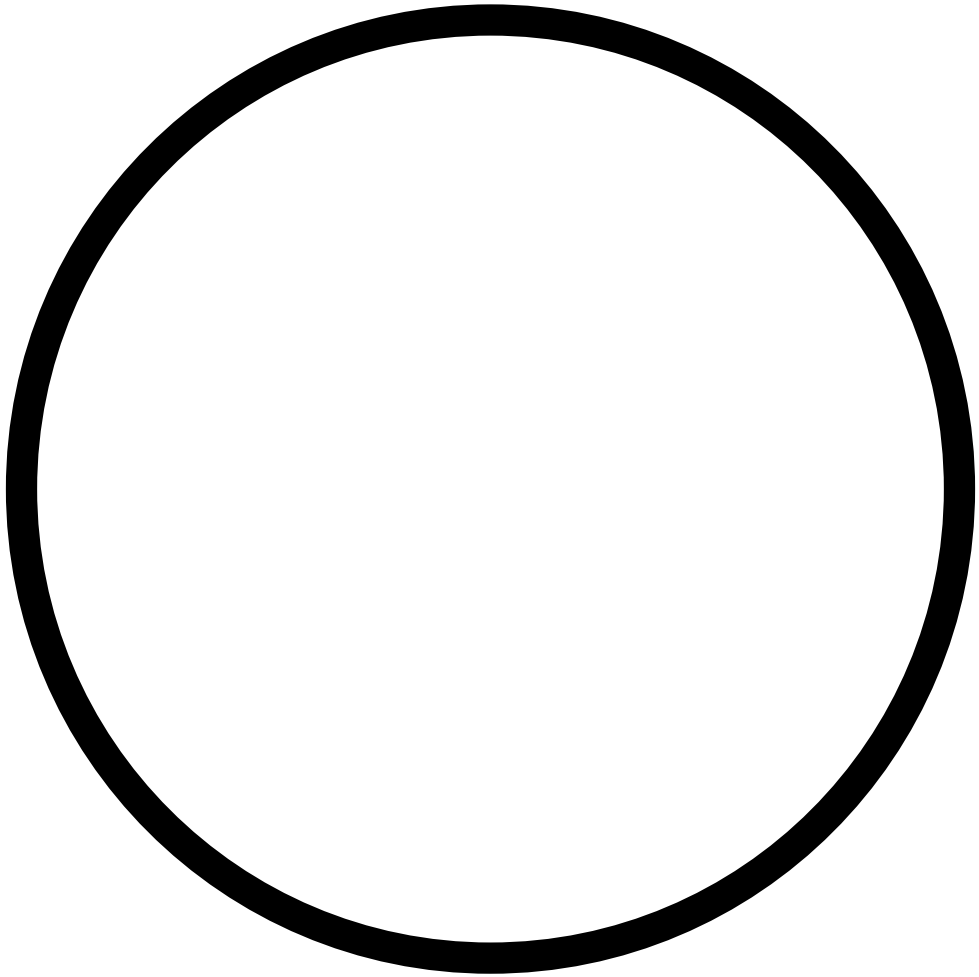
TDA is the application of tools from topology, geometry, algebra and other areas of mathematics to study the shape of data.

"Shape Matters"

- How to assemble discrete data into global geometric structures?
- What geometric properties of physical/biological/etc systems can we extract from data?
- How can we study high dimensional structures from low dimensional representations?

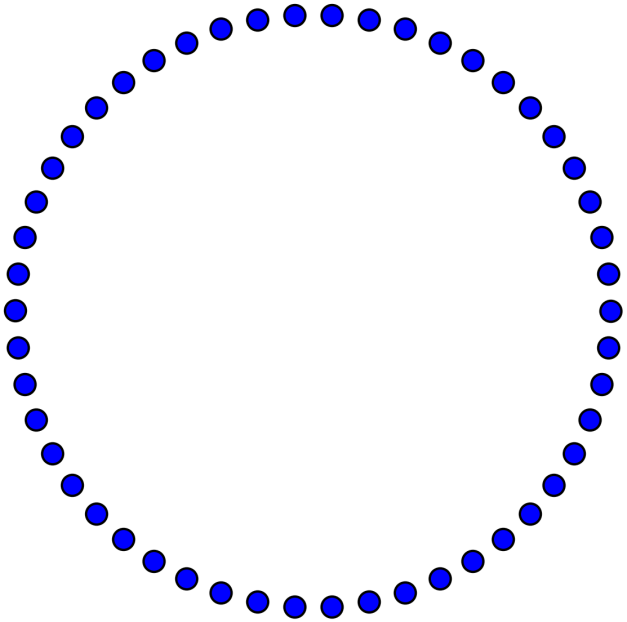
In this course, we will be concentrating on a technique called *persistent homology*.

Analysing shape: mathematics

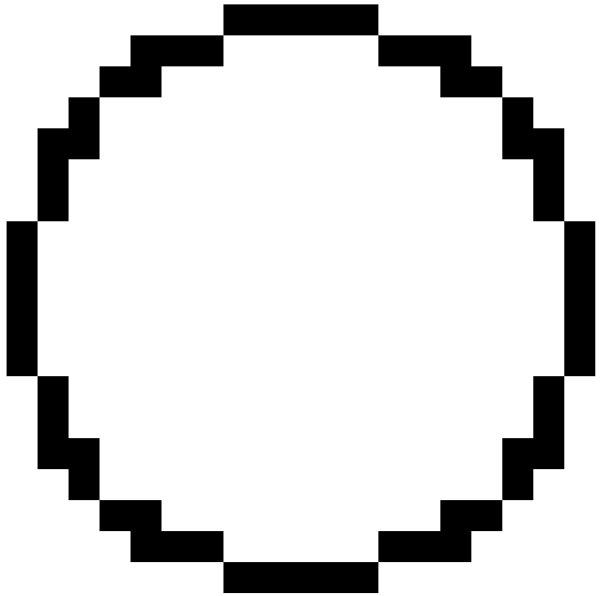


A circle: has 1 connected component and 1 “hole”.

Analysing shape: ideal data

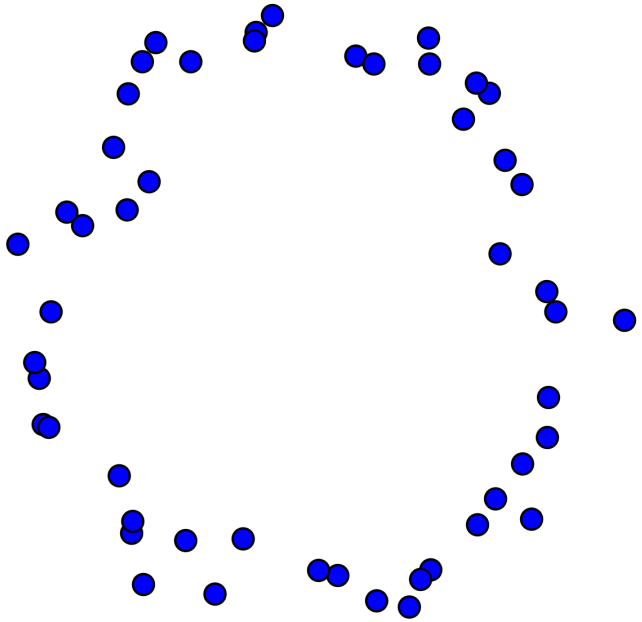


A point cloud.

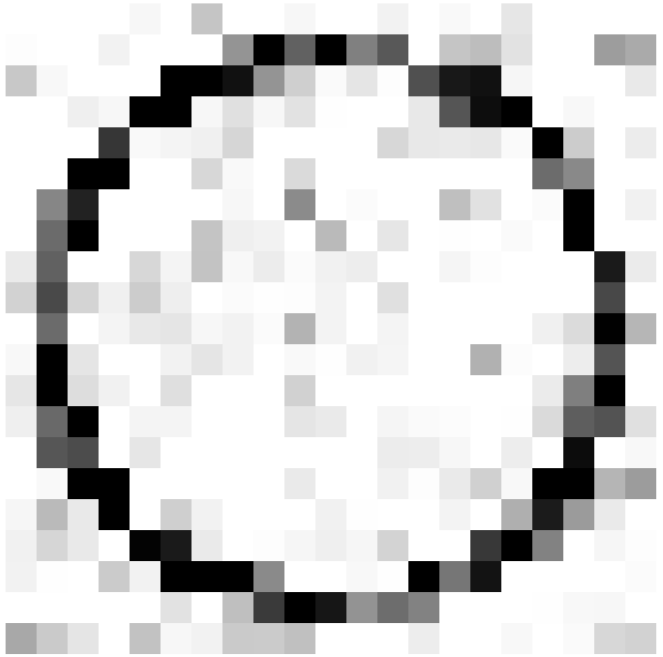


A picture.

Analysing shape: real data



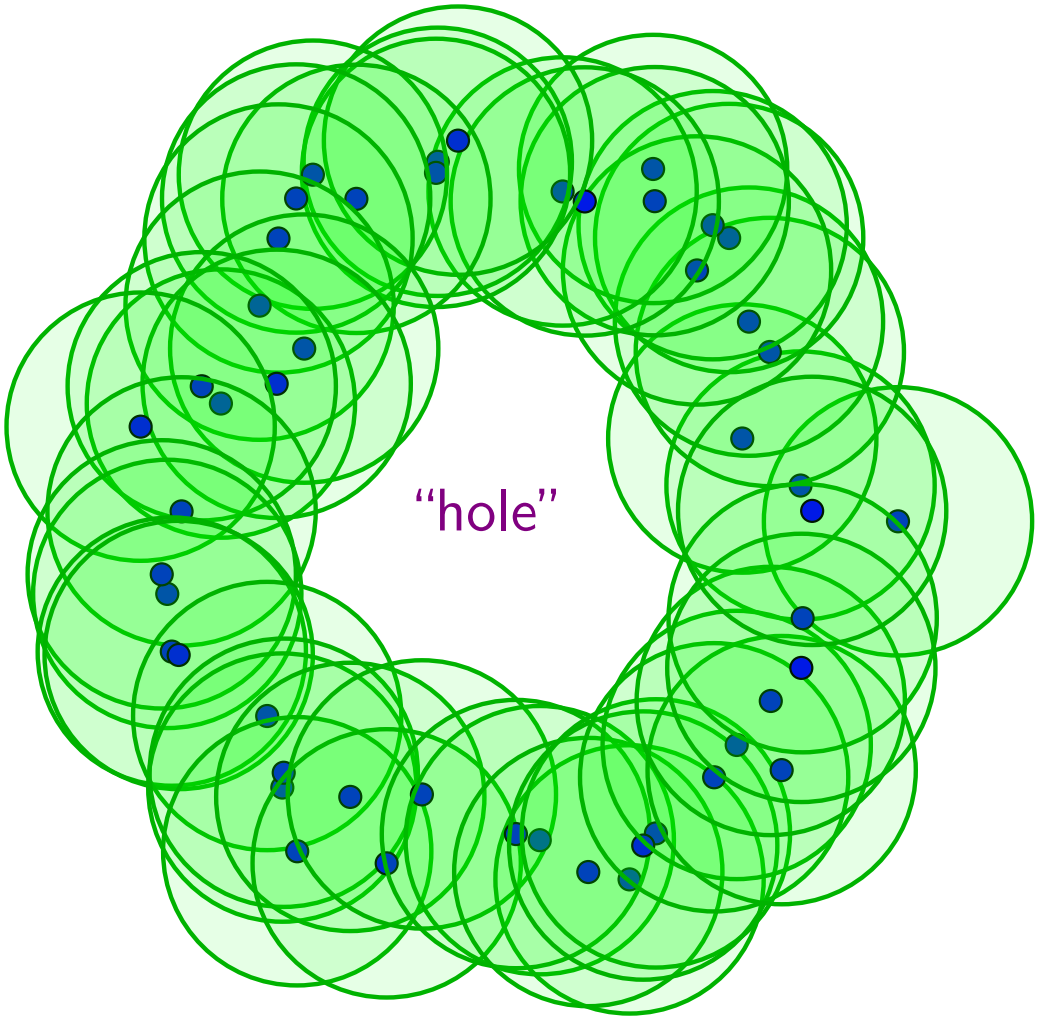
A point cloud.

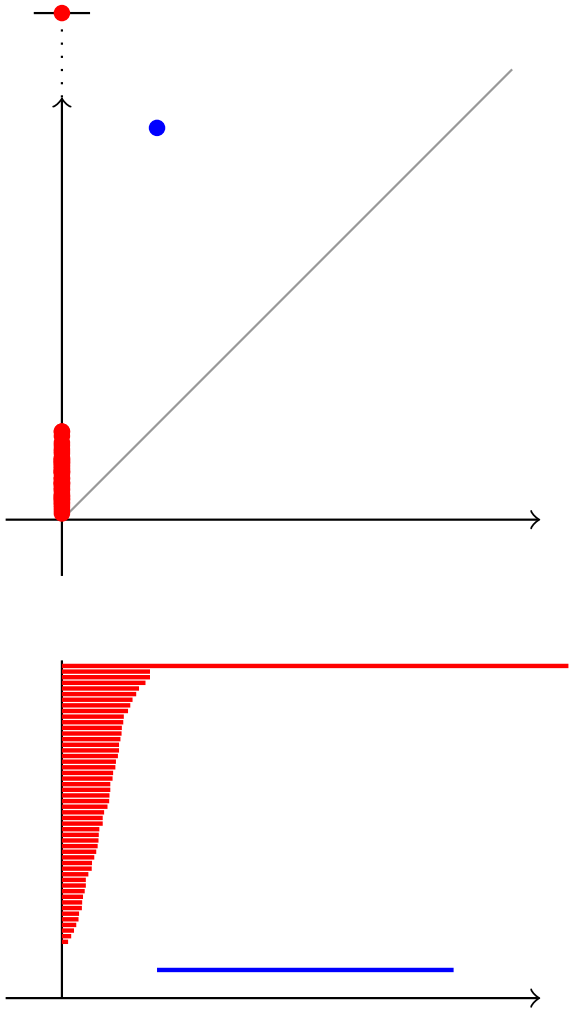


A picture.

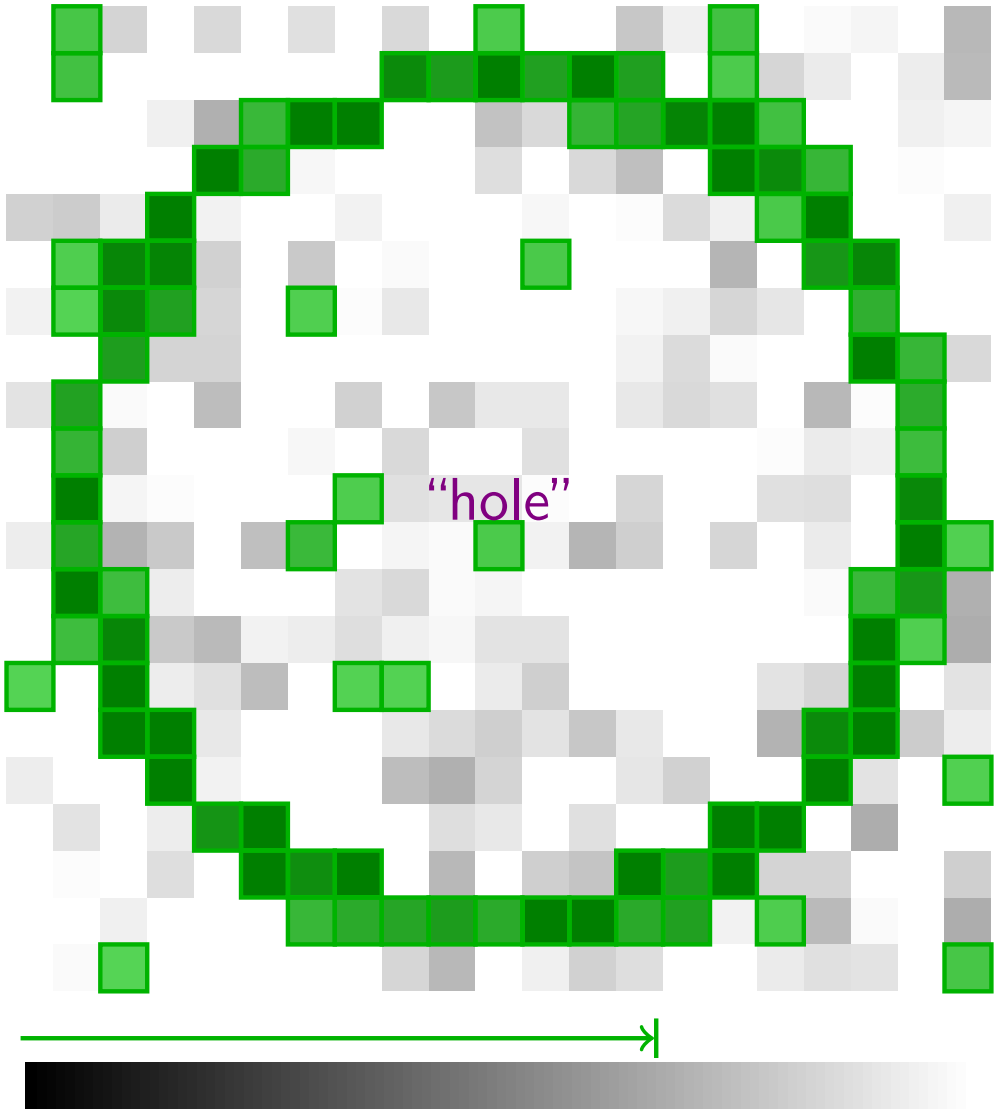
How can we recognise the properties of a circle from this?

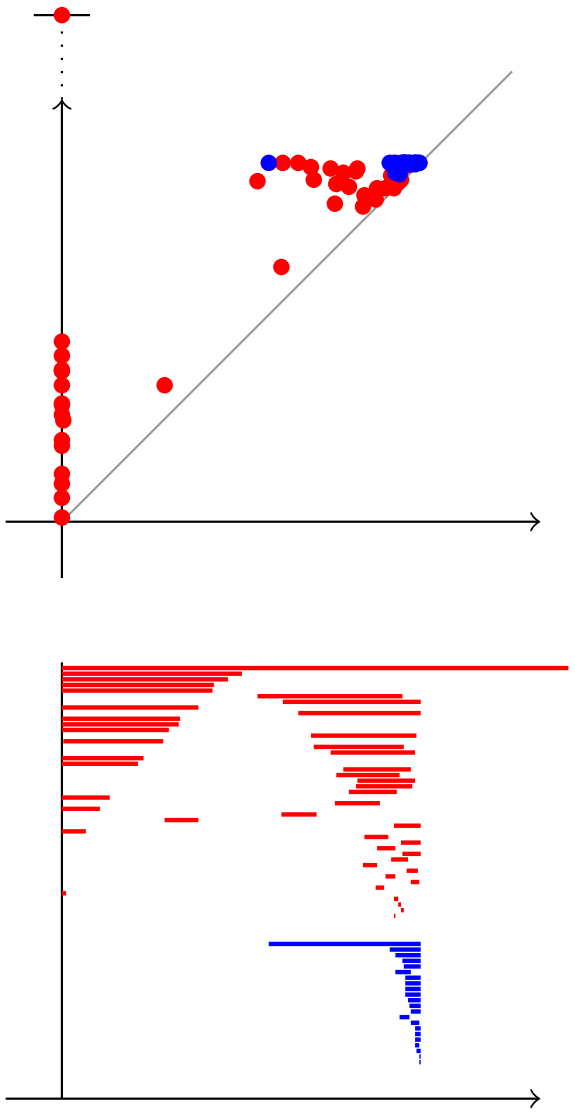
Čech filtration: growing balls





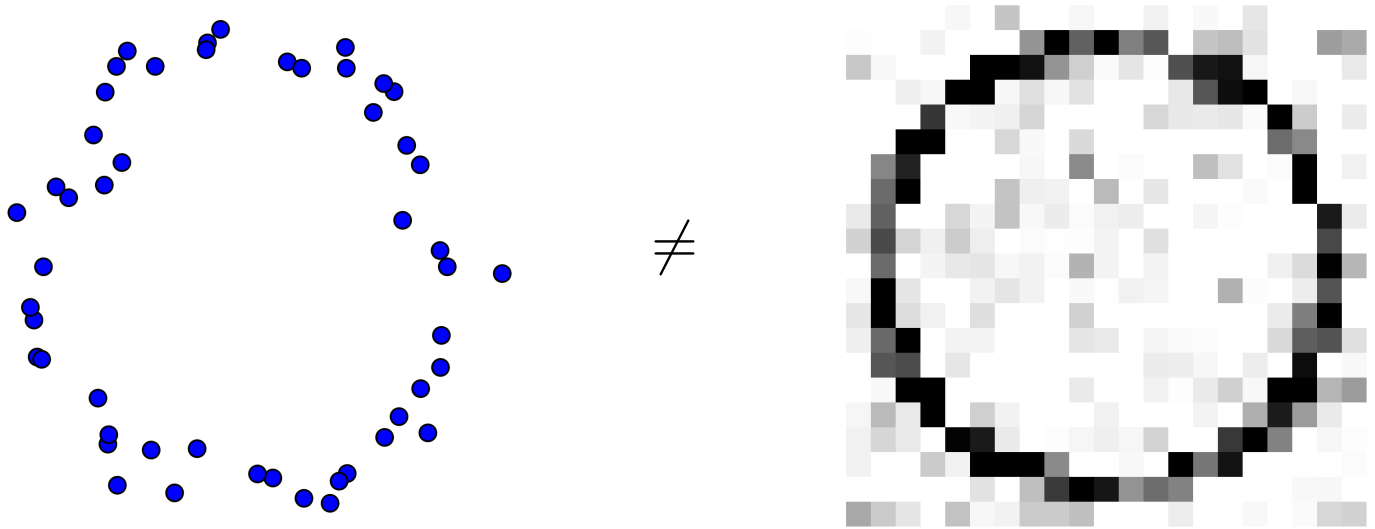
Morse filtration: reconstructing pictures



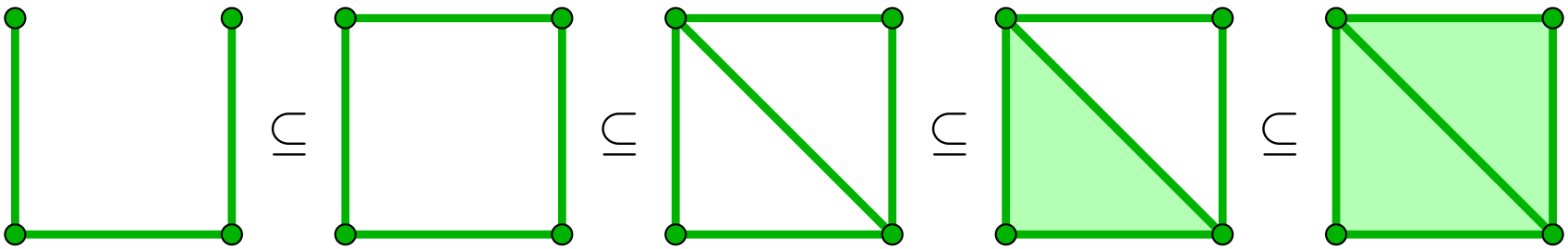


Problem 1

Q: How can we represent data uniformly in terms of a mathematical object?

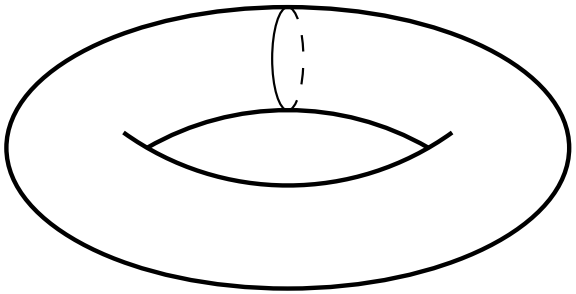


A: We use *filtrations of simplicial complexes*.

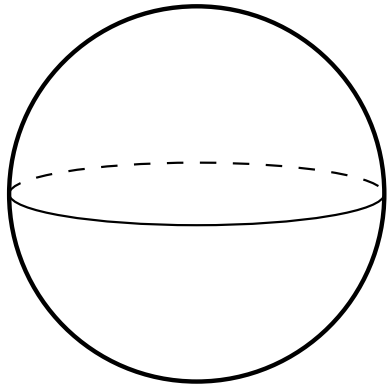


Problem 2

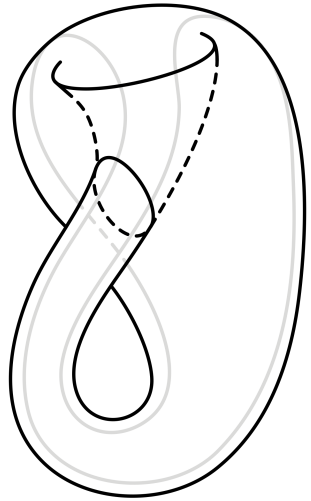
Q: How can we recognise the topological features (connected components, "holes", "voids", etc) of an object?



2 "holes"



1 "void"



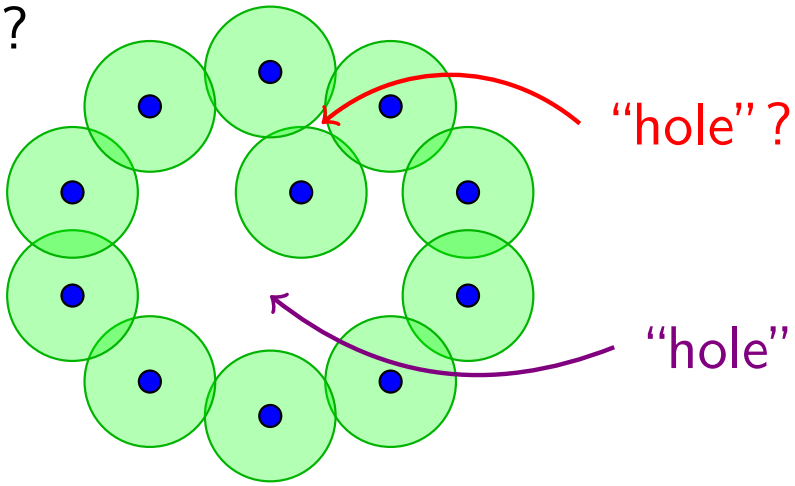
1 or 2 "holes"?

A: We use *homology*.

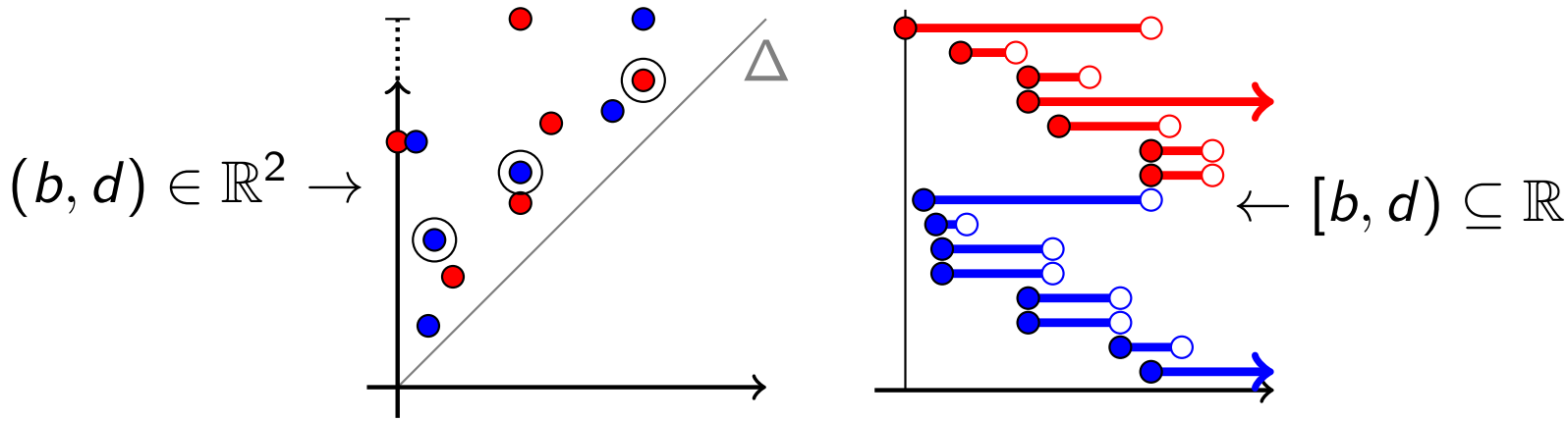
The dimension of a *homology group*, $\dim_{\mathbb{F}} H_n(X;\mathbb{F})$, represents the number of connected components ($n = 0$), "holes" ($n = 1$) or "voids" ($n = 2$) of X.

Problem 3

Q: How can we distinguish meaningful features of our data from any noise present?



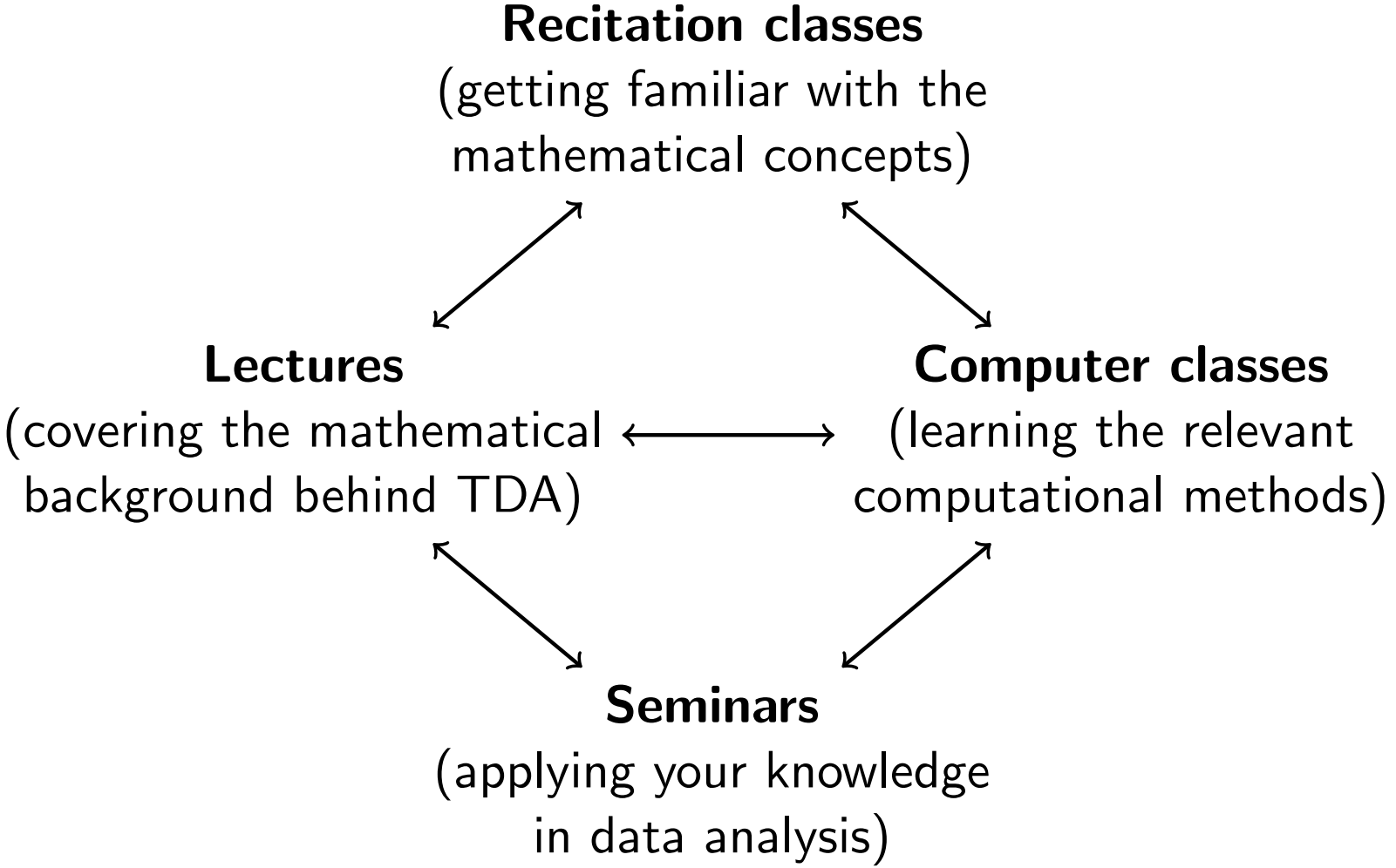
A: We use *persistence diagrams* or *barcodes*.



points far from Δ = features = long intervals

points close to Δ = noise = short intervals

Course structure



Prerequisites

The course will assume knowledge of:

- **Algebra 1:** fields and rings (especially $k[X]$ for a field k);
- **Algebra liniowa 2:** vector spaces (at least real vector spaces);
- **Analiza i topologia:** metric spaces.

All other required concepts will be briefly introduced in our lectures.