Persistent

Daniel Collin

January 27, 2021

Contents

1 Introduction

Using other measures than standard statistical analysis for datasets is an avenue that is explored in homology. From homology we can get a topological invariant of the data, interesting and can be further used as features in machine learning etc. Carlsson et al. describes it as exploring the shape of the data.

Persistent homology follows a basic principle that we can approximate a simplicial complex on a point cloud. This of course requires that the field is a metric space, because etc..

This thesis will serve as both an introduction the workings of persistent homology as well as (an?) example of persistent homology applied to a real dataset. Of course, since it is still quite a young field it is not entirely possible to answer to us what actual value persistent homology has as a tool within data science.

Etc etc.

2 Homology

Before go into what *persistent* homology it is well worth our time to clearly state what we mean by homology. (Why? Can this be skipped by experienced readers or are our definitions non-standard? Do we mostly follow hatcher?). In a general sense, homology is a particular of invariant of topological spaces. This has categorical reasons and others. Importantly we need to define simplicial complexes. There are other ways of defining this, notably singular homology, but for the computational aspect of persistent homology we do not have to dwell on this. For completion, we refer the reader to Hatcher for a more traditional treatment of homology.

2.1 Simplicial complexes

First we start with the simplex. The n-simplex is the smallest possible convex set in $\mathbb{R}^m$ containing the $n+1$ points $v_0, \dots, v_n$ such that the vectors $v_1 - v_0, \dots, v_n - v_0$ are linearily independent. The points $v_0, \dots, v_n$ are known as the *vertices* of the simplex. By $[v_0, \dots, v_n]$ we denote the simplex given by those very vertices. The *standard* n-simplex with vertices being the unit vectors along coordinate axes is defined as

$$\Delta^n := \{(t_0, \dots, t_n) \in R^{n+1} \mid \sum_i t_i = 1, t_i \geq 0 \quad \forall i\}$$

More to come.. Do we need orientations, for example?

Definition. A face is the $n-1$-simplex you get after removing a vertex from a n-simplex??

2.2 Simplicial complex

A Δ-complex on a given space X is a collection of maps $\sigma_{alpha} : \delta^n \to X$ such that:

1. Someting
2. Something
3. Something

Other definition. A simplicial complex X is a collection of simplices such that for every simplex Δ_1, Δ_2:

1. $\Delta_1, \Delta_2 \subseteq X$
2. $\Delta_1 \cap \Delta_2$ is either a face of both or the empty set.
3. $\Delta_1 \subseteq \Delta_2 \subset X \implies \Delta_1 \subset X$

So a pair of simplices in the complex can only touch at subsimplex, and all of the faces of a simplex is also in the complex.

This is the geometric definition of a simplicial complex. However, since we are working with topological spaces it is advantageous to think of an abstract simplicial complex without concerning ourselves with the geometric connotations:

Definition. An abstract simplicial complex is a consists of a set K and a collection of subsets $\Delta \subset K$ called simplices such that:

1. $v \in K$ then $\{v\} \in \Delta$
2. $\sigma \in \Delta$ and $\tau \subset \sigma$ then $\tau \in \Delta$

Definition (nlab). An abstract simplicial complex consists of

1. a set of objects $V(K)$ called the vertices
2. a set $S(K)$ of finite non-empty subsets of $V(K)$ called the simplices

such that the following holds:

1. if $\sigma \subset V(K)$ is a simplex, in other words $\sigma \in S(K)$, and $\tau \subset \sigma, \tau \neq \emptyset$ then $\tau \in S(K)$
2. For $v \in V(K)$ the singleton $\{v\}$ is a simplex.

Note how τ in this definition coincides with the geometric definition of a face of σ. Basically this abstract definition tells us that it is enough to define a simplicial complex and its corresponding simplices by the vertices alone and how they group together. If we want to recover an actual geometric simplex we look at the geometric realization of the simplicial complex.

Definition. Geometric Realization. A geometric realization $|K|$ of the abstract simplicial complex K is given by..

Since all simplices of same dimension are homeomorphic this concludes what we wanted etc.

2.3 Simplicial homology

For a simplicial complex K of dimension n we define a free abelian group C_k on the oriented $k-simplices$ of K. The elements of C_k are called k-chains and are formal sums of the type $\sum \alpha_i \sigma_i$ where α_i are coefficients in some ring R. Furthermore, we have a collection of homomorphisms, known as boundary maps, which together with the chain groups form a chain complex. The kth boundary map

$$\partial_k : C_k \to C_{k-1}$$

takes a k-simplex σ

$$\partial_k \sigma = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k]$$

where $\hat{v}_i$ signifies that this vertex has been omitted. This is a linear map so

$$\delta_k \sum \alpha_i \sigma_i = \sum \alpha_i \delta_k \sigma_i$$

Now a simplicial chain complex is a collection of chain groups together with their corresponding boundary maps as a sequence: Tikzcd diagrams.

Note that the boundary maps compose to become the zero map. From this definition we know that from every simplicial complex K we can associate a simplicial chain complex (this is a functor). We then define the homology group of K as the kernel quotioned by the image in the previous. What does this mean? Well, it's simply that we quotient cycles with boundaries. Note that the structure of H_k is in part dependent on the choice of ring R.

3 Persistent

What is persistent homology? What kind of spaces can we compute it in? What does computation entail?

3.1 Views

3.1.1 α Complex

3.1.2 Vietoris-Rips Complex

Theorem 1. *hehe*

3.1.3 Something Complex

3.2 Metrics

3.3 Persistence Diagrams

3.4 Barcodes

3.5 Computation of

4 Data