Introduction to Topological Data Analysis Worksheet

Peter Bubenik, January 2017

1. Simplicial complexes

Definition 1. A *k-simplex* is a k-dimensional polytope which is the convex hull of $k+1$ points in Euclidean space in general position.[1] The convex hull of a non-empty subset of these points is called a *face* of the simplex.

Definition 2. A *simplicial complex* is a set K of simplices such that

a. the face of any simplex in K is also in K; and
b. any non-empty intersection of two simplices is a face of both of them.

Example 3. The *standard k-simplex* is the convex hull of the first $k+1$ standard basis vectors.

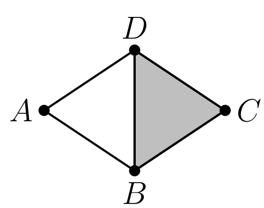
$$\Delta^k = \{(x_1, \ldots, x_{k+1}) \mid \sum_{i=1}^{k+1} x_i = 1, x_i \geq 0 \text{ for all } i\}$$

Definition 4. An *abstract simplicial complex* is a set K of non-empty subsets of a fixed set such that

- if $A \in K$ then all nonempty subsets of A are also in K.

Remark 5. Given a simplicial complex, there is a corresponding abstract simplicial complex. Given an abstract simplicial complex, there is a canonical corresponding simplicial complex constructed by gluing together standard simplices.

Exercise 1. Consider the following simplicial complex.



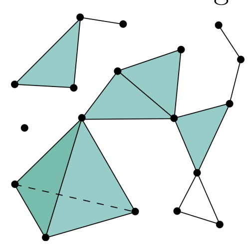
Write down the corresponding abstract simplicial complex.

[1] The points $x_1, \ldots, x_{k+1}$ are *affinely independent*, which means that the vectors $x_2 - x_1, \ldots, x_{k+1} - x_1$ are linearly independent.

2. Betti numbers

Definition 6. The k-th Betti number, β_k, counts the number of k-dimensional features of a simplicial complex.[2] For example, β_0 gives the number of connected components, β_1 gives the number of holes, and β_2 gives the number of voids.

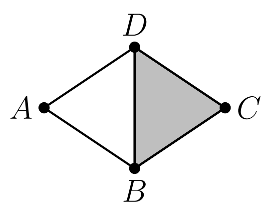
Exercise 2. Give the Betti numbers of the following simplicial complex.



3. Homology

Definition 7. Given an (abstract) simplicial complex, K, the *k-chains*, C_k, is the vector space over the field $\mathbb{Z}/2\mathbb{Z}$ with basis given by the k-simplices in K. The *boundary map*, $\partial_k : C_k \to C_{k-1}$, is given by the matrix whose entry in the ith row and jth column is 1 if the ith $(k-1)$-simplex is a face of the jth k-simplex, and otherwise 0. The *k-cycles* are the k-chains α such that $\partial_k(\alpha) = 0$. The *k-boundaries* are the k-chains α that are in the image of ∂_{k+1}. The k-cycles and k-boundaries form subspaces Z_k and B_k. It is an exercise to check that $B_k \subset Z_k$. This is the crucial part! We can now define *k-homology* to be the quotient space $H_K = Z_k/B_k$. Then the *k-th Betti number* is given by $\beta_k = \dim(H_k) = \dim(Z_k) - \dim(B_k) = \text{nullity}(\partial_k) - \text{rank}(\partial_{k+1})$.

Exercise 3. Consider the following simplicial complex.



a. Write down a basis for C_0, C_1 and C_2.
b. Write down the boundary matrices ∂_1 and ∂_2.
c. Use these matrices together with $\partial_0 = 0$ and $\partial_3 = 0$ to calculate the Betti numbers β_0, β_1 and β_2.

[2] We will make this precise in the next section.

4. Čech complex

Definition 8. Given a set of points $X = \{X_1, \dots, X_N\}$ in a metric space and $r \geq 0$, the *r-Čech complex*, $\check{C}_r(X)$, is the abstract simplicial complex on the set X consisting of those subsets $\{X_{i_1}, \dots, X_{i_m}\}$ such that the intersection of all of the closed balls of radius r centered at the X_{i_j} is nonempty.

Exercise 4. Given the set of points $X = \{(0,0), (0,1), (1,0), (1,1)\} \subset \mathbb{R}^2$, draw a picture of $\check{C}_{\frac{1}{2}}(X)$.

5. Persistent homology

Definition 9. A *filtered simplicial complex* consists of a nested sequence of simplicial complexes.

$$K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n$$

Applying homology, we obtain a sequence of vector spaces and linear maps called a *persistence module.*

$$H_k(K_0) \to H_k(K_1) \to H_k(K_2) \to \cdots \to H_k(K_n)$$

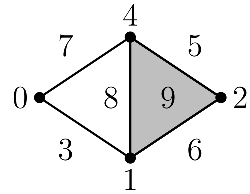
The images of (compositions of) these maps are called *persistent homology* vector spaces.

Theorem 10 (Fundamental Theorem of Persistent Homology)**.** *There is a choice of pairs of filtration values (corresponding to the birth and death of persistent homology classes) that completely describes the persistence module (up to isomorphism).*

Remark 11. When two homology classes born at filtration times b_1 and b_2 become homologous at filtration time d, follow the *Elder Rule*: the old survive; that is, pair d with the larger of b_1 and b_2. *This is not a choice, it follows from the algebra!* Homology classes born at filtration value b that never die are given the birth-death pair (b, ∞).

Definition 12. The *barcode* consists of intervals whose endpoints are these filtration values. The *persistence diagram* is a graph of points corresponding to these ordered pairs of filtration values.

Exercise 5. Consider the following filtered simplicial complex.



That is,

$$X_0 \hookrightarrow X_1 \hookrightarrow X_2 \hookrightarrow X_3 \hookrightarrow X_4 \hookrightarrow X_5 \hookrightarrow X_6 \hookrightarrow X_7 \hookrightarrow X_8 \hookrightarrow X_9$$

What happens to the Betti numbers at each time step? (use + for increases, use − for decreases)

Time	0	1	2	3	4	5	6	7	8	9
Betti number effect										

a. Give the birth-death pairs for H_0 and H_1.
b. Plot the corresponding barcode and persistence diagram.

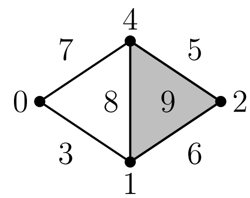
6. Persistence landscape

Definition 13. For $b < d$, let $f_{b,d} : \mathbb{R} \to \mathbb{R}$ be the piecewise-linear "hat" function given by

$$f_{b,d}(t) = \begin{cases} t - b & \text{if } b \leq t \leq \frac{d-b}{2} \\ d - t & \text{if } \frac{d-b}{2} < t \leq d \\ 0 & \text{otherwise.} \end{cases}$$

Let M be a persistence module. For $k = 1, 2, , 3, \ldots$, the *k-th persistence landscape function*, $\lambda_k : \mathbb{R} \to \mathbb{R}$, is given by letting $\lambda_k(t)$ be equal to the kth largest value of $f_{b,d}(t)$ over the birth–death pairs of M. The sequence of these functions is called the *persistence landscape*.[3]

Exercise 6. Consider the following filtered simplicial complex.



The birth–death pairs for *reduced* homology in degree 0 (that is, ignoring the pair $(0, \infty)$) are $(1, 3)$, $(2, 6)$, $(4, 5)$. Graph the corresponding persistence landscape.

[3]Alternatively, it may be viewed as a function $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ where $\lambda(k, t) = \lambda_k(t)$.