Topological Data Analysis

7 November 2019

1 Simplicial complexes from point clouds

1.1 Geometric simplicial complexes

An *n-simplex* in $\mathbb{R}^N$ (where $0 \le n \le N$) is the convex hull of a collection of $n+1$ affinely independent points $p_0, \ldots, p_n$ of $\mathbb{R}^N$; that is, such that the vectors $p_1 - p_0, \ldots, p_n - p_0$ are linearly independent. The following notation will be used:

$$\Delta(p_0, \ldots, p_n) = \{x_0\, p_0 + \cdots + x_n\, p_n \in \mathbb{R}^N \mid x_0 + \cdots + x_n = 1,\ x_i \ge 0 \text{ for all } i\}.$$

Every subset $\{p_{i_0}, \ldots, p_{i_k}\} \subseteq \{p_0, \ldots, p_n\}$ with $0 \le k \le n$ spans a k-simplex $\Delta(p_{i_0}, \ldots, p_{i_k})$, which is called a *k-face* of $\Delta(p_0, \ldots, p_n)$.

The *standard n-simplex* Δ^n is the convex hull of the coordinate unit points $e_0, \ldots, e_n$ in $\mathbb{R}^{n+1}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at place i. Hence,

$$\Delta^n = \Delta(e_0, \ldots, e_n) = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid x_0 + \cdots + x_n = 1,\ x_i \ge 0 \text{ for all } i\}.$$

A *geometric simplicial complex* in $\mathbb{R}^N$ is a set X of simplices in $\mathbb{R}^N$ such that every face of a simplex of X is in X and such that any two simplices of X are either disjoint or intersect along one common face.

The *dimension* of a geometric simplicial complex is the maximum of the dimensions of its faces, where a k-face has dimension k.

Every geometric simplicial complex X has an *underlying* topological space $|X|$, which is the union of all the simplices in X endowed with the Euclidean topology. Then $|X|$ is called a *polyhedron*, and X is called a *triangulation* of $|X|$.

1.2 Abstract simplicial complexes

An *abstract simplicial complex* with *vertex set* $V = \{v_i\}_{i \in I}$ is a collection K of finite subsets $\{v_{i_0}, \ldots, v_{i_k}\} \subseteq V$ such that $\{v\} \in K$ for all $v \in V$ and such that, if $F \in K$ and $F' \subseteq F$, then $F' \in K$.

The elements of V are called *vertices* of K and the elements of K are called *faces* of K. For $k \ge 0$, a face $\{v_{i_0}, \ldots, v_{i_k}\}$ of cardinality $k+1$ is called a *k-face*. Hence, $\{v\}$ is a 0-face of K for every vertex v. The 1-faces are called *edges*. The collection of all k-faces of K for all $0 \le k \le m$ is an abstract simplicial complex, called the *m-skeleton* of K, for any m.

Every abstract simplicial complex is determined by its *maximal* faces, i.e., those not contained in any larger face.

An abstract simplicial complex with vertex set $V = \{v_i\}_{i \in I}$ where I is equipped with a total order is called *ordered*. If K is ordered, then, for simplicity of notation, we will often denote its faces by $(i_0 \cdots i_k)$ with $i_0 < \cdots < i_k$ instead of $\{v_{i_0}, \ldots, v_{i_k}\}$.

1.3 Geometric realization

If K is an ordered abstract simplicial complex with vertex set $V = \{v_{i_0}, \ldots, v_{i_n}\}$, then the *geometric realization* X_K of K is the geometric simplicial complex defined as follows. Let $\{e_0, \ldots, e_n\}$ be the set of coordinate unit points in $\mathbb{R}^{n+1}$. For each k-face $\{v_{i_0}, \ldots, v_{i_k}\}$ in K with $0 \le k \le n$, consider the corresponding k-simplex $\Delta(e_{i_0}, \ldots, e_{i_k})$ in $\mathbb{R}^{n+1}$, and let X_K be the set of all those simplices spanned by the faces of K. We will denote the underlying topological space $|X_K|$ by $|K|$ as well.

If K is unordered, then the space $|K|$ depends on a choice of an order on V and therefore it is determined up to a face-preserving homeomorphism.

If X is any geometric complex and V is the set of 0-faces of X, then X yields an abstract complex K_X with vertex set V and faces given by the simplices of X. In this situation, there is a face-preserving homeomorphism $|X| \cong |K_X|$.

Conversely, every abstract complex K admits a face-preserving bijective correspondence with the abstract complex determined by the geometric realization X_K.

1.4 Čech complexes and Vietoris–Rips complexes

A *point cloud* is an unordered finite collection $X = \{x_i\}_{i \in I}$ of points in $\mathbb{R}^N$ for some $N \ge 1$. Every point cloud X is a finite metric space with the Euclidean distance.

For each real number $\varepsilon > 0$, the *Čech complex* $C_\varepsilon(X)$ of X is the abstract simplicial complex with vertex set X whose k-faces are collections of points $\{x_{i_0}, \ldots, x_{i_k}\}$ such that the closed balls $\bar{B}_{\varepsilon/2}(x_{i_0}), \ldots, \bar{B}_{\varepsilon/2}(x_{i_k})$ have at least one point of common intersection.

Likewise, for each real number $\varepsilon > 0$, the *Vietoris–Rips complex* $R_\varepsilon(X)$ of X is the abstract simplicial complex with vertex set X whose k-faces are collections of points $\{x_{i_0}, \ldots, x_{i_k}\}$ of diameter at most ε; that is, such that $d(x_{i_r}, x_{i_s}) \le \varepsilon$ for all $r, s \in \{0, \ldots, k\}$.

For every X and every ε, there is an inclusion $C_\varepsilon \subseteq R_\varepsilon$. For sufficiently small values of ε, both complexes are equal and discrete (in bijection with X), while for large enough values of ε they are also equal and their geometric realization is a single n-simplex if X consists of $n+1$ points.

A *flag complex* is an abstract simplicial complex where every collection of pairwise adjacent vertices spans a face. Thus every flag complex is maximal among those with a given 1-skeleton, and this means that it can be stored by listing only its edges. The Vietoris–Rips complex of every point cloud is a flag complex.

Short exercises

(1) List the maximal faces of the Čech complex $C_\varepsilon(X)$ and the Vietoris–Rips complex $R_\varepsilon(X)$, depending on ε, if X is the set of vertices of a regular hexagon of radius 1.

(2) List the maximal faces of the Vietoris–Rips complex for every ε for the point cloud $X = \{(0,0,0),\ (1,1,1),\ (0,2,0),\ (0,0,-1)\}$ in $\mathbb{R}^3$.

Longer exercises

(1) Prove that if $C_\varepsilon(X)$ is the Čech complex of any point cloud X, then the space $|C_\varepsilon(X)|$ is homotopy equivalent to the union of all the closed balls of radius $\varepsilon/2$ centered at the points of X. Show by means of a counterexample that the Vietoris–Rips complex does not share this property.

(2) Prove that $R_\varepsilon(X) \subseteq C_{\varepsilon\sqrt{2}}(X)$ for every ε and every point cloud X. You may check it easily if X has cardinality smaller than or equal to 4. For a proof in the general case, see Theorem 2.5 in [V. de Silva, R. Ghrist, Coverage in sensor networks via persistent homology, *Algebr. Geom. Topol.* 7 (2007), 339–358].

(3) Find and discuss the definition of the *alpha complex* $A_\varepsilon(X)$ of a point cloud X, including examples.

(4) Find information about GUDHI (*Geometry Understanding in Higher Dimensions*) at http://gudhi.gforge.inria.fr. The GUDHI library is a generic open source C++ library, with a Python interface, for topological data analysis and higher dimensional geometry understanding. The library offers algorithms to construct simplicial complexes and compute persistent homology.

(5) Find information about the R package TDA (*Topological Data Analysis*) at https://cran.r-project.org/web/packages/TDAstats/index.html. You may find useful to read [R. R. Wadhwa et al., TDAstats: R pipeline for computing persistent homology in topological data analysis, *J. Open Source Software* 3(28), 860 (2018)] and [B. T. Fasy et al., Introduction to the R package TDA, 2015, https://hal.inria.fr/hal-01113028].