

Seminario de Aprendizaje Topológico 2025-2

A Survey of Topological Machine Learning Methods

por

Felix Hensel, Michael Moor and Bastian Rieck

Fernando H Ferreyra

26 de febrero de 2025

Contenidos

Homología simplicial

Homología singular

Filtraciones

Métodos

Homología simplicial

Sea K un complejo simplicial y $s \in K$ un q -simplejo $s = \{v_0, \dots, v_q\}$, la frontera de s consta de sus $(q - 1)$ caras, es decir,

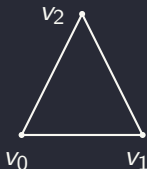
$$\partial s = \{v_1, v_2, \dots, v_q\} \cup \{v_0, v_2, \dots, v_q\} \cup \dots \cup \{v_1, v_2, \dots, v_{q-1}\}$$

Homología simplicial

Sea K un complejo simplicial y $s \in K$ un q -simplejo
 $s = \{v_0, \dots, v_q\}$, la frontera de s consta de sus $(q - 1)$ caras, es decir,

$$\partial s = \{v_1, v_2, \dots, v_q\} \cup \{v_0, v_2, \dots, v_q\} \cup \dots \cup \{v_1, v_2, \dots, v_{q-1}\}$$

Consideremos el siguiente 2-simplejo:



$$s = \{v_0, v_1, v_2\}$$

$$\partial s = \{v_1, v_2\} \cup \{v_0, v_2\} \cup \{v_0, v_1\}$$

Siguiendo el artículo y por simplicidad, trabajemos sobre el campo \mathbb{Z}_2 .

Consideremos $C_q(K)$ el \mathbb{Z}_2 –espacio vectorial generado por q –simplejos de K .

Definimos el operador frontera $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$ donde para cada $s = \{v_0, \dots, v_q\} \in K$ simplejo,

$$\partial_q(s) = \{\hat{v}_0, v_1, \dots, v_q\} + \{v_0, \hat{v}_1, \dots, v_q\} + \dots + \{v_0, v_1, \dots, \hat{v}_q\}$$

donde \hat{v}_i quiere decir que omitimos el vértice v_i para todo $i \in \{0, \dots, q\}$

Siguiendo el artículo y por simplicidad, trabajemos sobre el campo \mathbb{Z}_2 .

Consideremos $C_q(K)$ el \mathbb{Z}_2 –espacio vectorial generado por q –simplejos de K .

Definimos el operador frontera $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$ donde para cada $s = \{v_0, \dots, v_q\} \in K$ simplejo,

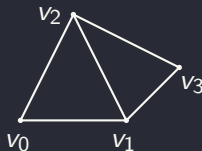
$$\partial_q(s) = \{\hat{v}_0, v_1, \dots, v_q\} + \{v_0, \hat{v}_1, \dots, v_q\} + \dots + \{v_0, v_1, \dots, \hat{v}_q\}$$

donde \hat{v}_i quiere decir que omitimos el vértice v_i para todo $i \in \{0, \dots, q\}$

Retomando el ejemplo anterior, si $s = \{v_0, v_1, v_2\}$, entonces

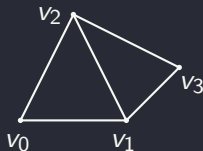
$$\partial(s) = \{v_1, v_2\} + \{v_0, v_2\} + \{v_0, v_1\}$$

Consideremos ahora la triangulación de la esfera



La cadena que representa el simplejo anterior está dada por
 $c = \{v_0, v_1, v_3\} + \{v_1, v_2, v_3\} + \{v_0, v_1, v_2\} + \{v_0, v_2, v_3\}$.

Consideremos ahora la triangulación de la esfera



La cadena que representa el simplejo anterior está dada por $c = \{v_0, v_1, v_3\} + \{v_1, v_2, v_3\} + \{v_0, v_1, v_2\} + \{v_0, v_2, v_3\}$. Podemos calcular su frontera la cual es

$$\begin{aligned}
 \partial_2(c) &= \partial_2(\{v_0, v_1, v_3\}) + \partial_2(\{v_1, v_2, v_3\}) + \partial_2(\{v_0, v_1, v_2\}) + \partial_2(\{v_0, v_2, v_3\}) \\
 &= \{v_1, v_3\} + \{v_0, v_3\} + \{v_0, v_1\} + \{v_2, v_3\} + \{v_1, v_3\} + \{v_1, v_2\} \\
 &\quad + \{v_1, v_2\} + \{v_0, v_2\} + \{v_0, v_1\} + \{v_2, v_3\} + \{v_0, v_3\} + \{v_0, v_2\} \\
 &= 0
 \end{aligned}$$



Para cada q , diremos que las cadenas en $Im\partial_{q+1}$ son las fronteras y los elementos en $ker\partial_q$ son los ciclos.

Para cada q , diremos que las cadenas en $Im\partial_{q+1}$ son las fronteras y los elementos en $ker\partial_q$ son los ciclos.

Resulta que cada frontera es un ciclo, es decir,
 $Im\partial_{q+1} \subset ker\partial_q$.

Para ver lo anterior basta ver que $\partial_q \circ \partial_{q+1} = 0$.

Para cada q , diremos que las cadenas en $Im\partial_{q+1}$ son las fronteras y los elementos en $ker\partial_q$ son los ciclos.

Resulta que cada frontera es un ciclo, es decir,

$$Im\partial_{q+1} \subset ker\partial_q.$$

Para ver lo anterior basta ver que $\partial_q \circ \partial_{q+1} = 0$.

Sea $s = \{v_0, \dots, v_{q+1}\}$, entonces

$$\begin{aligned} \partial_q(\partial_{q+1}\{v_0, \dots, v_{q+1}\}) &= \partial_{q+1} \left(\sum_{i=0}^{q+1} \{v_0, \dots, \hat{v}_i, \dots, v_{q+1}\} \right) \\ &= \sum_{i=0}^{q+1} \partial_q \{v_0, \dots, \hat{v}_i, \dots, v_{q+1}\} \\ &= \sum_{i=0}^{q+1} \sum_{\substack{j=0 \\ i \neq j}}^q \{v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_{q+1}\} \\ &= 0 \end{aligned}$$

Homología simplicial

Definimos la q -homología simplicial de K con coeficientes en \mathbb{Z}_2 al módulo

$$H_q(K; \mathbb{Z}_2) = \ker \partial_q / \operatorname{Im} \partial_{q+1}$$

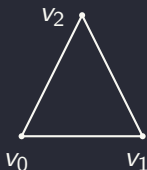
Homología simplicial

Definimos la q -homología simplicial de K con coeficientes en \mathbb{Z}_2 al módulo

$$H_q(K; \mathbb{Z}_2) = \ker \partial_q / \text{Im} \partial_{q+1}$$

¿Qué mide la homología?

Calculemos la 1-homología del siguiente complejo simplicial



$$H_1(K; \mathbb{Z}_2) = \ker \partial_1 / \text{Im} \partial_2$$

Notemos que $\text{Im} \partial_2$ es trivial, y $\ker \partial_1$ está generado por una cadena, así

$$H_1(K; \mathbb{Z}_2) = \mathbb{Z}_2$$

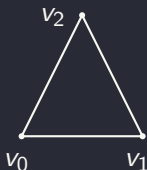
Homología simplicial

Definimos la q -homología simplicial de K con coeficientes en \mathbb{Z}_2 al módulo

$$H_q(K; \mathbb{Z}_2) = \ker \partial_q / \text{Im} \partial_{q+1}$$

¿Qué mide la homología?

Calculemos la 1-homología del siguiente complejo simplicial



$$H_1(K; \mathbb{Z}_2) = \ker \partial_1 / \text{Im} \partial_2$$

Notemos que $\text{Im} \partial_2$ es trivial, y $\ker \partial_1$ está generado por una cadena, así

$$H_1(K; \mathbb{Z}_2) = \mathbb{Z}_2$$

¡La homología mide hoyos!

Contenidos

Homología simplicial

Homología singular

Filtraciones

Métodos

Homología singular

Hasta el momento, solo sabemos calcular la homología a complejos simpliciales.

Sin embargo, ¿qué pasa con espacios más generales?

Homología singular

Hasta el momento, solo sabemos calcular la homología a complejos simpliciales.

Sin embargo, ¿qué pasa con espacios más generales?

Si un espacio topológico X es homeomorfo a la realización geométrica de algún complejo simplicial K , $X \cong |K|$ podemos definir $H_q(X; \mathbb{Z}_2) := H_q(K, \mathbb{Z}_2)$ en virtud de la invarianza topológica de la homología simplicial.

Homología singular

Hasta el momento, solo sabemos calcular la homología a complejos simpliciales.

Sin embargo, ¿qué pasa con espacios más generales?

Si un espacio topológico X es homeomorfo a la realización geométrica de algún complejo simplicial K , $X \cong |K|$ podemos definir $H_q(X; \mathbb{Z}_2) := H_q(K, \mathbb{Z}_2)$ en virtud de la invarianza topológica de la homología simplicial.

Problema. No todo espacio topológico es triangulable.

Para extender el concepto de homología, dado X un espacio topológico consideremos al conjunto de funciones continuas de la forma

$$\sigma : \Delta^q \rightarrow X$$

que llamaremos los q —**simplejos singulares de X** .

Para extender el concepto de homología, dado X un espacio topológico consideremos al conjunto de funciones continuas de la forma

$$\sigma : \Delta^q \rightarrow X$$

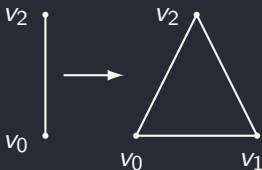
que llamaremos los q -**simplejos singulares de X** .

Como antes, necesitamos definir un operador frontera

$$\partial_q$$

Primero consideremos la inclusión de la i -ésima cara de Δ^q que denotaremos

$$F_i^q : \Delta^{q-1} \rightarrow \Delta^q, F_q^i(t_0, \dots, t_{q-1}) = (t_0, \dots, 0, t_i, \dots, t_{q-1})$$



Sea X un espacio arbitrario y $S_q(X)$ el \mathbb{Z}_2 –espacio vectorial generado por los q –simplejos singulares, definimos

$$\partial_q : S_q(X) \rightarrow S_{q-1}(X),$$

$$\partial_q(\sigma) = \sum_{i=0}^q \sigma \circ F_q^i$$

Sea X un espacio arbitrario y $S_q(X)$ el \mathbb{Z}_2 –espacio vectorial generado por los q –simplejos singulares, definimos

$$\partial_q : S_q(X) \rightarrow S_{q-1}(X),$$

$$\partial_q(\sigma) = \sum_{i=0}^q \sigma \circ F_q^i = \sum_{i=0}^q \sigma|_{[e_0, \dots, \hat{e}_i, \dots, e_q]}$$

Es fácil ver que, al igual que en el caso simplicial, $\partial_q \circ \partial_{q+1} = 0$ y tenemos una sucesión

$$\rightarrow S_q(X) \xrightarrow{\partial_q} S_{q-1}(X) \xrightarrow{\partial_{q-1}} \dots \rightarrow S_1(X) \xrightarrow{\partial_1} S_0(X) \xrightarrow{\partial_0} 0$$

Definimos la q -ésima homología singular de X con coeficientes en \mathbb{Z}_2 como

$$H_q(X; \mathbb{Z}_2) = \ker \partial_q / \operatorname{Im} \partial_{q+1}$$

A los elementos $\ker \partial_q = Z_q(X)$ les llamaremos q - *ciclos* y los elementos $\operatorname{Im} \partial_{q+1} = B_q(X)$ les decimos q - *fronteras*

Nuevamente, es necesario preguntarnos

¿qué mide la homología singular?

Nuevamente, es necesario preguntarnos

¿qué mide la homología singular?

Primero, podemos hablar de la 0–homología, en la cual tenemos los siguientes resultados:

- ▶ Sea X un espacio topológico 0–conexo. Entonces $H_0(X; R) \cong \mathbb{Z}_2$
- ▶ Sea X un espacio y $\{X_\alpha\}_{\alpha \in \Lambda}$ la familia de sus componentes por trayectorias. Entonces $H_q(X; \mathbb{Z}_2) \cong \bigoplus_{\alpha \in \Lambda} H_q(X_\alpha; \mathbb{Z}_2)$

Nuevamente, es necesario preguntarnos

¿qué mide la homología singular?

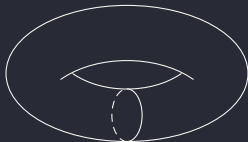
Primero, podemos hablar de la 0–homología, en la cual tenemos los siguientes resultados:

- ▶ Sea X un espacio topológico 0–conexo. Entonces $H_0(X; \mathbb{Z}_2) \cong \mathbb{Z}_2$
- ▶ Sea X un espacio y $\{X_\alpha\}_{\alpha \in \Lambda}$ la familia de sus componentes por trayectorias. Entonces $H_q(X; \mathbb{Z}_2) \cong \bigoplus_{\alpha \in \Lambda} H_q(X_\alpha; \mathbb{Z}_2)$

De los dos resultados anteriores se sigue que:

Sea X un espacio topológico, entonces $H_0(X; \mathbb{Z}_2)$ es isomorfo a la suma directa de copias de \mathbb{Z}_2 , una copia por cada componente por trayectorias.

A continuación presentamos la figura de un toro, \mathbb{T} .



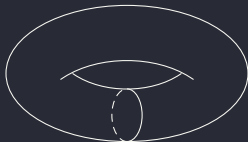
Con un poco de esfuerzo, podemos probar que

$$H_1(\mathbb{T}, \mathbb{Z}_2) \cong H_1(S^2 \times S^2; \mathbb{Z}_2) \cong H_1(S^2; \mathbb{Z}_2) \oplus H_1(S^2; \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2.$$

Es decir,

$$H_1(\mathbb{T}, \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$$

A continuación presentamos la figura de un toro, \mathbb{T} .



Con un poco de esfuerzo, podemos probar que

$$H_1(\mathbb{T}, \mathbb{Z}_2) \cong H_1(S^2 \times S^2; \mathbb{Z}_2) \cong H_1(S^2; \mathbb{Z}_2) \oplus H_1(S^2; \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2.$$

Es decir,

$$H_1(\mathbb{T}, \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$$

Entonces, tenemos que la dimensión de $H_1(\mathbb{T}, \mathbb{Z}_2)$ es 2, lo cual coincide con nuestra intuición de cuántos hoyos tiene el toro.

Ya que los grupos de homología son invariantes topológicos, es decir, permanecen invariantes bajo homeomorfismos (salvo isomorfismo) son relevantes para el estudio y clasificación de espacios.

Terminaremos la sección definiendo una invariante importante derivada de los módulos de homología son los números de Betti. Sea $q \geq 0$, definimos el q -ésimo número de Betti como

$$\beta_q := \dim H_q(X; \mathbb{Z}_2)$$

Ya que los grupos de homología son invariantes topológicos, es decir, permanecen invariantes bajo homeomorfismos (salvo isomorfismo) son relevantes para el estudio y clasificación de espacios.

Terminaremos la sección definiendo una invariante importante derivada de los módulos de homología son los números de Betti. Sea $q \geq 0$, definimos el q -ésimo número de Betti como

$$\beta_q := \dim H_q(X; \mathbb{Z}_2)$$

Tal como se expreso anteriormente, β_0 determina el número de componentes conexas de un espacio y β_1 cuenta el número de hoyos.

Contenidos

Homología simplicial

Homología singular

Filtraciones

Métodos

Nervio

Dado un conjunto de datos una tarea común es lograr clasificar los puntos en colecciones con base a características similares.

Veamos como podemos aproximarnos a esta tarea por medio de complejos simpliciales.

Nervio

Dado un conjunto de datos una tarea común es lograr clasificar los puntos en colecciones con base a características similares.

Veamos como podemos aproximarnos a esta tarea por medio de complejos simpliciales.

Sea X un espacio topológico y $\mathcal{U} = \{U_\alpha\}_{\alpha \in \Lambda}$ una cubierta abierta de subconjuntos no vacíos. Definimos un complejo simplicial $N\mathcal{U}$, llamado el nervio de la cubierta como sigue:

- ▶ Los vértices son $V_{N\mathcal{U}} = \Lambda$; y
- ▶ los q -simplejos son las colecciones

$$\{\alpha_0 \alpha_1 \dots \alpha_q \mid U_{\alpha_0} \cap U_{\alpha_1} \cap \dots \cap U_{\alpha_q} \neq \emptyset\}$$

Nervio

Dado un conjunto de datos una tarea común es lograr clasificar los puntos en colecciones con base a características similares.

Veamos como podemos aproximarnos a esta tarea por medio de complejos simpliciales.

Sea X un espacio topológico y $\mathcal{U} = \{U_\alpha\}_{\alpha \in \Lambda}$ una cubierta abierta de subconjuntos no vacíos. Definimos un complejo simplicial $N\mathcal{U}$, llamado el nervio de la cubierta como sigue:

- ▶ Los vértices son $V_{N\mathcal{U}} = \Lambda$; y
- ▶ los q -simplejos son las colecciones

$$\{\alpha_0 \alpha_1 \dots \alpha_q \mid U_{\alpha_0} \cap U_{\alpha_1} \cap \dots \cap U_{\alpha_q} \neq \emptyset\}$$

Notemos que cubiertas distintas generan nervios distintos y, para nuestros fines, algunos serán más útiles que otros.

Dada $X \subset Y \subset \mathbf{R}^n$, para cada $\varepsilon > 0$ y $x \in X$ definimos $B_\varepsilon(x) = \{y \in Y \mid d(x, y) < \varepsilon\}$.

Sea $S \subset Y$ un conjunto finito de puntos y $\varepsilon > 0$, tomamos la cubierta abierta $\mathcal{U} = \{B_\varepsilon(x)\}_{x \in S}$. Definimos el complejo de Cech, denotado $C_\varepsilon(S)$, como el nervio de la cubierta \mathcal{U} .

Dada $X \subset Y \subset \mathbf{R}^n$, para cada $\varepsilon > 0$ y $x \in X$ definimos $B_\varepsilon(x) = \{y \in Y \mid d(x, y) < \varepsilon\}$.

Sea $S \subset Y$ un conjunto finito de puntos y $\varepsilon > 0$, tomamos la cubierta abierta $\mathcal{U} = \{B_\varepsilon(x)\}_{x \in S}$. Definimos el complejo de Cech, denotado $C_\varepsilon(S)$, como el nervio de la cubierta \mathcal{U} .

¿Qué tan buenas son nuestras aproximaciones mediante complejos simpliciales? Borsuk (1948) respondió esta pregunta en el siguiente teorema

Sea $X \subset \mathbf{R}^n$ y $\mathcal{U} = \{U_\alpha\}_{i=1}^n$ una cubierta finita de X , Si cualquier intersección $U_{i_1} \cap \dots \cap U_{i_x}$ es vacía o contraíble, entonces $X \simeq N\mathcal{U}$.

Rips

Aunque la aproximación por medio de complejos de Čech es buena, puede ser computacionalmente muy costoso. Resulta conveniente definir entonces el complejo de Rips.

Rips

Aunque la aproximación por medio de complejos de Čech es buena, puede ser computacionalmente muy costoso. Resulta conveniente definir entonces el complejo de Rips.

Sea $\{v_1, \dots, v_n\} \subset Y \subset \mathbf{R}^n$ y $\varepsilon > 0$, llamaremos a $G_\varepsilon = \{S, E_\varepsilon\}$ la gráfica de ε -vecindades en S , donde

$$\{\{v, u\} \mid d(v, u) \leq \varepsilon, u \neq v \in S\}$$

Rips

Aunque la aproximación por medio de complejos de Cech es buena, puede ser computacionalmente muy costoso.

Resulta conveniente definir entonces el complejo de Rips.

Sea $\{v_1, \dots, v_n\} \subset Y \subset \mathbf{R}^n$ y $\varepsilon > 0$, llamaremos a $G_\varepsilon = \{S, E_\varepsilon\}$ la grafica de ε -vecindades en S , donde

$$\{\{v, u\} \mid d(v, u) \leq \varepsilon, u \neq v \in S\}$$

Se denomina complejo de clanes al complejo simplicial cuyos simplejos son los clanes maximales de una gráfica. Al complejo de clanes de la gráficas de ε -vecindades se le denomina complejo de Vietoris-Rips.

El complejo de Cech es bueno porque nos podemos acercar mucho a X .

Mientras que el complejo de Rips es bueno porque es computacionalmente eficiente.

Pero, ¿mantenemos la información del espacio X ?

El complejo de Cech es bueno porque nos podemos acercar mucho a X .

Mientras que el complejo de Rips es bueno porque es computacionalmente eficiente.

Pero, ¿mantenemos la información del espacio X ?

¡Sí!, en virtud de esta proposición:

Sea $S \subset \mathbf{R}^n$ finito, entonces para todo $\varepsilon > 0$,

$$C_{\frac{\varepsilon}{2}}(S) \subset R_{\varepsilon}(S) \subset C_{\varepsilon}(S)$$

Sea $S \subset \mathbf{R}^n$ finito, entonces para todo $\varepsilon > 0$,

$$C_{\frac{\varepsilon}{2}}(S) \subset R_{\varepsilon}(S) \subset C_{\varepsilon}(S)$$

Demostración,

Sea $\sigma = \{v_0, \dots, v_m\} \in C_{\frac{\varepsilon}{2}}$ Entonces hay $p \in B_{\frac{\varepsilon}{2}}(v_i) \cap B_{\frac{\varepsilon}{2}}(v_j)$, así

$$d(v_i, v_j) \leq d(v_i, p) + d(p, v_j) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Por lo tanto, $\sigma \in R_{\varepsilon}(S)$.

Sea $S \subset \mathbf{R}^n$ finito, entonces para todo $\varepsilon > 0$,

$$C_{\frac{\varepsilon}{2}}(S) \subset R_{\varepsilon}(S) \subset C_{\varepsilon}(S)$$

Demostración,

Sea $\sigma = \{v_0, \dots, v_m\} \in C_{\frac{\varepsilon}{2}}$ Entonces hay $p \in B_{\frac{\varepsilon}{2}}(v_i) \cap B_{\frac{\varepsilon}{2}}(v_j)$, así

$$d(v_i, v_j) \leq d(v_i, p) + d(p, v_j) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Por lo tanto, $\sigma \in R_{\varepsilon}(S)$.

Sea $\sigma = \{v_0, \dots, v_m\} \in R_{\varepsilon}(S)$, entonces $v_i \in B_{\varepsilon}(v_j)$, entonces $B_{\varepsilon}(v_i) \cap B_{\varepsilon}(v_j) \neq \emptyset$.

Por lo tanto, $\sigma \in C_{\varepsilon}(S)$.

Tenemos el siguiente corolario

Sea $S \subset \mathbf{R}^n$ finito, entonces para todo $\varepsilon > 0$,

$$R_\varepsilon(S) \subset C_\varepsilon(S) \subset R_{2\varepsilon}$$

Tenemos el siguiente corolario

Sea $S \subset \mathbf{R}^n$ finito, entonces para todo $\varepsilon > 0$,

$$R_\varepsilon(S) \subset C_\varepsilon(S) \subset R_{2\varepsilon}$$

Por lo tanto, podemos acercarnos tanto como queramos a complejos cech mediante complejos de Rips.

Notemos que en los dos complejos que definimos anteriormente, el complejo generado depende fuertemente del epsilon escogido.

Tenemos que si $\varepsilon_1 < \varepsilon_2$, entonces

$$C_{\varepsilon_1}(S) \subset C_{\varepsilon_2}(S) \qquad R_{\varepsilon_1}(S) \subset R_{\varepsilon_2}(S)$$

Lo anterior motiva el siguiente concepto

Dado un complejo simplicial K , una filtración es un conjunto ordenado por contención de subcomplejos K_i de K con $i \in I$.

Es decir, si $i < j$ en I , entonces $K_i \subset K_j$.

El orden de estos subcomplejos se llama filtro.

Dado un complejo simplicial K , una filtración es un conjunto ordenado por contención de subcomplejos K_i de K con $i \in I$.

Es decir, si $i < j$ en I , entonces $K_i \subset K_j$.

El orden de estos subcomplejos se llama filtro.

Notemos que si tenemos S un conjunto de puntos, entonces los complejos de cech y rips cambian en a lo más un número finito, así podemos pensar $I = \mathbb{Z}^+$.

Cada un complejo de cadenas C , una filtración es un conjunto de complejos ordenados C_i de C . Tenemos el siguiente diagramas

$$\begin{array}{ccccccc}
 C^{i-1} : & \dots & \longrightarrow & C_{k+1}^{i-1} & \xrightarrow{\partial_{k+1}^{i-1}} & C_k^{i-1} & \xrightarrow{\partial_k^{i-1}} C_{k-1}^{i-1} \longrightarrow \dots \\
 \downarrow & & & \downarrow & & \downarrow & \downarrow \\
 C^i : & \dots & \longrightarrow & C_{k+1}^i & \xrightarrow{\partial_{k+1}^i} & C_k^i & \xrightarrow{\partial_k^i} C_{k-1}^i \longrightarrow \dots \\
 \downarrow & & & \downarrow & & \downarrow & \downarrow \\
 C^{i+1} : & \dots & \longrightarrow & C_{k+1}^{i+1} & \xrightarrow{\partial_{k+1}^{i+1}} & C_k^{i+1} & \xrightarrow{\partial_k^{i+1}} C_{k-1}^{i+1} \longrightarrow \dots
 \end{array}$$

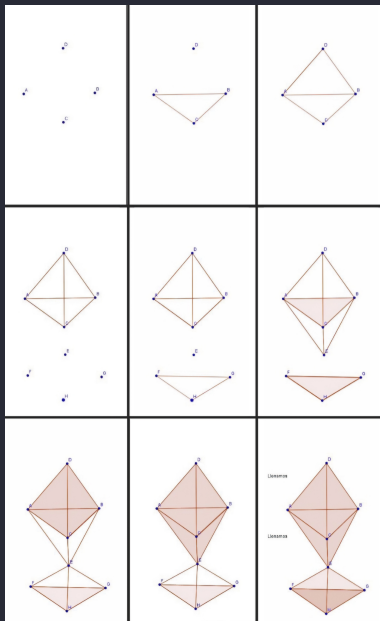
Donde C_k^i representa las i -ésimos q -cadenas.

Denotamos por Z_q^i a los i -ésimos q -ciclos y B_q^i a las i -ésimas q -fronteras.

Definimos la p —persistencia del k —ésimo módulo de homología como

$$H_k^{i,p} = \frac{Z_k^i}{B_k^{i+p} \cap Z_k^i}$$

Lo anterior nos da el número de clases que estaban en la i —ésima k —homología que sigue existiendo en la $i + p$ —ésima.



En 2007, Cohen-Steiner, H. Edelsbrunner y J. Harer demostraron el **teorema de estabilidad para diagramas de persistencia**.

El teorema establece que pequeñas perturbaciones en el conjunto de datos inducen pequeñas perturbaciones en la homología persistente que se obtendrá de ellos.

Contenidos

Homología simplicial

Homología singular

Filtraciones

Métodos

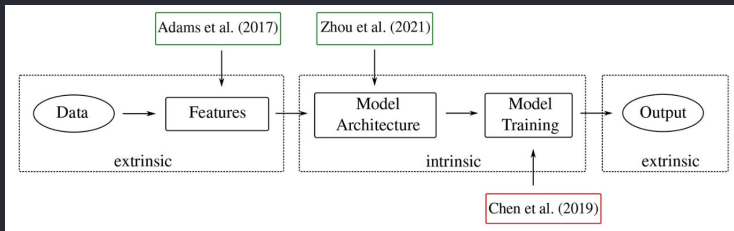
Los metodos de análisis de datos topológicas se agrupan en dos categorías:

- ▶ **Métodos de propiedades topológicas extrínsecas.**

- ▶ **Métodos de propiedades topológicas intrínsecas.**

Estos, a su vez, se subdividen en

- ▶ métodos de observación, y
- ▶ métodos de intervención



Propiedades topológicas extrínsecas

Métodos para obtener representaciones de propiedades topológicas para usarse en modelos de entrenamiento.

Aunque los diagramas de persistencia son muy útiles para la representación de la homología persistente, no pueden ser usados como *input* de modelos de entrenamiento dada la naturaleza de su estructura de datos.

La más simples representaciones son por medio de indicadores estadísticos, por ejemplo, la persistencia total usada para medir la estabilidad topológica del conjunto de datos.

Curvas de Betti

Una representación tanto simple y estable de los diagramas de persistencia como útil para tareas de *machine learning* son las **curvas de Betti**.

Curvas de Betti

Una representación tanto simple y estable de los diagramas de persistencia como útil para tareas de *machine learning* son las **curvas de Betti**.

Sea D un diagrama de persistencia y $w : \mathbf{R}^2 \rightarrow \mathbf{R}$ una función de peso, definimos la curva de Betti función como la función $\beta : \mathbf{R} \rightarrow \mathbf{R}$ definida como

$$\beta(t) := \sum_{(b,d) \in D} w(b,d) 1_{[b,d]}(t)$$

Curvas de Betti

Una representación tanto simple y estable de los diagramas de persistencia como útil para tareas de *machine learning* son las **curvas de Betti**.

Sea D un diagrama de persistencia y $w : \mathbf{R}^2 \rightarrow \mathbf{R}$ una función de peso, definimos la curva de Betti función como la función $\beta : \mathbf{R} \rightarrow \mathbf{R}$ definida como

$$\beta(t) := \sum_{(b,d) \in D} w(b,d) 1_{[b,d]}(t)$$

La ventaja de las curvas de Betti sobre los diagramas, es que al ser funciones permite comparar fácilmente similitudes entre diagramas.

Aplicaciones. Las curvas de Betti se han utilizado para para el análisis de datos (Umeda, 2017), Chevyrev et al. (2018) usaron esta representación para resolver problemas de clasificación.



Carrière et al. (2015) desarrollaron un método vectorial para representar los diagramas de persistencia.

Para cada par de punto $(p, q) \in D$, calculamos

$$m(p, q) = \min\{d_\infty(p, q), d_\infty(p, \Delta), d_\infty(\Delta, q)\}$$

y le asociamos a D un vector cuyas entradas son los $m(p, q)$ ordenados de manera descendiente.

Carrière et al. (2015) desarrollaron un método vectorial para representar los diagramas de persistencia.

Para cada par de punto $(p, q) \in D$, calculamos

$$m(p, q) = \min\{d_\infty(p, q), d_\infty(p, \Delta), d_\infty(\Delta, q)\}$$

y le asociamos a D un vector cuyas entradas son los $m(p, q)$ ordenados de manera descendiente.

Este método posiblemente no funcione bien con *datasets* demasiado largo, además que ha sido poco adoptado dentro de la comunidad.

Referencias