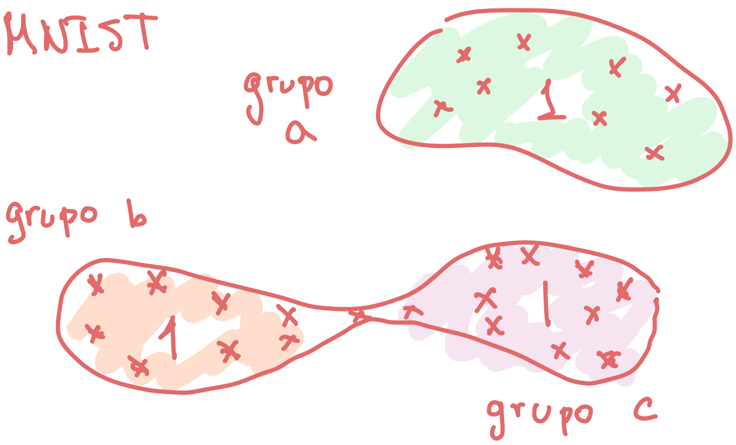
Clase 2 (11/11/2025)

EAFIT - TDA

Ayer:

* Datos (X, d_X), where X is the "conjunto" and d_X is the "distancia"

* MNIST



Topología	Ciencia de datos
Componentes arco-conexas	Clusters (agrupamientos)

Hoy:

* Componentes arco-conexas
* 0-homología
* Clustering (agrupamiento)

— // —

Def: Un grafo (no dirigido) es un par $G = (V, E)$ donde

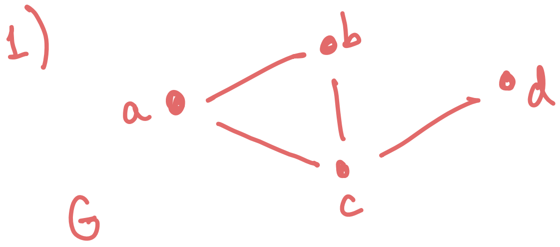
$V \leftarrow$ conjunto de vertices

$$E \subseteq \{ \{v, v'\} \mid v, v' \in V, \; v \neq v' \}$$

$\uparrow$ conjunto de aristas

Ejemplos

1)



$V = \{a, b, c, d\}$

$E = \{\{a,b\}, \{a,c\}, \{b,c\}, \{c,d\}\}$

2) Sea (X, d_X) un conjunto de datos y $\varepsilon > 0$.

Sea $E_\varepsilon(X) = \{ \{x, x'\} \subseteq X \mid 0 < d_X(x, x') < \varepsilon \}$

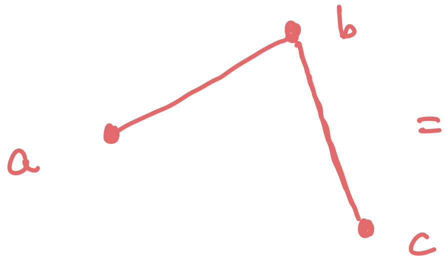
y defina

$$\underbrace{G_\varepsilon(X, d_X)}_{\text{grafo de } \varepsilon\text{-vecindad}} = (\underbrace{X}_{\text{vertices}}, \underbrace{E_\varepsilon(X)}_{\text{aristas}})$$

e.g. $X = \{a, b, c\}$,

d_X	a	b	c
a	0	1	2
b	1	0	1.5
c	2	1.5	0

, $\varepsilon = 2$



$= G_2(X)$

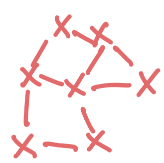
3) Sea (X, d_X) un espacio métrico finito,

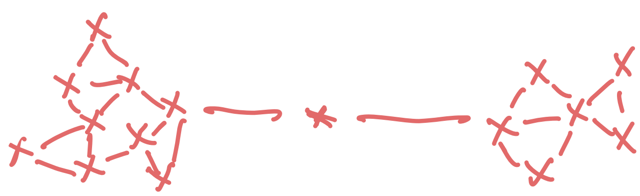
y $1 \leq k < \#(X)$. Para $x \in X$ sea

$$E_k(X) = \left\{ \{x, x'\} \subseteq X \;\middle|\; \begin{array}{l} x' \neq x \text{ es uno de los} \\ k \text{ vecinos mas cercanos} \\ \text{a } x \end{array} \right\}$$

$$\underset{\substack{\uparrow \\ \text{grafo de} \\ k\text{-vecinos} \\ \text{mas cercanos}}}{G_k(X, d_X)} = \Big(\underset{\substack{\uparrow \\ \text{vertices}}}{X} \;,\; \underset{\substack{\uparrow \\ \text{aristas}}}{E_k(X)} \Big)$$

e.g. $k = 2$ $\quad X$





Def: Un camino en un grafo $G = (V, E)$ es una sucesión finita

$$\Gamma = \big(\{v_0, v_1\}, \{v_1, v_2\}, \ldots, \{v_{n-1}, v_n\} \big)$$

de aristas $\{v_j, v_{j+1}\} \in E$, $j = 0, \ldots, n-1$.

$v_0 = i(\Gamma) \leftarrow$ inicio, $\quad v_n = f(\Gamma) \leftarrow$ final.

Def: Dado un grafo $G=(V,E)$, diremos que dos vértices $v, v' \in V$ están conectados por caminos ($v \sim v'$) si $v = v'$, o $v \neq v'$ y existe un camino Γ con $v = i(\Gamma)$ y $v' = f(\Gamma)$.

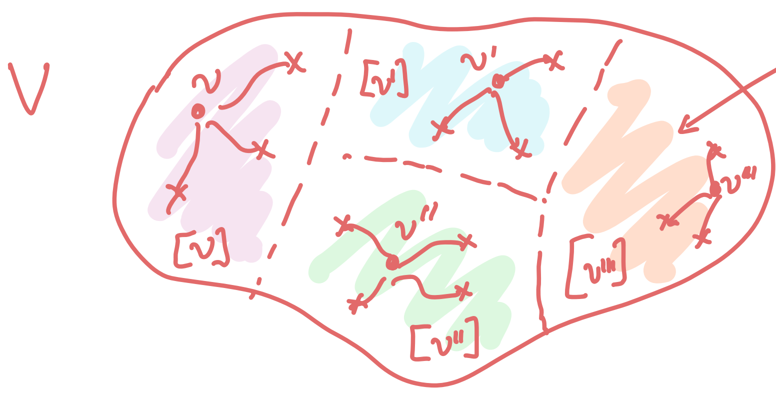
Prop: $\sim$ es una relación de equivalencia en V.

→ reflexiva: $v \sim v$

→ simétrica: si $v \sim v'$ entonces $v' \sim v$

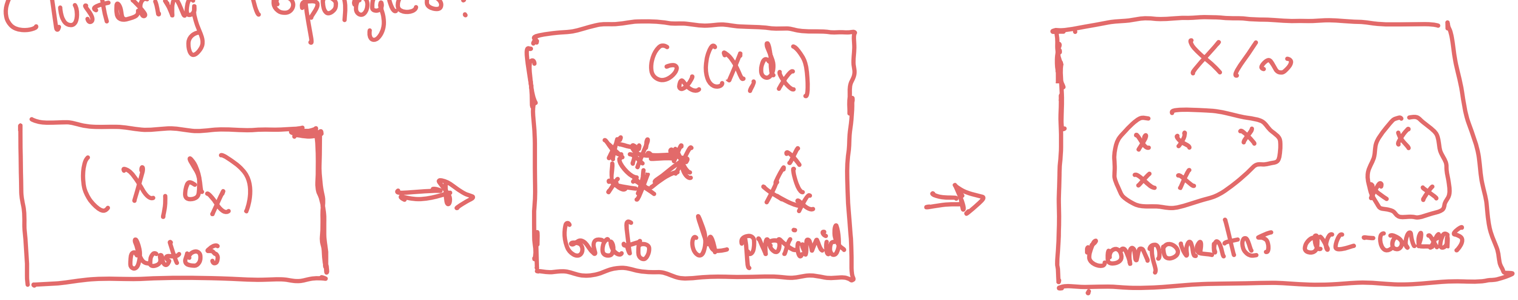
→ transitiva: si $v \sim v'$ y $v' \sim v''$ entonces $v \sim v''$

Nota: $\sim$ particiona V en clases (disjuntas) de equivalencia, llamadas componentes arco-conexas.



$V/\sim$ = conjunto de componentes arco-conexas

Clustering Topológico:



Algoritmos: $G = (V, E)$ finito con $n_E = \#(E)$.

El algoritmo de Kruskal (para encontrar maximum spanning trees con union-find) toma

$$\mathcal{O}(n_E \cdot \log(n_E)) \text{ tiempo.}$$

<u>Versión algebraica</u>: <u>0-homología</u>

Sea $(\mathbb{F}, +, *)$ un campo (e.g. $\mathbb{R}, \mathbb{Q}, \mathbb{C}$, o $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ para $p \in \mathbb{N}$ primo) y para $G = (V, E)$

sean:
$$C_0(G; \mathbb{F}) = \mathrm{Span}_{\mathbb{F}}(V)$$
$$= \left\{ \lambda_1 v_1 + \cdots + \lambda_n v_n \;\middle|\; \begin{array}{l} n \in \mathbb{N},\ \lambda_i \in \mathbb{F} \\ v_i \in V,\ 1 \le i \le n \end{array} \right\}$$

0-cadenas (vertices)

= Espacio vectorial sobre $\mathbb{F}$ con base (i.e., generado por) V.

$$C_1(G; \mathbb{F}) = \mathrm{Span}_{\mathbb{F}}(E)$$
$$= \left\{ \beta_1 e_1 + \cdots + \beta_m e_m \;\middle|\; \begin{array}{l} m \in \mathbb{N},\ \beta_j \in \mathbb{F} \\ e_j \in E,\ 1 \le j \le m \end{array} \right\}$$

1-cadenas. (aristas)

= Espacio vectorial sobre $\mathbb{F}$ generado por E.

Fije un <u>buen orden</u> $\preceq$ en V (todo subconjunto no vacío tiene un elemento mínimo = Teorema de Zermelo $\Leftrightarrow$ Axioma de elección)

$\Rightarrow \quad e \in E \quad \Rightarrow$

$e = \{ i(e), f(e) \} \subseteq V$

$i(e) \preceq f(e)$

Defina la <u>frontera</u> de e como

$$\partial(e) = f(e) - i(e) \in C_0(G; \mathbb{F})$$

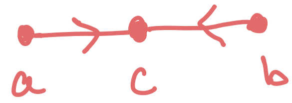
Como E es una base para $C_1(G;\mathbb{F})$, entonces ∂ define una única transformación lineal

$$\partial : C_1(G;\mathbb{F}) \longrightarrow C_0(G;\mathbb{F})$$

Dada por:

$$\partial\left(\sum_{j=1}^{m} \beta_j \cdot e_j \right) = \sum_{j=1}^{m} \beta_j \cdot \partial(e_j) = \sum_{j=1}^{m} \beta_j \cdot (f(e_j) - i(e_j))$$

$$= \sum_{j=1}^{m} \beta_j \cdot f(e_j) - \beta_j \cdot i(e_j)$$

Ejemplo:  G , $\partial(\{a,c\} - \{b,c\}) = b - a$.

Teorema: Sea $G = (V, E)$ un grafo y $v, v' \in V$.

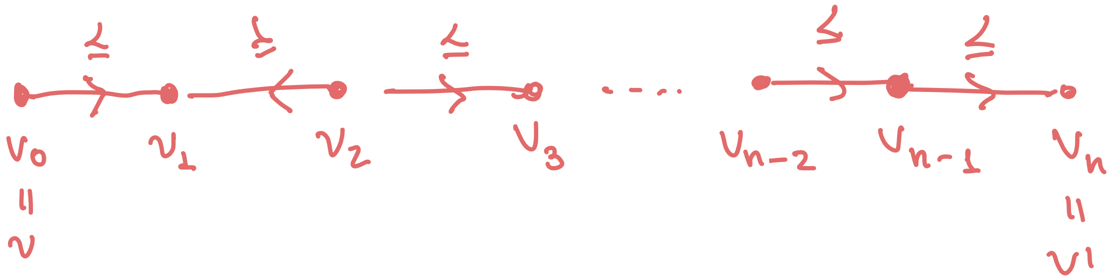
$v \sim v'$ (conectados por caminos) si y solo si

$v' - v \in \operatorname{img}(\partial)$.

Prueba: $(\Rightarrow)$ Si $v \sim v'$, entonces existe

un camino $\Gamma = (\{v_0, v_1\}, \{v_1, v_2\}, \dots, \{v_{n-1}, v_n\})$

en G con $v = v_0 = i(\Gamma)$ y $v' = v_n = f(\Gamma)$.



Sea

$$s_j = \begin{cases} 1 & \text{si } v_{j-1} \leq v_j \\ -1 & \text{si } v_{j-1} \geq v_j \end{cases}$$

note que $\partial(s_j\{v_{j-1}, v_j\}) = v_j - v_{j-1}$

$j = 1, \dots, n$

Si $\gamma = \sum_{j=1}^{n} s_j \cdot \{v_{j-1}, v_j\} \in C_1(G;\mathbb{F})$,

entonces

$$\operatorname{img}(\partial) \ni \partial(\gamma) = \partial\left(\sum_{j=1}^{n} s_j \cdot \{v_{j-1}, v_j\}\right)$$

$$= \sum_{j=1}^{n} \partial\left(s_j \{v_{j-1}, v_j\}\right)$$

$$= \sum_{j=1}^{n} v_j - v_{j-1} = v_n - v_0$$

$$= v' - v$$

$$\Rightarrow v' - v \in \operatorname{img}(\partial).$$

($\Leftarrow$) Si $v' - v = \partial\left(\sum_{j=1}^{m} \beta_j e_j\right)$, $\beta_j \in \mathbb{F}$, $e_j \in E$

entonces uno puede utilizar algunos de los e_j's para construir un camino Γ de v a v'

(ver: Munkres, Elements of Algebraic Topology, Thm 7.1)

$\blacksquare$

Corolario: El espacio vectorial cociente

base = vertices de G

$$H_0(G;\mathbb{F}) := C_0(G;\mathbb{F}) \Big/ \operatorname{img}(\partial)$$

difieren por un camino

0-homología de G con coeficientes en $\mathbb{F}$

tiene dimensión

$$\#(V/\sim) = \#$$ de componentes arco-conexas de G.