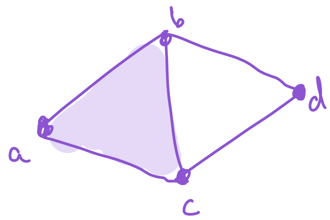
Hasta Ahora

- Complejos Simpliciales



K

$\{a, b, c, d, ab, bc, cd, ac, bd, abc\}$

  - ε-Complejo de Rips

$$R_\varepsilon(X, d_X) = \{\sigma \subseteq X \mid \sigma \neq \phi \text{ y } \operatorname{diam}(\sigma) < \varepsilon\}$$

  - Dada $f: K^{(0)} \longrightarrow \mathbb{R}$

$$K^{\leq r} = \{\sigma \in K \mid \max_{x \in \sigma} f(x) \leq r\} \quad r\text{-subnivel}$$

$$K^{\geq r} = \{\sigma \in K \mid \min_{x \in \sigma} f(x) \geq r\} \quad r\text{-supernivel}$$

- Homología de K con coeficientes en un campo $\mathbb{F}$

$$\cdots \xrightarrow{\partial_{n+1}} C_n(K;\mathbb{F}) \xrightarrow{\partial_n} C_{n-1}(K;\mathbb{F}) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_1(K;\mathbb{F}) \xrightarrow{\partial_0} 0$$

$$\sigma = \{x_0, \ldots, x_n\} \longmapsto \sum_{i=0}^{n} (-1)^i \, \sigma \setminus \{x_i\}$$

$$\underbrace{\operatorname{Img}(\partial_{n+1}) = B_n(K;\mathbb{F})}_{\text{fronteras}} \subseteq \underbrace{Z_n(K;\mathbb{F}) = \operatorname{Ker}(\partial_n)}_{\text{Ciclos}}$$

Def:

$$\boxed{H_n(K;\mathbb{F}) = Z_n(K;\mathbb{F}) \big/ B_n(K;\mathbb{F})}$$

$$[z] \in H_n(K;\mathbb{F}) \Rightarrow [z] = z + B_n(K;\mathbb{F})$$

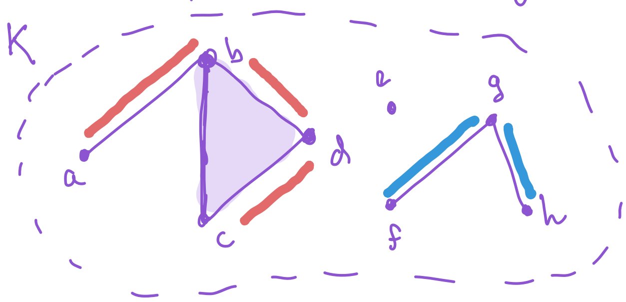
$$z \in Z_n(K;\mathbb{F})$$

Hoy

- $\dim(H_0(K;\mathbb{Z}_2))$ = # componentes conexas por caminos de K.
- Homología persistente para agrupamientos (clustering)

Def: Sea K un complejo simplicial y escriba los vertices de K como $x \in K^{(0)}$, en vez de $\{x\} \in K^{(0)}$.

Decimos que $x, x' \in K^{(0)}$ estan conectados via un camino, si existen $\{x_0,x_1\}, \{x_1,x_2\}, \dots, \{x_{n-1},x_n\} \in K^{(1)}$ tal que $x = x_0$ y $x_n = x'$. Notación: $x \sim x'$



- $a \sim c$: $\{a,b\}, \{b,d\}, \{d,c\}$
- $f \sim h$: $\{f,g\}, \{g,h\}$
- $a \not\sim h$

Nota: "Estar conectados por caminos" es una relación de equivalencia en $K^{(0)}$ ($x \sim x'$ es simetrica, reflexiva, trans). Las clases de equivalencia $I \subset K^{(0)}$ son llamadas las componentes conexas por caminos de K.

Prop: Sean $\{I_j\}_{j\in J}$ las componentes de K, y para cada $j \in J$ fije $b_j \in I_j$.

La colección

$$\mathcal{B} = \{ [b_j] \in H_0(K; \mathbb{Z}_2) \mid j \in J \}$$

genera $\underline{H_0(K;\mathbb{Z}_2)}$ sobre $\mathbb{Z}_2$.

Proof:

$$C_1(K;\mathbb{Z}_2) \xrightarrow{\partial_1} C_0(K;\mathbb{Z}_2) \xrightarrow{\partial_0} 0$$

$\forall j \in J$, $b_j \in \underbrace{K^{(0)} \subseteq C_0(K;\mathbb{Z}_2)}_{\text{base}} = \ker(\partial_0) = \underline{Z_0(K;\mathbb{Z}_2)}$

$\{ [x] \in H_0(K;\mathbb{Z}_2) \mid x \in K^{(0)} \}$ genera $H_0(K;\mathbb{Z}_2) = \dfrac{Z_0(K;\mathbb{Z}_2)}{B_0(K;\mathbb{Z}_2)}$

Es suficiente mostrar que $\{ [x] \mid x \in K^{(0)} \}$ y $\{ [b_j] \mid j \in J \}$ generan lo mismo.

Sea $x \in K^{(0)} = \bigcup\limits_{j \in J} \Gamma_j$ (componentes conexas por caminos) $\Rightarrow$ existe $j(x) \in J$ tal que

$x \in \Gamma_{j(x)} \ni b_{j(x)} \Rightarrow x \sim b_{j(x)}$

$\Rightarrow \exists \{x_0, x_1\}, \{x_1, x_2\}, \dots, \{x_{n-1}, x_n\} \in K^{(1)}$

$x = x_0$, $b_{j(x)} = x_n$

$\Rightarrow$ Sea $q = \{x_0,x_1\} + \{x_1,x_2\} + \dots + \{x_{n-1},x_n\} \in C_1(K;\mathbb{Z}_2)$

$\partial_1(q) = \partial(\{x_0,x_1\}) + \partial_1(\{x_1,x_2\}) + \dots + \partial_1(\{x_{n-1},x_n\})$

$= \cancel{x_1} - x_0 + \cancel{x_2} - \cancel{x_1} + \cancel{\dots} + x_n - \cancel{x_{n-1}}$

$$= x_n - x_0 = b_{j(x)} - x$$

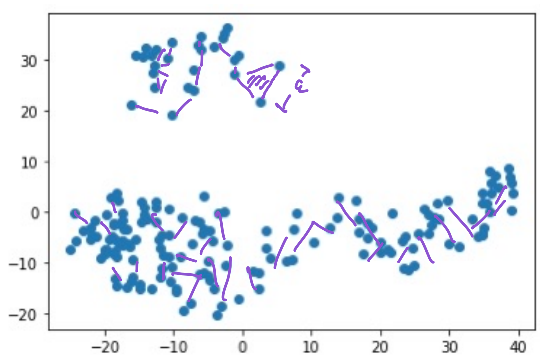
$$\Rightarrow b_{j(x)} - x \in B_0(K;\mathbb{Z}_2) = \mathrm{Im}(\partial_1)$$

$$\Rightarrow [b_{j(x)}] = [x] \text{ en } H_0(K;\mathbb{Z}_2) \quad \blacksquare$$

Prop: $\mathcal{B} = \{[b_j] \in H_0(K;\mathbb{Z}_2) \mid j \in J\}$ es linealmente independiente en $H_0(K;\mathbb{Z}_2)$.

Thm: $\dim(\boxed{H_0(K;\mathbb{Z}_2)}) = \#(J) =$ # componentes conexas por caminos de K.

Agrupamientos (Clustering)



Pregunta 1: Dado un conjunto de datos (X, d_X), cuantos clusters hay? Quienes son?

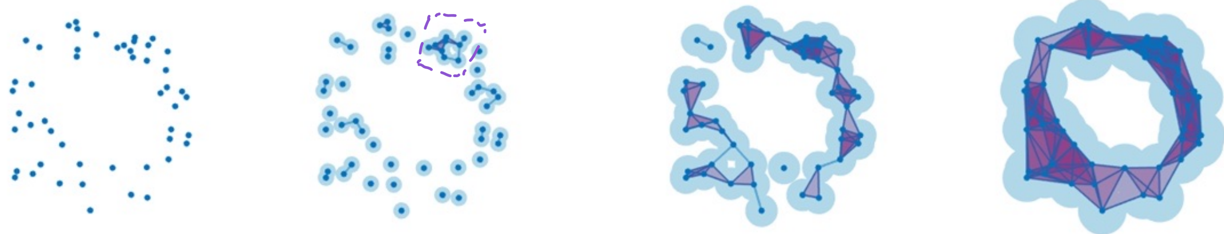
Idea 1: Calcule el ε-complejo de Rips (ε apropiado)

$$R_\varepsilon(X, d_X) = \{\sigma \subseteq X \mid \sigma \neq \emptyset \text{ y } \mathrm{diam}(\sigma) < \varepsilon\}.$$

$H_0(R_\varepsilon(X);\mathbb{Z}_2) \rightarrow$ dim = # clusters

base = clusters

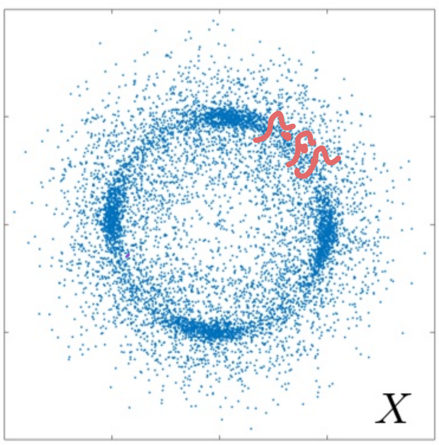
Pregunta 2: Como fijo $\varepsilon > 0$?



$X \qquad R_\varepsilon(X) \subseteq R_{\varepsilon'}(X) \subseteq R_{\varepsilon''}(X)$

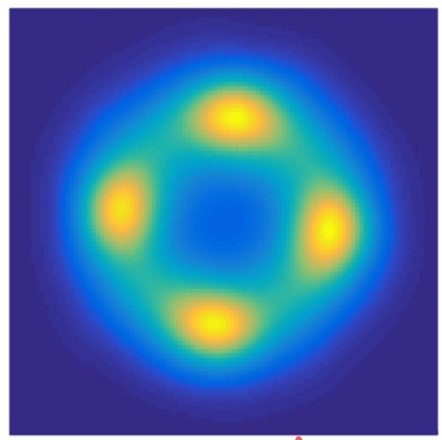
$0 < \varepsilon < \varepsilon' < \varepsilon''$

Pregunta 3:



Que pasa si hay ruido?

Idea 3: Considere $R_\varepsilon(X)^{\delta,r}$ para $f: X \to [0,\infty)$



un kernel-estimador de densidad:

$$f(x) = \frac{1}{h \cdot \#(X)} \sum_{x' \in X} Q\left(\frac{d_X(x,x')}{h}\right)$$

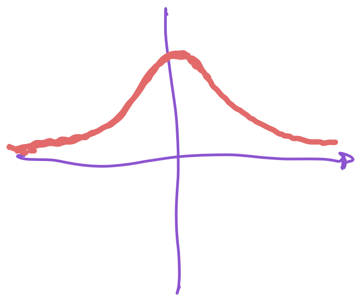
donde $h > 0$ es el ancho de banda (como fijar?)
y $Q : \mathbb{R} \to [0,\infty)$ es el kernel:

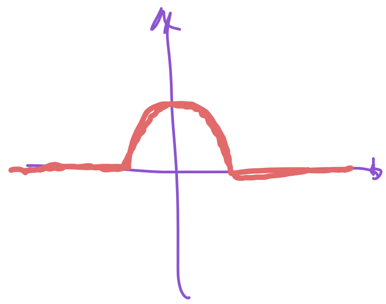
Gaussian:

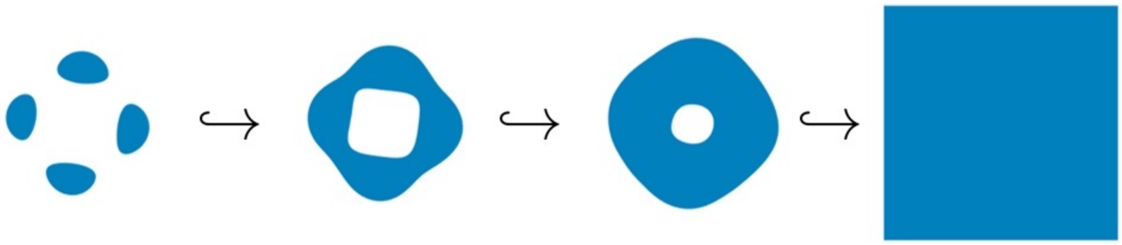
$$Q(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Parabolico (Epanechnikov)

$$Q(u) = \frac{3}{4} \max\{0, 1-u^2\}$$







$$R_\varepsilon(X)^{\geq r''} \subseteq R_\varepsilon(X)^{\geq r'} \subseteq R_\varepsilon(X)^{\geq r}$$

$$r < r' < r''$$

Problema: Entender

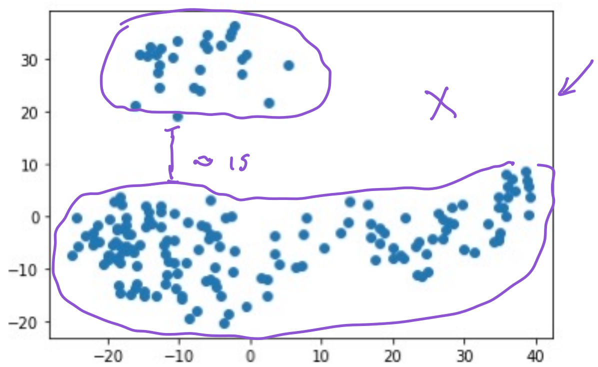
Persistencia Homológica

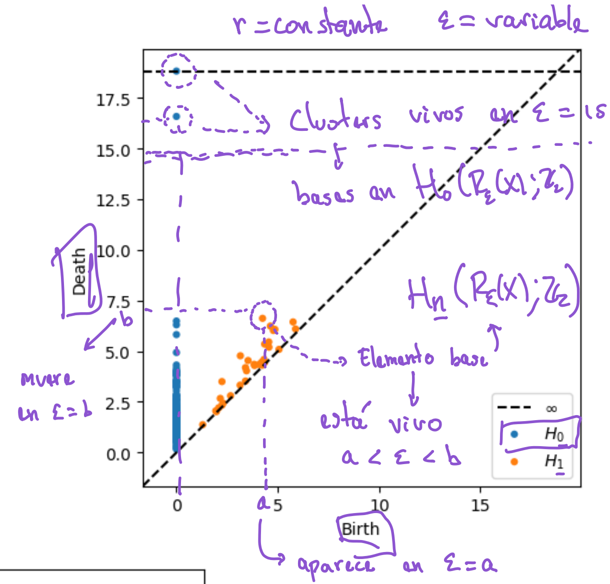
$$H_0(\underline{R_\varepsilon(X)^{\geq r}}; \mathbb{Z}_2)$$

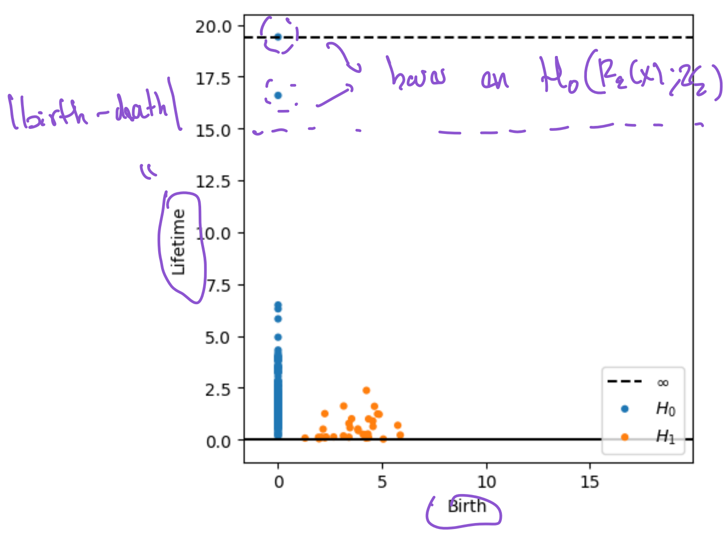
clusters

ε	r
variable	cotante
constante	variable.
variable	variable.

Diagramas de Persistencia.


30
20
10
0
−10
−20
−20
−10
0
10
20
30
40
≈ 15
X


r = constante   ε = variable
Clusters vivos en ε = 15
bases en $H_0(R_\varepsilon(X);\mathbb{Z}_2)$
$H_n(R_\varepsilon(X);\mathbb{Z}_2)$
Elemento base
está vivo
a < ε < b
muere en ε = b
aparece en ε = a
Death
Birth
∞
H_0
H_1
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
5
10
15


|birth - death|
"
bases en $H_0(R_\varepsilon(X);\mathbb{Z}_2)$
Lifetime
Birth
∞
H_0
H_1
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
5
10
15