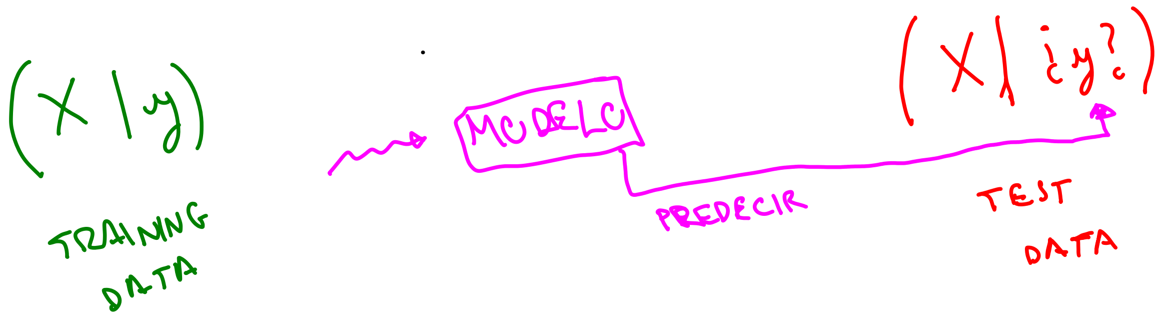
# K - Medias

## § 1. Aprendizaje no supervisado

Aprendizaje supervisado: Lo que hemos hecho hasta ahora. Teníamos unos training data de los que conocíamos los valores de la variable objetivo y los usábamos para predecir valores de la variable objetivo en los test data.




Aprendizaje no supervisado: No conocemos los valores de la variable objetivo, ni siquiera en los training data.

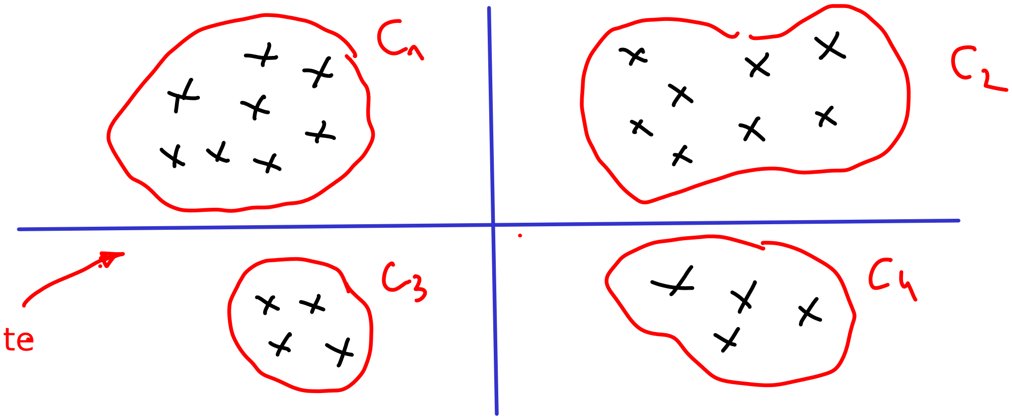
$(X \mid ¿y?)$

Ejemplo: Quiero detectar productos defectuosos en una línea de producción. Puedo hacer fotos de todos pero no puedo ir uno por uno distinguiendo defectuosos de no defectuosos. Busco un algoritmo que directamente aprenda a distinguirlos.

Nosotros concretamente vamos a estudiar algorimos de agrupamiento (clustering):

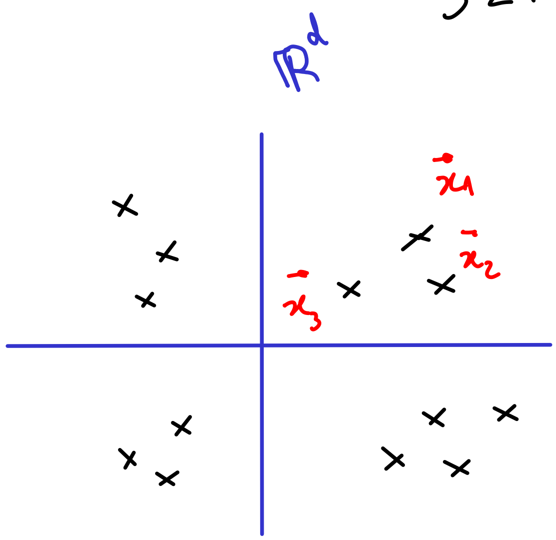
Supongamos que vemos muchas flores. No sé de qué especie son pero puedo tratar de agruparlas por especies según sus relaciones.

Idea gráfica:




El clustering permite identificar estos grupos

## §2. K-Medias




$\{\vec{x}_1, \dots, \vec{x}_m\} \subset \mathbb{R}^d$

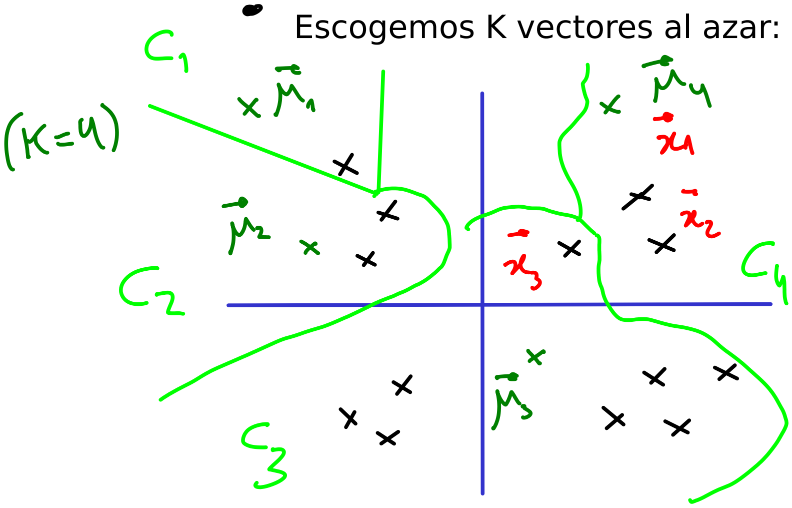
↑ DATOS

Tengo que especificar el número K de agrupamientos que quiero obtener

- Escogemos K vectores al azar:

$\vec{\mu}_1, \dots, \vec{\mu}_K$ ← "CENTROIDES"

$C_1, \dots, C_K$ ← CLUSTERS




Asignamos unas etiquetas:

$$r_{ij} = \begin{cases} 1 & \text{si } \vec{x}_i \in C_j \\ 0 & \text{si } \vec{x}_i \notin C_j \end{cases}$$

$i = 1, \dots, m$

$j = 1, \dots, K$

OJO: $\text{Si } r_{ij} = 1 \Rightarrow r_{ij'} = 0 \quad \forall\, j' \neq j$ (✱) CONDICIÓN

Función de coste: La suma de los cuadrados de las distancias de cada dato al representante de su agrupamiento.

$$J = \sum_{i=1}^{n} \sum_{j=1}^{K} r_{ij} \underbrace{\|\vec{x}_i - \vec{\mu}_j\|^2}_{= d(\vec{x}_i, \vec{\mu}_j)^2}$$

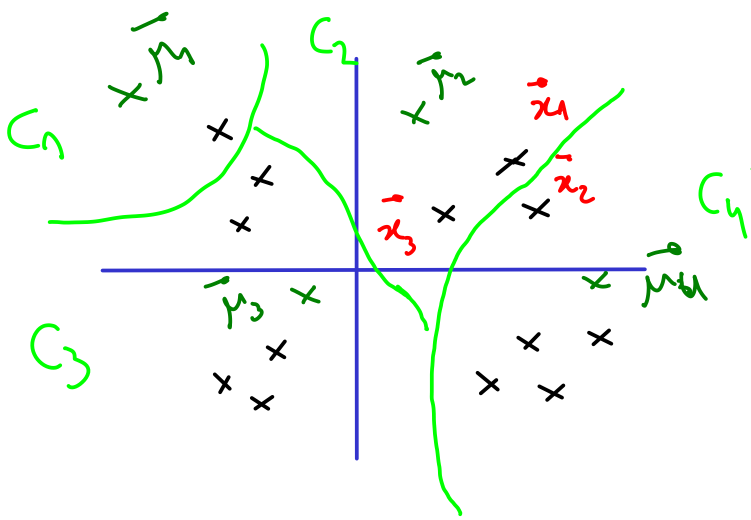
La minimización de la función de coste nos va a dar un algoritmo:

1. ASIGNACIÓN: Cómo asignar los $\vec{x}_i$ a los $\vec{\mu}_j$

   Es decir, cómo construir los agrupamientos.
   O, lo que es lo mismo, hallar los $r_{ij}$ óptimos

$$r_{ij} = \begin{cases} 1 & \text{si} \quad j = \underset{j}{\operatorname{argmin}} \|\vec{x}_i - \vec{\mu}_j\|^2 \\ 0 & \text{en otro caso} \end{cases}$$

La idea: Asignar cada xi al centroide que tenga más cerca




Se forman diagramas de Voronoi

2. ACTUALIZACIÓN: Mejorar en los $\vec{\mu}_j$

Buscamos unos nuevos centroides que se ajusten mejor a los grupos creados.
Para ello buscamos el mínimo de la función de coste en los centroides.

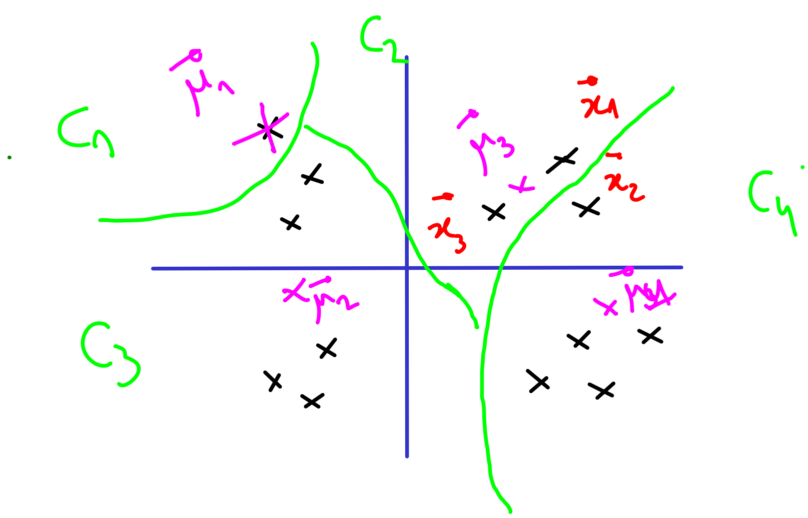
$$J = \sum_{i=1}^{n} \sum_{j=1}^{K} r_{ij} \, \|\vec{x}_i - \vec{\mu}_j\|^2$$

$$\frac{\partial J}{\partial \vec{\mu}_j} = 2 \sum_{i=1}^{n} r_{ij} \, (\vec{x}_i - \vec{\mu}_j) = 0$$

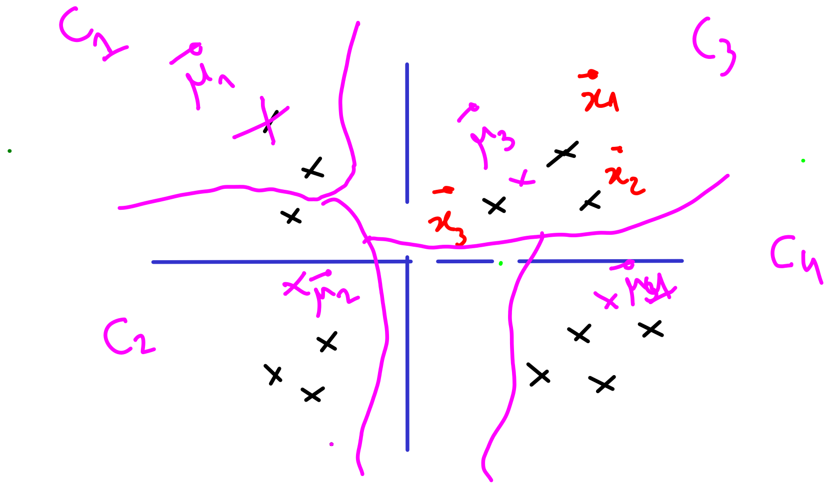
$$\Downarrow$$

$$\boxed{\vec{\mu}_j = \frac{\sum_{i=1}^{m} r_{ij} \, \vec{x}_i}{\underbrace{\sum_{i=1}^{m} r_{ij}}_{\text{puntos en } C_K}}}$$

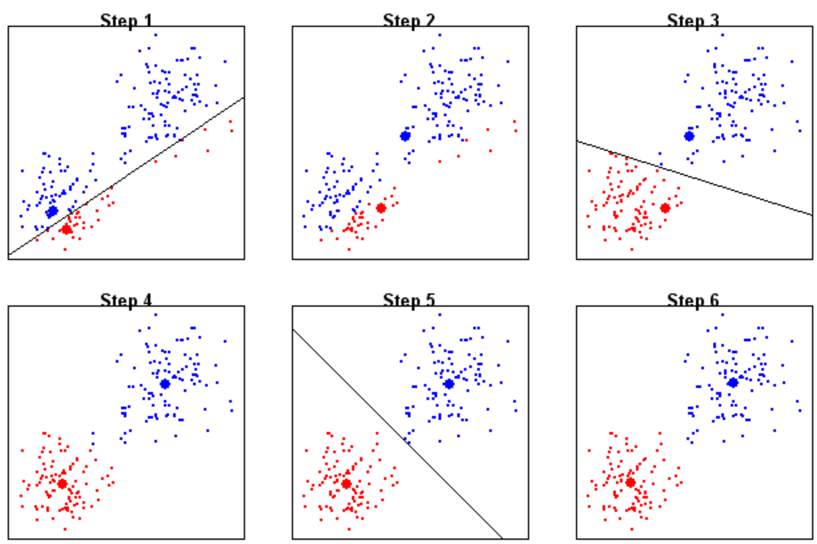
Los centroides óptimos son simplemente las medias de los agrupamientos (de ahí el nombre K-Medias).




Ahora volveríamos al paso 1




Seguimos iterando hasta alcanzar la convergencia. Esto es, hasta que los centroides (y los agrupamientos) ya no cambien más.



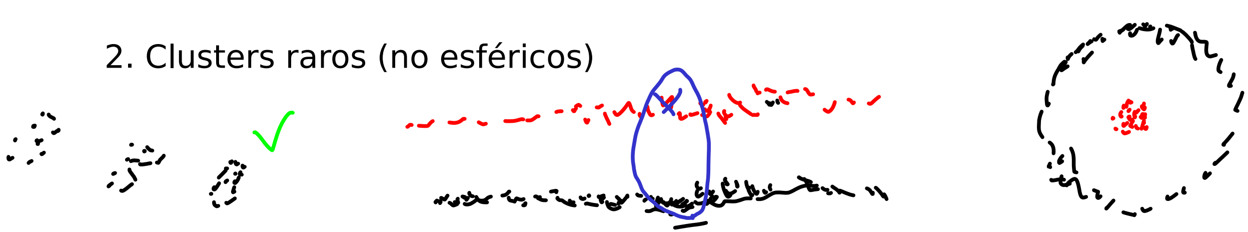

Ver vídeo: https://www.youtube.com/watch?v=5I3Ei69I40s

Ver vídeo: https://www.youtube.com/watch?v=yR7k19YBqiw&t

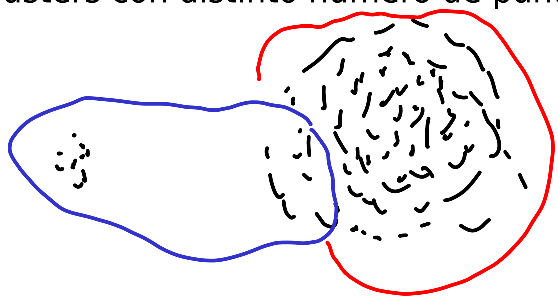
## §3. Problemas

1. ¿Número de agrupamientos? ¿Cuál es el K óptimo?

2. Clusters raros (no esféricos)



3. Clusters con distinto número de puntos



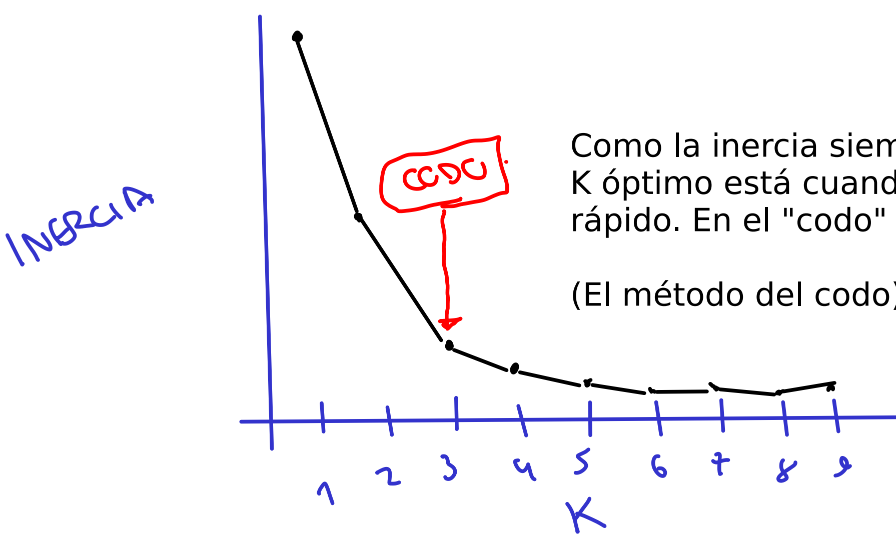
## § 4. ¿Cómo sabemos el K óptimo?

Inercia = $J_{mín}$

La inercia es el mínimo que alcanza la función coste. Es decir el valor de J en la convergencia.

La inercia esencialmente depende de los vectores de inicialización y del número de clusters.

El problema es que la inercia siempre disminuye con el número de clusters, sin embargo no queremos hacer más clusters de los que realmente hay. ¿Cómo lo resolvemos?




Como la inercia siempre disminuye, el K óptimo está cuando deja de disminuir rápido. En el "codo"

(El método del codo)