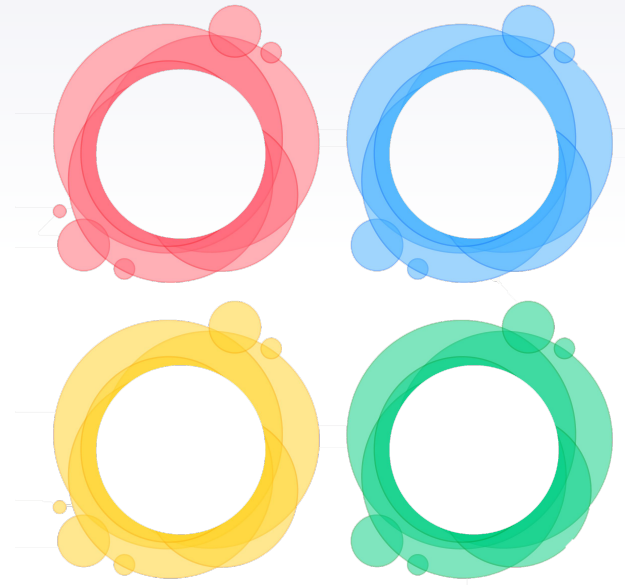




Clustering con incertidumbre

Correa López, Florentina
Esteban Castrillo, Álvaro

Universidad de Sevilla
Ciencias de la Computación e Inteligencia Artificial
Curso 2019/2020



Contenido

- ▶ Introducción
- ▶ Desarrollo del código
- ▶ Casos experimentales
- ▶ Conclusiones

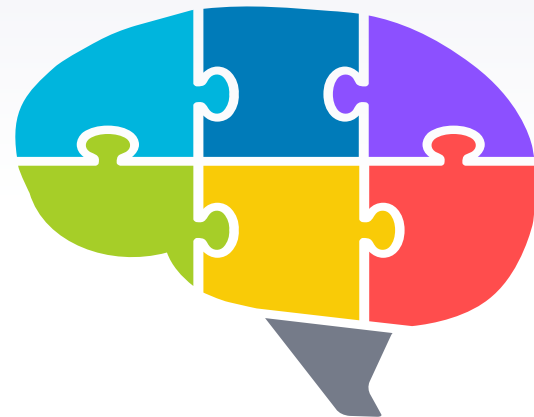


1

Introducción

¿Qué es el Clustering?

¿En qué está basado este trabajo?





Es un subproblema real del experimento LHCb en el CERN.





“Es un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio”

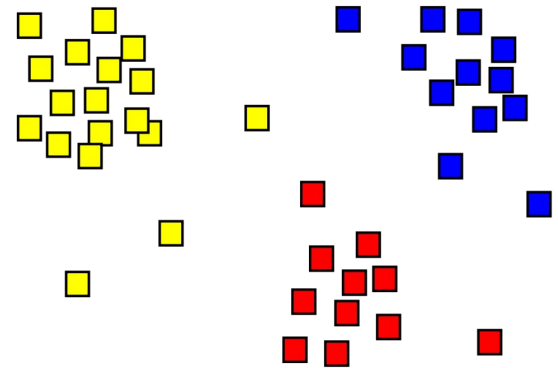
Ref. Wikipedia.



WIKIPEDIA
The Free Encyclopedia

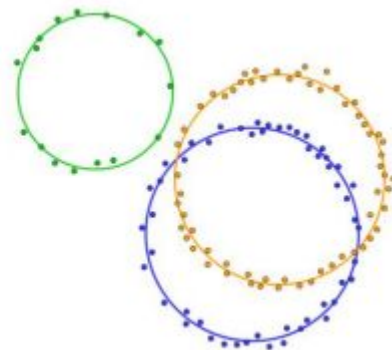
Tipos de algoritmos clustering

- ▶ K-medias (centroides)
 - ▶ Fuzzy Clustering (agrupamiento difuso)
- ▶ DBSCAN (modelo de densidad)
- ▶ Mean Shift (modelo de gradientes)
- ▶ AGNES



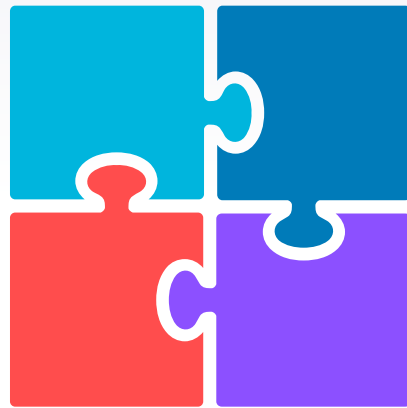
► Algoritmo propuesto

El objetivo es identificar el centro y el radio automáticamente de los clusters y asignar los puntos a todas las circunferencias aunque con distinto "grado de pertenencia".



2

Desarrollo del código



► ¿Por dónde empezamos?

- ▶ Resolución de circunferencias separadas
- ▶ Número de iteraciones preestablecidas
- ▶ Cálculo de cercanía a cada centro
- ▶ Cálculo de nuevos centros y radios

¿Qué cambios realizamos?

- ▶ Creación de nuevos métodos
- ▶ Calcular el grado de pertenencia
- ▶ Refinar la forma de iterar en el algoritmo
- ▶ Mejora en la actualización de nuevos centros y radios

► Métodos creados

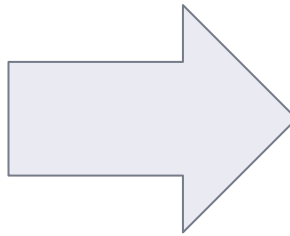
- ▶ `Media_puntos(V[puntos])`
- ▶ `Radio(V[puntos], centro)`
- ▶ `Grado_pertenencia(V[clusteres], V[puntos])`
- ▶ `Grado_pertenencia_cluster(V[clusteres], punto)`
- ▶ Otros.

► Cálculo del grado de pertenencia

$$P = |d(\text{punto}, \text{centro}) - \text{radio}|$$

► Iteraciones del algoritmo

Número de
iteraciones
preestablecido

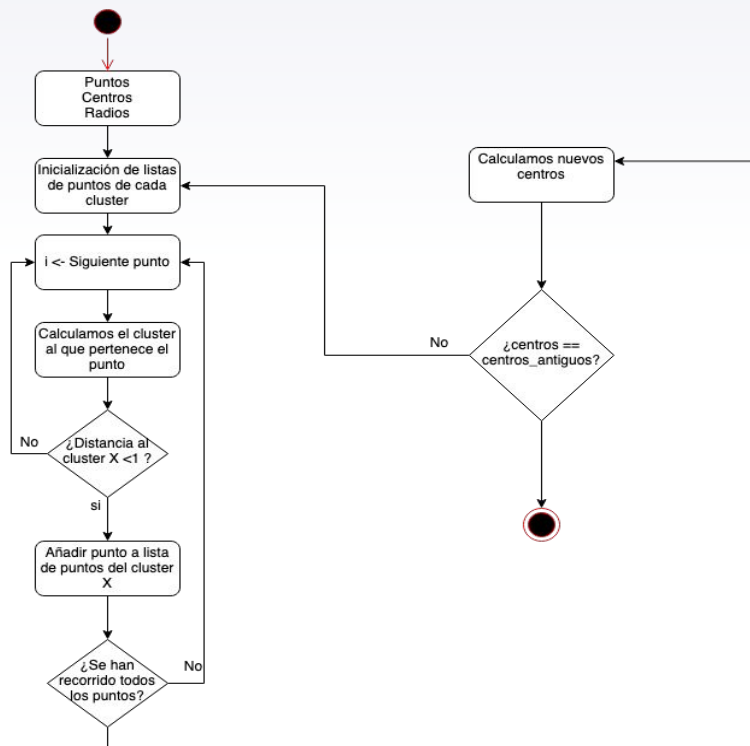


Itera hasta que
se fijan los
centros

► Cálculo de nuevos centros y radios

- ▶ Eliminación de ruido
- ▶ Tomamos los puntos pertenecientes al cluster

► Resultado final



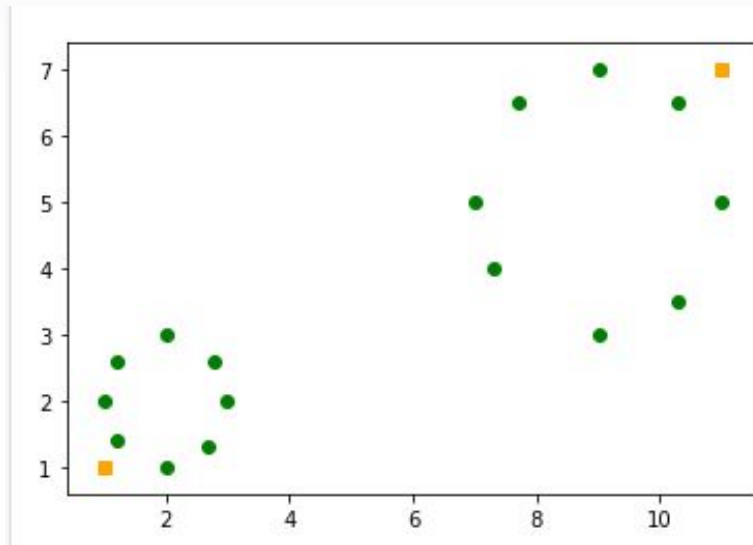
3

Casos experimentales



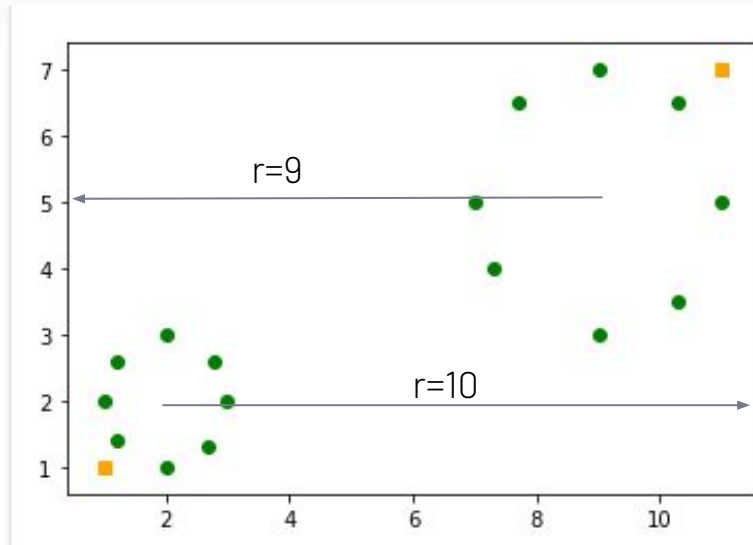
Anillos separados

Los valores iniciales de centro se han calculado con los métodos *punto_mayor* y *punto_menor*.



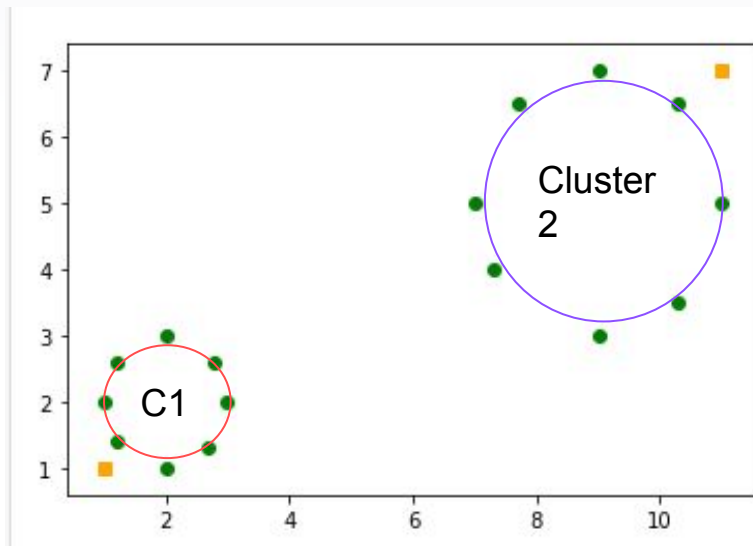
Anillos separados

Los valores iniciales de los radios han sido escogidos mediante una función de aleatoriedad `RandInt` de la librería `Random`



Anillos separados

Para la asignación a los cluster se ha utilizado la distancia del punto al centro de la circunferencia.

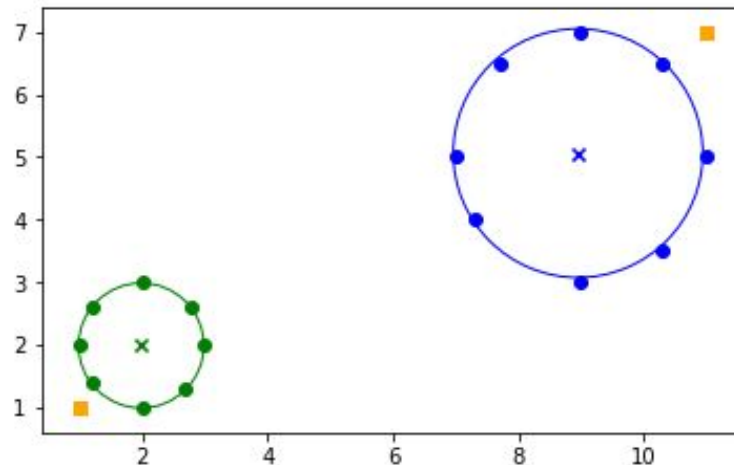


Resultado esperado

Anillos separados

Resultado
experimental:

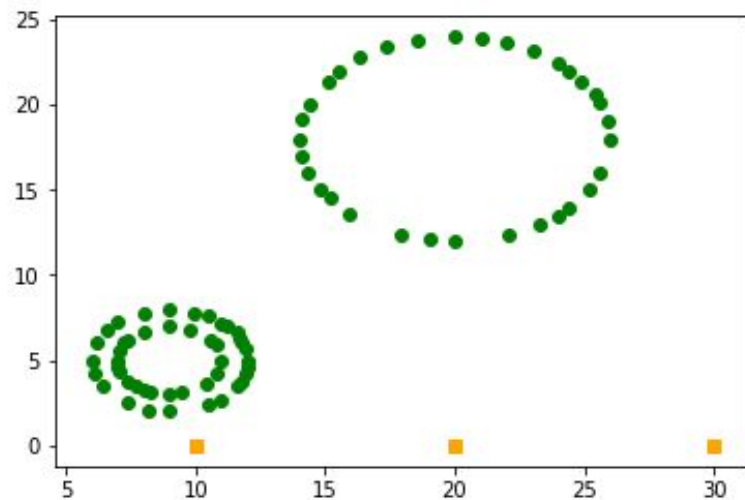
```
Centro 1 inicial: (1, 1)
Radio 1 inicial: 9
Centro 2 inicial: (11, 7)
Radio 2 inicial: 10
Centro 1 final: (1.9874999999999998, 1.9875)
Radio 1 final: 0.9985099503625606
Centro 2 final: (8.95, 5.0625)
Radio 2 final : 1.9884534426156921
```



Anillos concéntricos y un anillo exterior

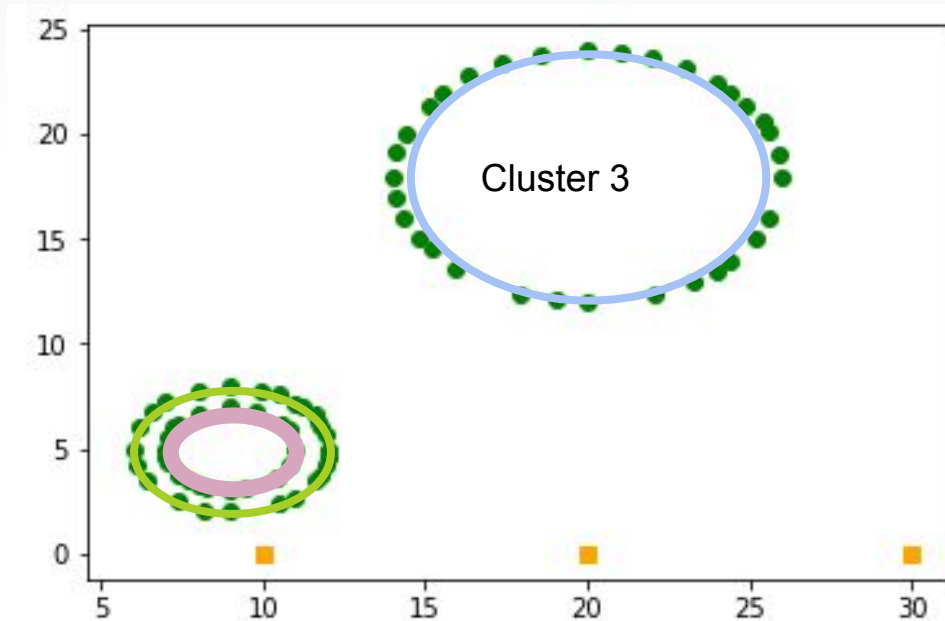
Los valores iniciales de centro y radio se han prefijado para obtener un mejor resultado.

Se ha utilizado el grado de pertenencia.



Anillos concéntricos y un anillo exterior

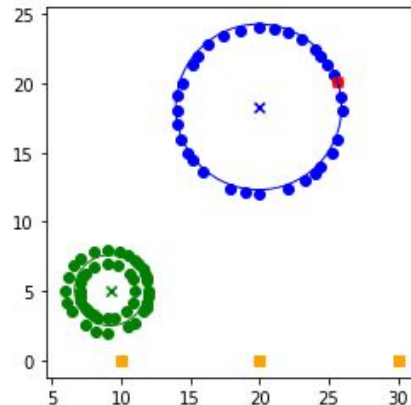
Resultado
esperado:



Anillos concéntricos y un anillo exterior

Resultado experimental:

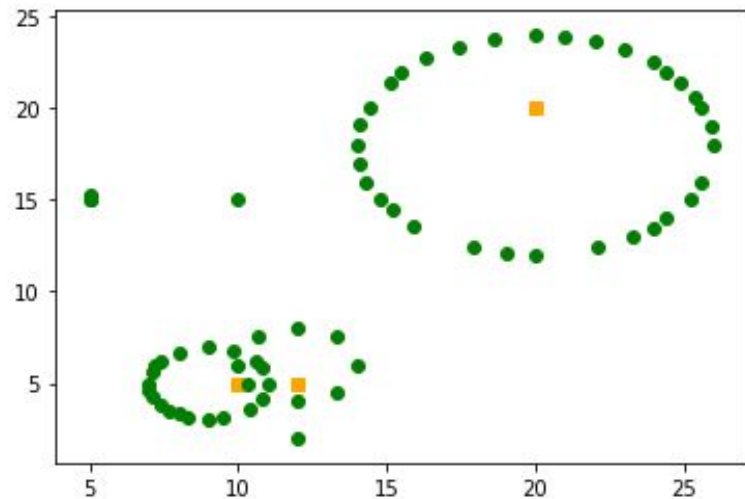
```
Centro 1 inicial: (10, 0)
Radio 1 inicial: 2
Centro 2 inicial: (20, 0)
Radio 2 inicial: 2
Centro 3 inicial: (30, 0)
Radio 3 inicial: 2
Centro 1 final: (9.20434782608696, 5.054347826086956)
Radio 1 final: 2.559197513396158
Centro 2 final: (19.931250000000002, 18.293750000000003)
Radio 2 final : 5.980972491273866
Centro 3 final: (25.6, 20.1)
Radio 3 final: 0.0
```



Anillos con intersecciones y un anillo exterior

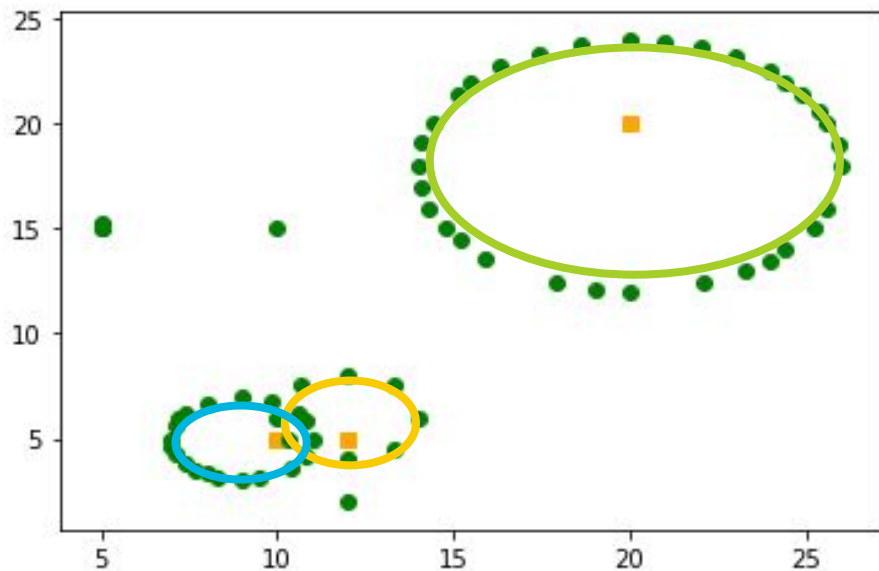
Los valores iniciales de centros y radios han sido prefijados.

Se ha utilizado el grado de pertenencia y además la distancia del punto a la circunferencia.



Anillos con intersecciones y un anillo exterior

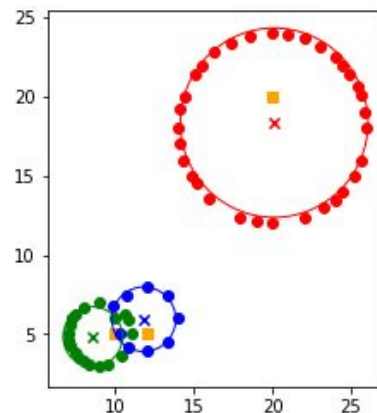
Resultado
esperado:



Anillos con intersecciones y un anillo exterior

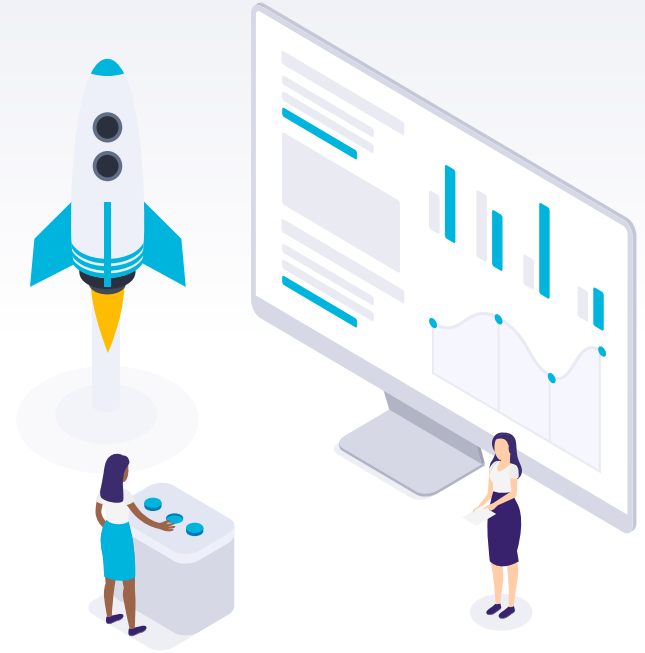
Resultado experimental:

```
Centro 1 inicial: (10, 5)
Radio 1 inicial: 2
Centro 2 inicial: (12, 5)
Radio 2 inicial: 2
Centro 3 inicial: (20, 20)
Radio 3 inicial: 2
Centro 1 final: (8.55263157894737, 4.836842105263156)
Radio 1 final: 1.8902424263392141
Centro 2 final: (11.799999999999999, 5.944444444444445)
Radio 2 final: 2.0364267988520064
Centro 3 final: (20.1030303030303, 18.348484848484848)
Radio 3 final: 5.975421519460662
```



4

Conclusiones



► Línea del tiempo



Investigación

Anillos de Cherenkov,
clustering, tipos de
clustering, gráficas

Desarrollo del código

Métodos propios:
GradoPertenencia

Gráficas

Experimentación

Obtener resultados del
funcionamiento del
algoritmo

Documentación

Redactar el artículo
científico

► Valoraciones propias

- Puntos débiles:

Dificultad para obtener el resultado óptimo del algoritmo

Poca información para el desarrollo del código

- Puntos fuertes:

Implicación y trabajo en equipo

Buena gestión del tiempo

► Posibles mejoras

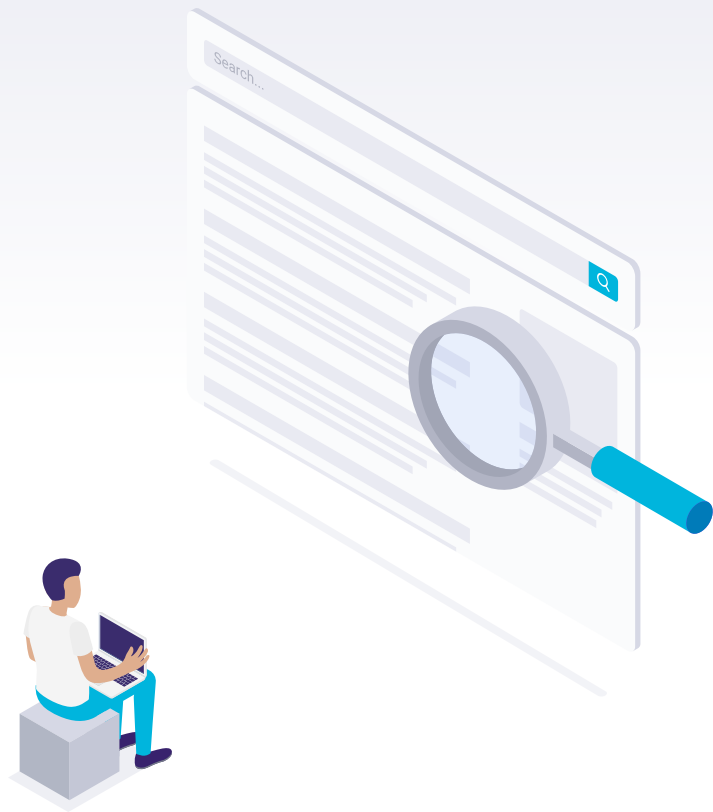
- ▶ Implementar el método del Codo de Jambú
- ▶ Generador automático de puntos en torno a una circunferencia
- ▶ Implementar otro tipo de algoritmo de clustering como DBSCAN

Gracias

¿Alguna pregunta?

Trabajo realizado por:

- ▶ Florentina Correa López
flocorlop@alum.us.es
- ▶ Álvaro Esteban Castrillo
alvestcas@alum.us.es



► Créditos

El diseño de esta presentación ha sido obtenida gratuitamente:

- ▶ Plantilla de presentación por [SlidesCarnival](#)
- ▶ Ilustraciones por [Sergei Tikhonov](#), [Freepik](#)