

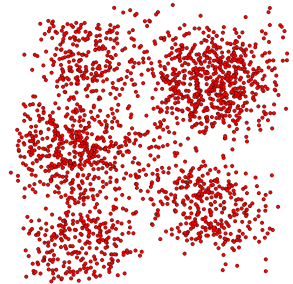
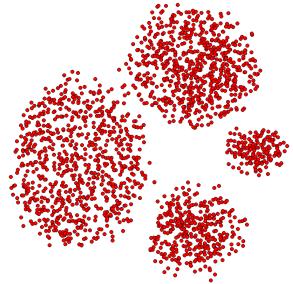
Aprendizaje no supervisado

Agrupamiento basado en densidad - DBSCAN

Javier Sevilla

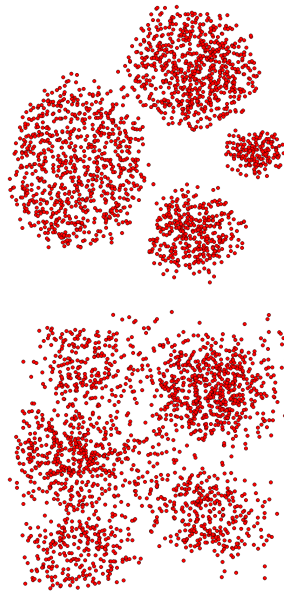
Tipos de algoritmos de agrupamiento

- ▶ Basados en particiones
- ▶ Jerárquicos
- ▶ Espectrales
- ▶ Basados en densidad
- ▶ Probabilísticos



Tipos de algoritmos de agrupamiento

- ▶ Basados en particiones
- ▶ Jerárquicos
- ▶ Espectrales
- ▶ **Basados en densidad**
- ▶ Probabilísticos



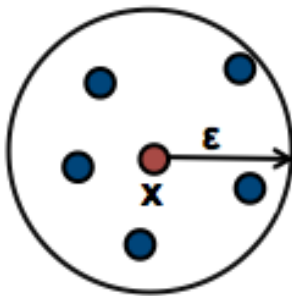
Agrupamiento basado en densidad

Conceptos

Densidad

¿Cómo definir densidad?

- Vecindario (parámetro ϵ)



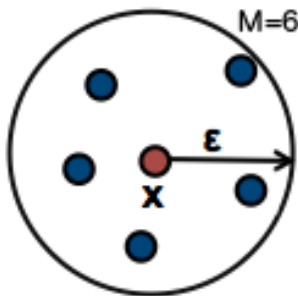
Agrupamiento basado en densidad

Conceptos

Densidad

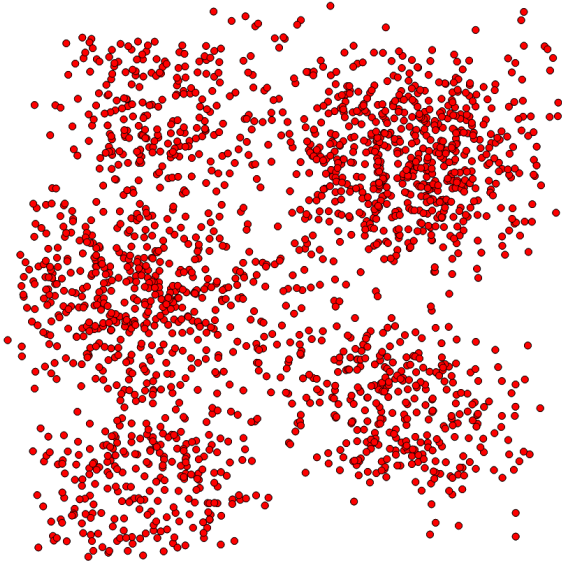
¿Cómo definir densidad?

- ▶ Vecindario (parámetro ϵ)
- ▶ Punto nuclear (parámetros ϵ , M)



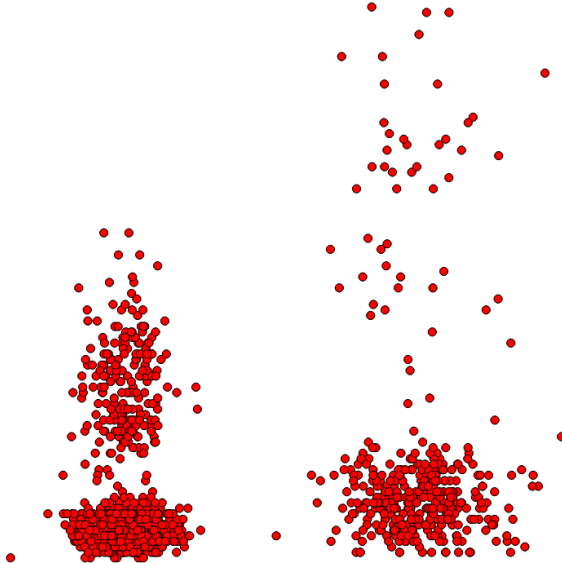
Agrupamiento basado en densidad

Conceptos



Agrupamiento basado en densidad

Conceptos



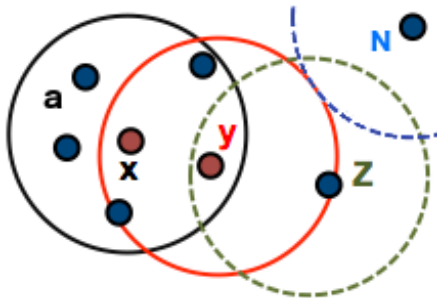
Agrupamiento basado en densidad

DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- Punto nuclear (x o y)



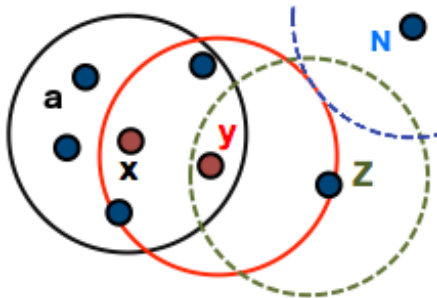
Agrupamiento basado en densidad

DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- ▶ Punto nuclear (x o y)
- ▶ Punto directamente denso-alcanzable (y desde x , z desde y)



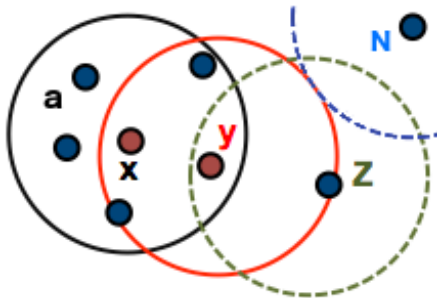
Agrupamiento basado en densidad

DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- ▶ Punto nuclear (x o y)
- ▶ Punto directamente denso-alcanzable (y desde x , z desde y)
- ▶ Punto borde (z)



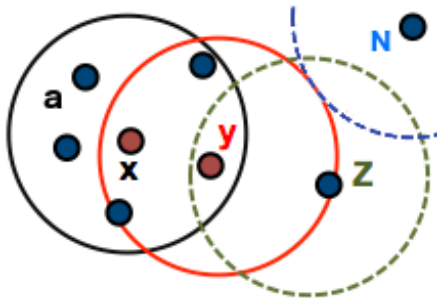
Agrupamiento basado en densidad

DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- ▶ Punto nuclear (x o y)
- ▶ Punto directamente denso-alcanzable (y desde x , z desde y)
- ▶ Punto borde (z)
- ▶ Punto ruido (n)



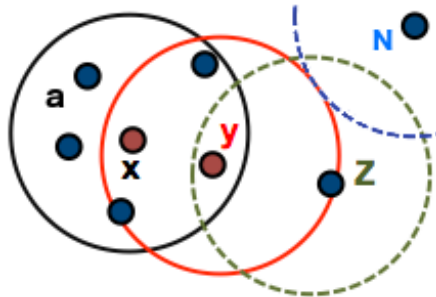
Agrupamiento basado en densidad

DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- Punto **directamente** denso-alcanzable (y desde x , z desde y)



Agrupamiento basado en densidad

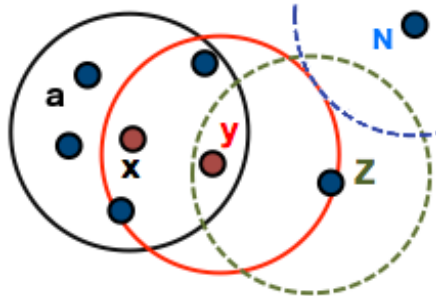
DBSCAN

Densidad

En base al valor de los parámetros ϵ y M , se definen:

- ▶ Punto **directamente** denso-alcanzable (y desde x, z desde y)
- ▶ Punto denso-alcanzable (z desde x)

** Relación asimétrica **



Agrupamiento basado en densidad

DBSCAN

Densidad

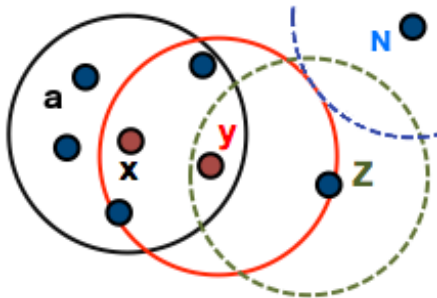
En base al valor de los parámetros ϵ y M , se definen:

- ▶ Punto **directamente** denso-alcanzable (y desde x, z desde y)
- ▶ Punto denso-alcanzable (z desde x)

** Relación asimétrica **

- ▶ Puntos denso-conectados (a y z)

** Relación simétrica **

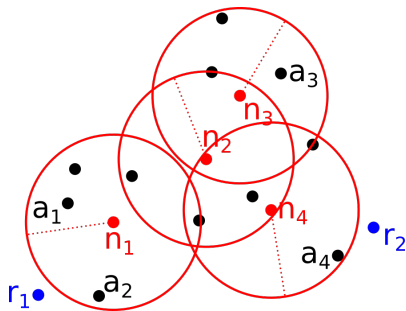


Agrupamiento basado en densidad

DBSCAN

Clúster

Conjunto de puntos nucleares denso-conectados y el resto de puntos (directamente) denso-alcanzables desde ellos.



Agrupamiento basado en densidad

DBSCAN

DBSCAN

1. $C = 1$

2. Para todo ejemplo, x_i

2.1. Si x_i ya está asignado, continuar (volver a 2)

2.2. Calcular vecindario V de x_i (dado ϵ)

2.3. Si $|V| < \mathcal{M}$, asignar x_i como ruido y continuar (volver a 2)

2.4. Crear clúster número C y asignarle x_i

2.5. Para todo ejemplo $x_j \in V$

2.5.1. Si x_j está asignado como ruido, asignarlo al clúster C y continuar (volver a 2.5)

2.5.2. Si x_j tiene otra asignación, continuar (volver a 2.5)

2.5.3. Asignar x_j al clúster C

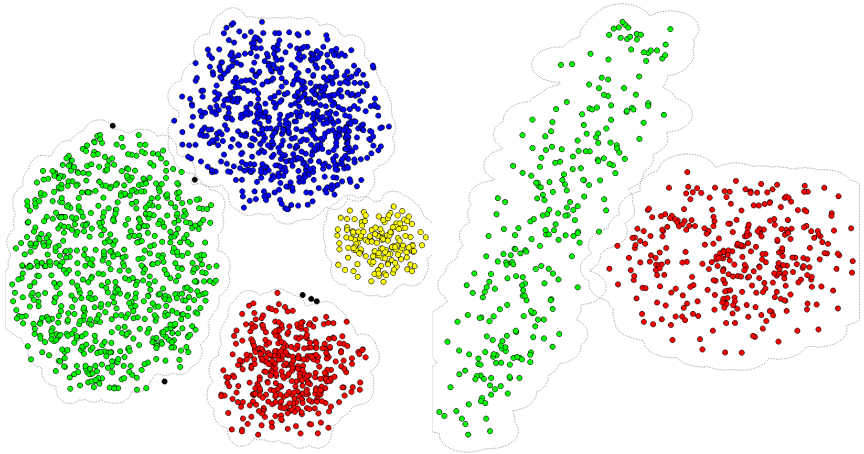
2.5.4. Calcular vecindario V' de x_j (dado ϵ)

2.5.5. Si $|V'| \geq \mathcal{M}$, $V = V \cup V'$

2.6. $C = C + 1$

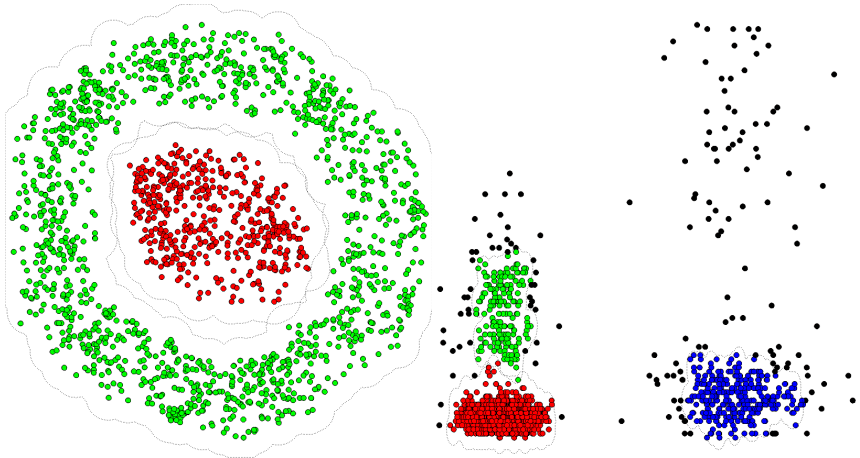
Agrupamiento basado en densidad

DBSCAN



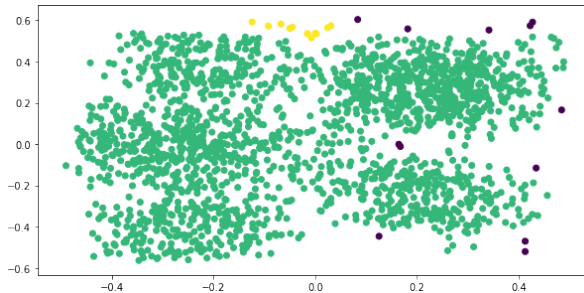
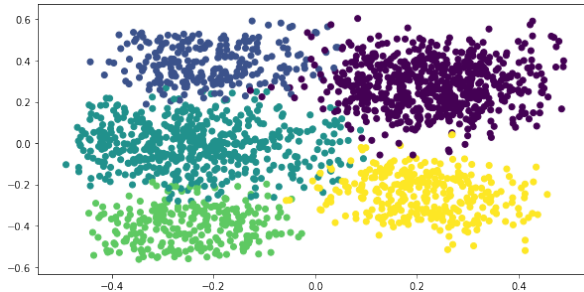
Agrupamiento basado en densidad

DBSCAN



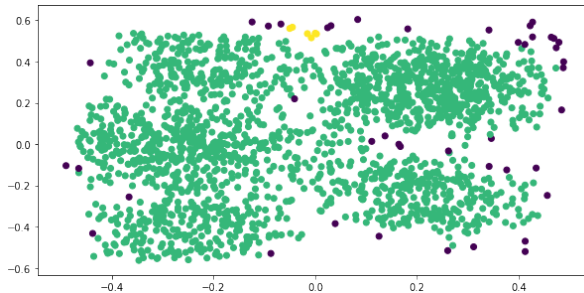
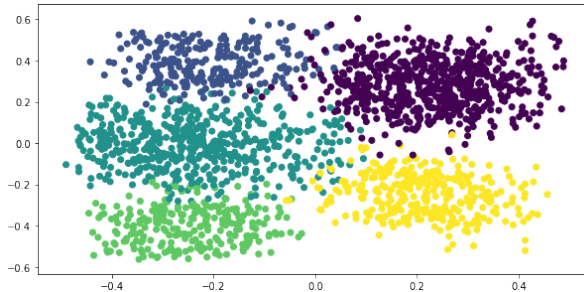
Agrupamiento basado en densidad

DBSCAN: Efecto de M



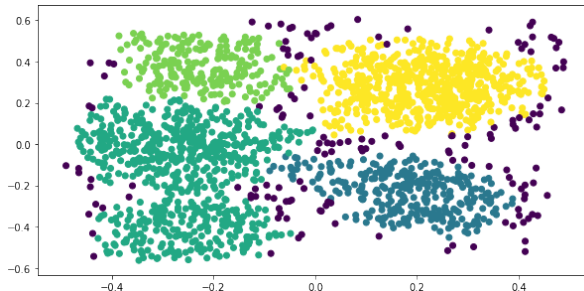
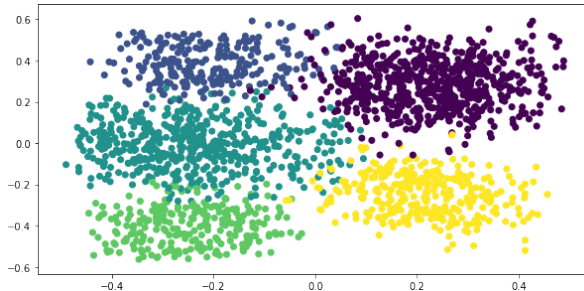
Agrupamiento basado en densidad

DBSCAN: Efecto de M



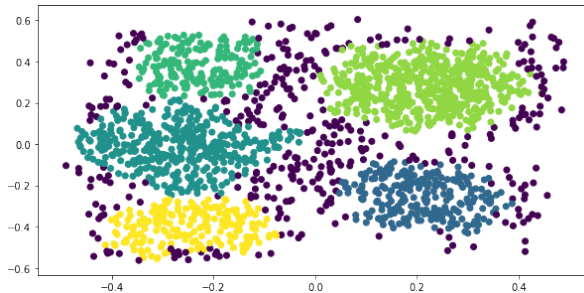
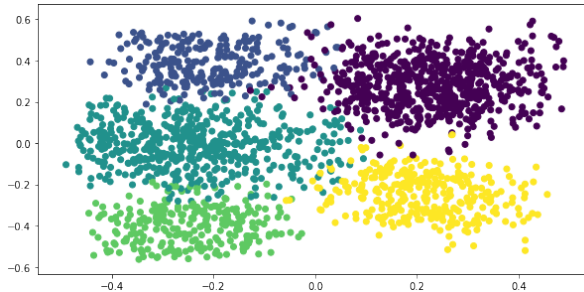
Agrupamiento basado en densidad

DBSCAN: Efecto de M



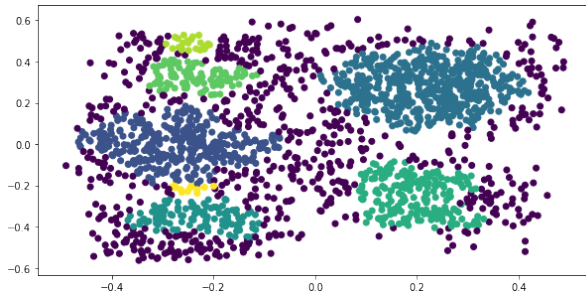
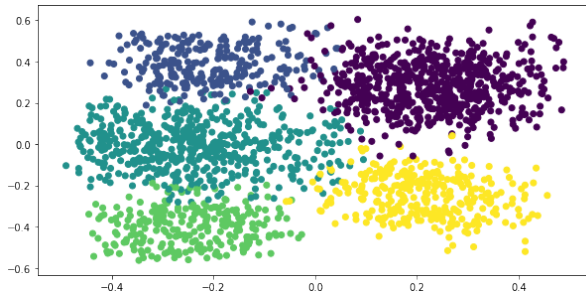
Agrupamiento basado en densidad

DBSCAN: Efecto de M



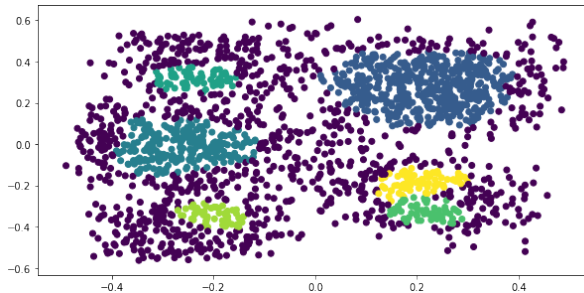
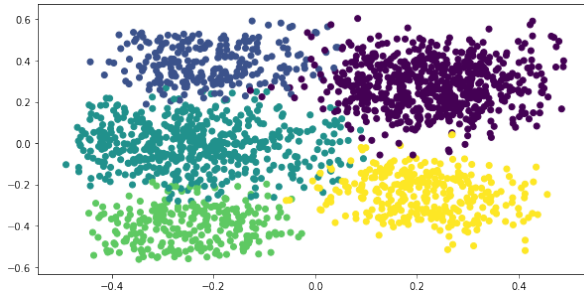
Agrupamiento basado en densidad

DBSCAN: Efecto de M



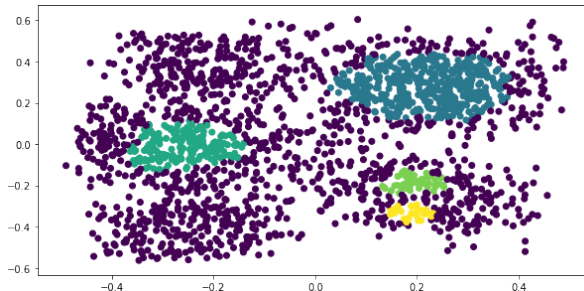
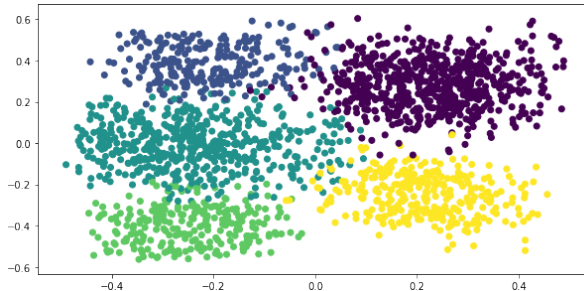
Agrupamiento basado en densidad

DBSCAN: Efecto de M



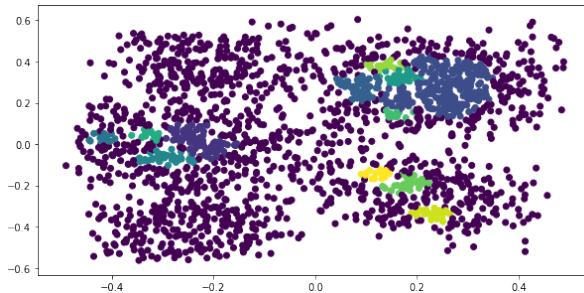
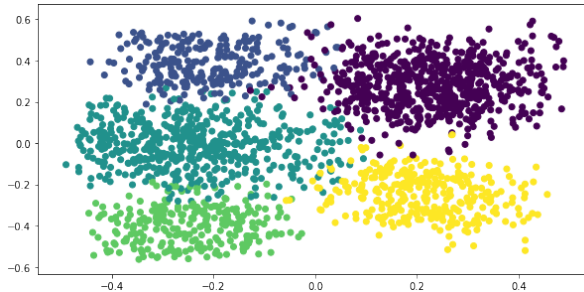
Agrupamiento basado en densidad

DBSCAN: Efecto de M



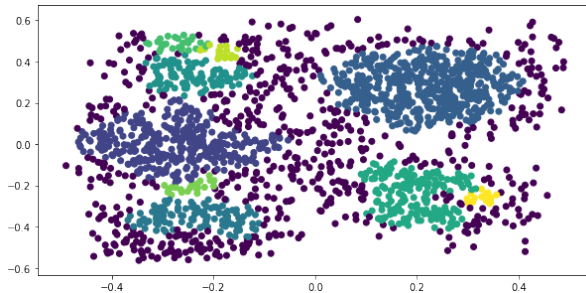
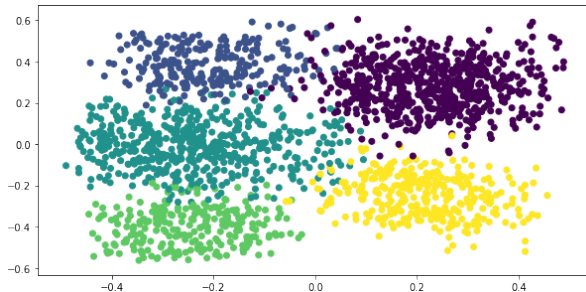
Agrupamiento basado en densidad

DBSCAN: Efecto de ϵ



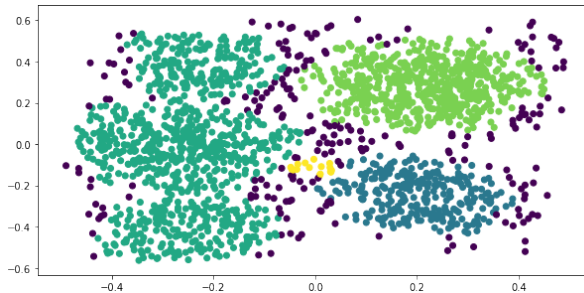
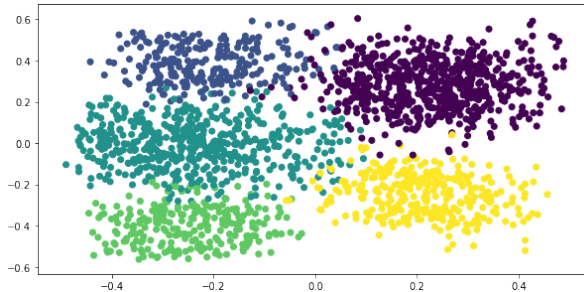
Agrupamiento basado en densidad

DBSCAN: Efecto de ϵ



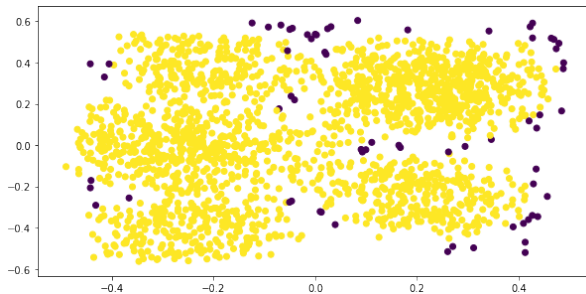
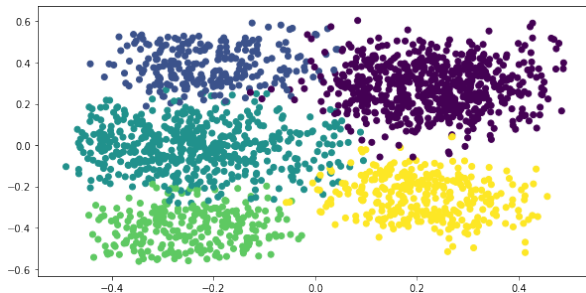
Agrupamiento basado en densidad

DBSCAN: Efecto de ϵ



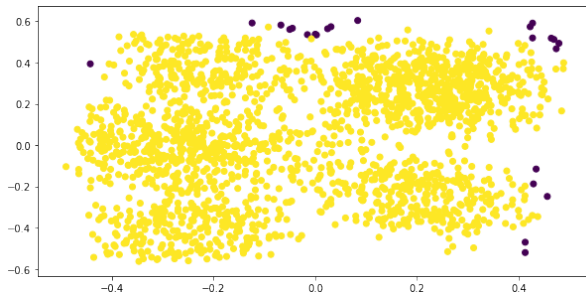
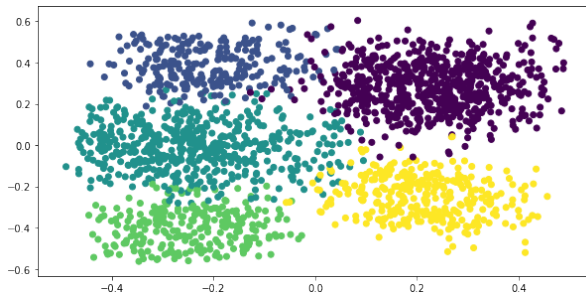
Agrupamiento basado en densidad

DBSCAN: Efecto de ϵ



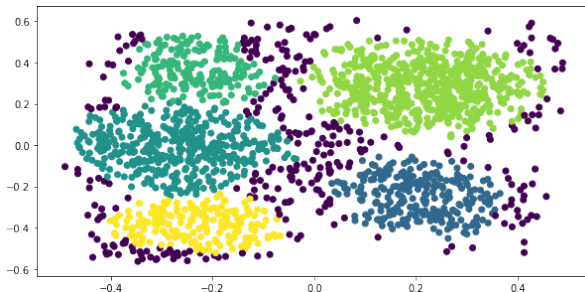
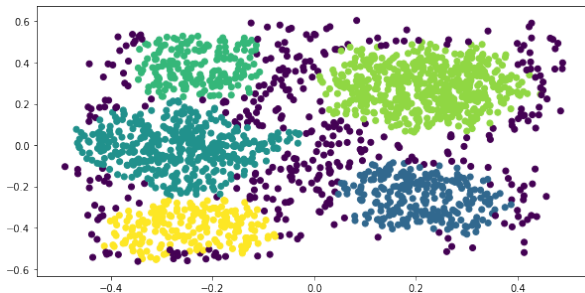
Agrupamiento basado en densidad

DBSCAN: Efecto de ϵ



Agrupamiento basado en densidad

DBSCAN: Efecto de ambos, M y ϵ



Ventajas

- ▶ No es necesario especificar K
- ▶ Definición basada en densidad
- ▶ Funciona con clústeres de diferente tamaño y formas
- ▶ Generalizable a otros conceptos de densidad
- ▶ Puede funcionar con diferentes medidas de distancia

Desventajas

- ▶ Definición compleja
- ▶ Problemas al lidiar con clústeres de diferente densidad
- ▶ Dos parametros interdependientes a ajustar

Gracias