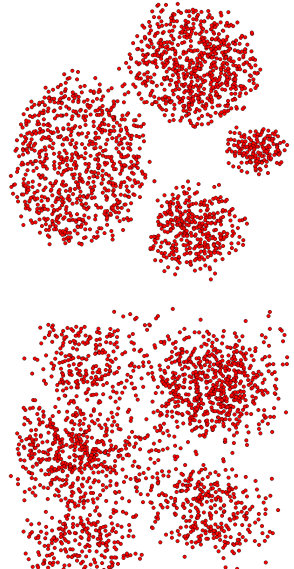


Agrupamiento

Tipos de algoritmos de agrupamiento

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



Agrupamiento espectral

Puntos básicos

1. Obtener un grafo y su matriz de adyacencias
2. Obtener una representación alternativa de los datos
3. Aplicar un algoritmo de agrupamiento estándar (K -means)

Agrupamiento espectral

Puntos básicos

1. **Obtener un grafo y su matriz de adyacencias**
2. Obtener una representación alternativa de los datos
3. Aplicar un algoritmo de agrupamiento estándar (K -means)

Agrupamiento espectral

Consideraciones previas

Matriz de similitud

- ▶ El algoritmo no funciona con la matriz original
- ▶ Se usa la **matriz de similitud**

Se transforma la matriz de ejemplos D ($n \times v$) en...
matriz de distancias, M ($n \times n$), tal que:

$$M_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$$

y ésta, a su vez, en una matriz de similitudes, S ($n \times n$):

$$S_{ij} = \exp(-M_{ij}^2 / 2 \cdot \sigma^2)$$

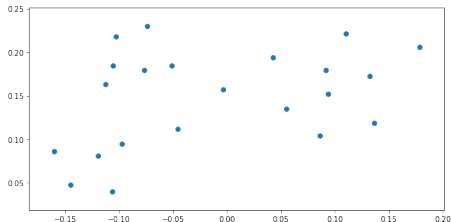
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Todos los nodos están conectados con todos
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo completo



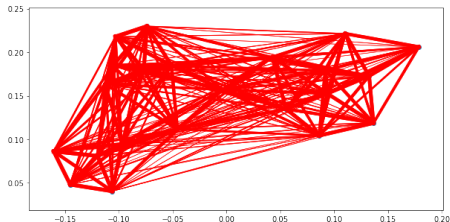
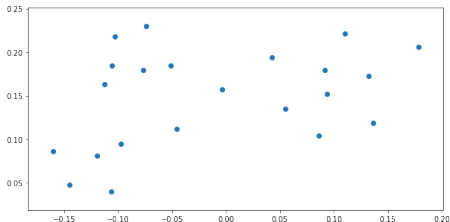
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Todos los nodos están conectados con todos
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo completo



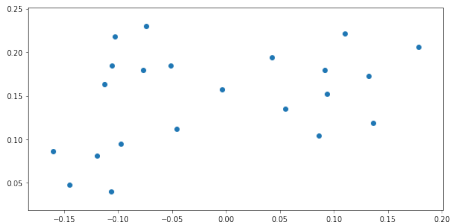
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Existe una arista entre dos nodos si hay un mínimo de similitud entre ambos, $S_{ij} > \epsilon$
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo umbral



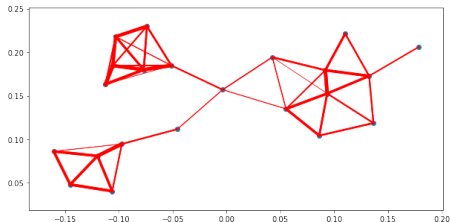
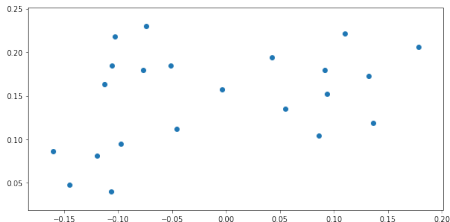
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Existe una arista entre dos nodos si hay un mínimo de similitud entre ambos, $S_{ij} > \epsilon$
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo umbral



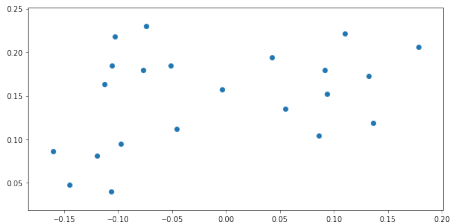
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Un nodo i tiene una arista con otro nodo j si j es uno de los K 'vecinos' más similar de i (o viceversa)
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo KNN



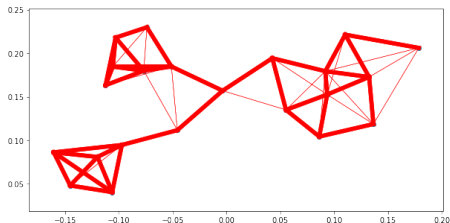
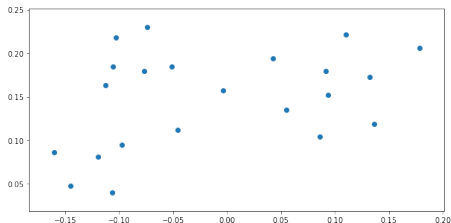
Agrupamiento espectral

Obtener un grafo y su matriz de adyacencias

Procedimientos de generación de un grafo

- ▶ Cada ejemplo de D es un nodo del grafo
- ▶ Un nodo i tiene una arista con otro nodo j si j es uno de los K 'vecinos' más similar de i (o viceversa)
- ▶ El peso de la arista entre dos nodos es su similitud, $W_{ij} = S_{ij}$

Grafo KNN



Agrupamiento espectral

Puntos básicos

1. Obtener un grafo y su matriz de adyacencias
2. Obtener una representación alternativa de los datos
3. Aplicar un algoritmo de agrupamiento estándar (K -means)

Agrupamiento espectral

Puntos básicos

1. Obtener un grafo y su matriz de adyacencias
2. **Obtener una representación alternativa de los datos**
3. Aplicar un algoritmo de agrupamiento estándar (K -means)

Agrupamiento espectral

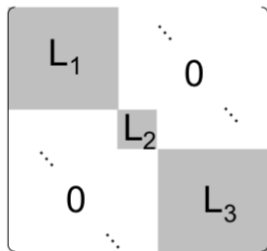
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = D - W$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$

Matriz Laplaciana básica



Agrupamiento espectral

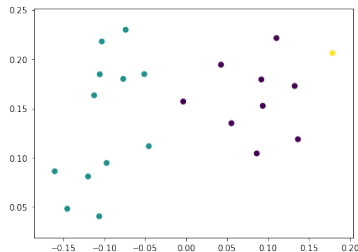
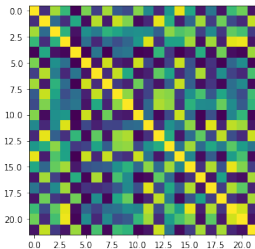
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = D - W$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$

Matriz Laplaciana básica



Agrupamiento espectral

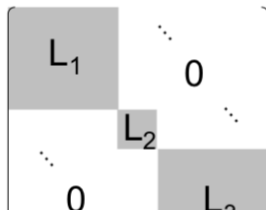
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = I - D^{-1}W$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$
- ▶ I : matriz identidad

Matriz Laplaciana normalizada (RW)



Agrupamiento espectral

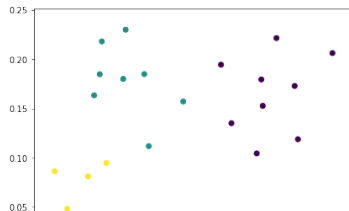
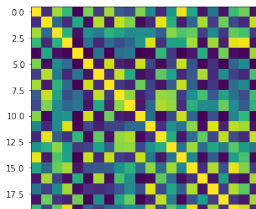
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = I - D^{-1}W$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$
- ▶ I : matriz identidad

Matriz Laplaciana normalizada (RW)



Agrupamiento espectral

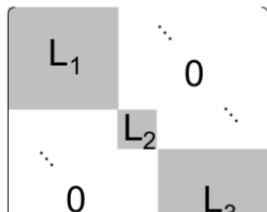
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = I - D^{-1/2} W D^{-1/2}$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$
- ▶ I : matriz identidad

Matriz Laplaciana normalizada simétrica



Agrupamiento espectral

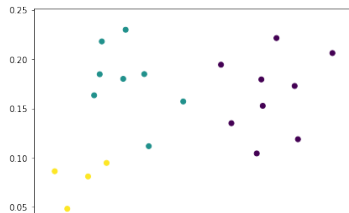
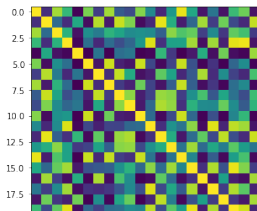
Obtener una representación alternativa de los datos

Procedimientos de la matriz Laplaciana

$$L = I - D^{-1/2} W D^{-1/2}$$

- ▶ W : matriz representativa del grafo (paso anterior)
- ▶ D : matriz diagonal con $D_{ii} = \sum_j W_{ij}$
- ▶ I : matriz identidad

Matriz Laplaciana normalizada simétrica



Agrupamiento espectral

Obtener una representación alternativa de los datos

Obtener datos transformados

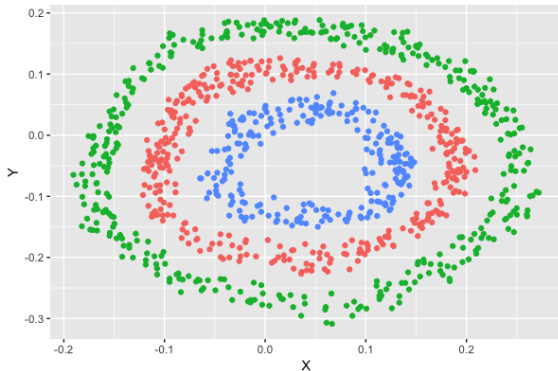
1. Descomponer la matriz L en vectores propios
2. Ordenar los vectores propios según el valor propio correspondiente (ascendente)
3. Seleccionar los primeros K vectores propios

Cada vector propio es una “variable” en el dataset transformado, que tiene n filas (tantas como ejemplos) y K variables (vectores propios):

$$D(n \times v) \rightarrow M(n \times K)$$

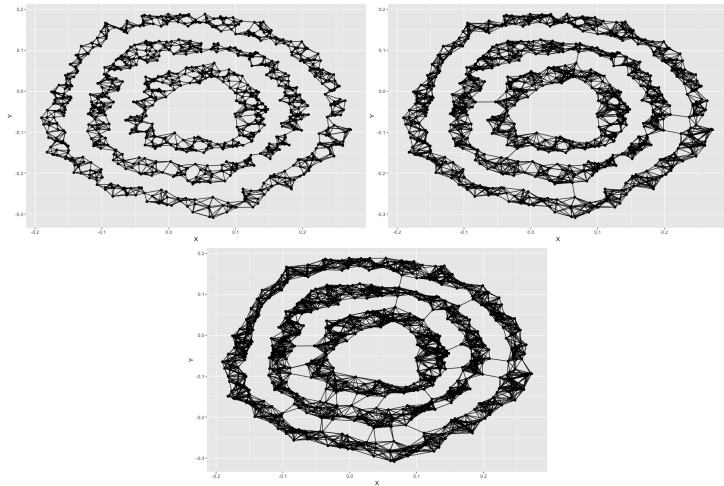
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



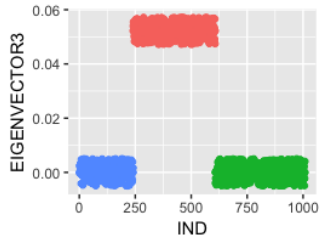
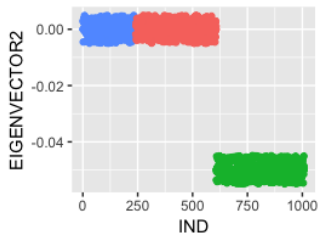
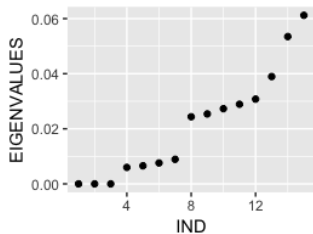
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



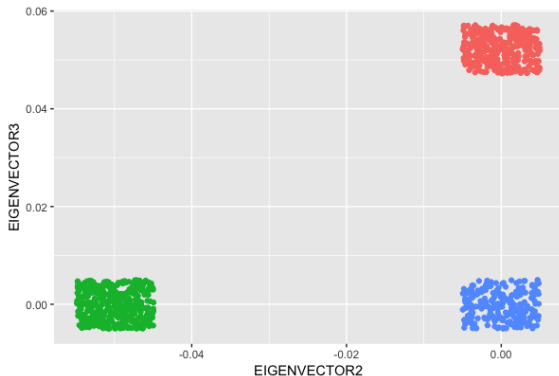
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



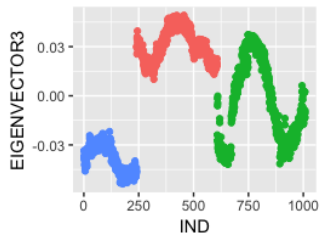
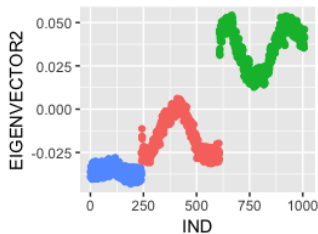
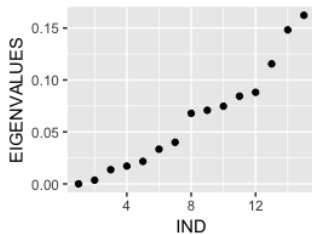
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



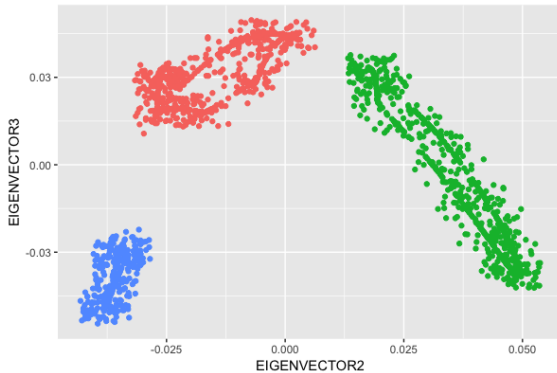
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



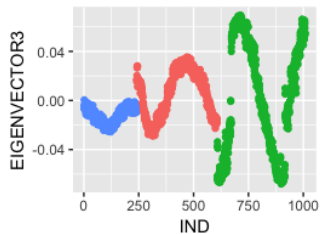
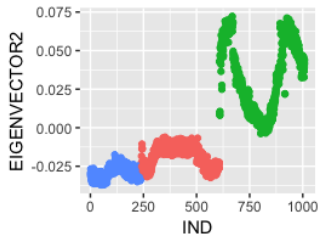
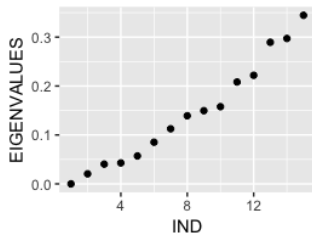
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



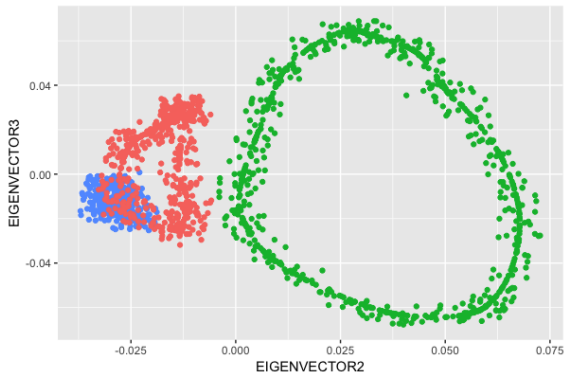
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



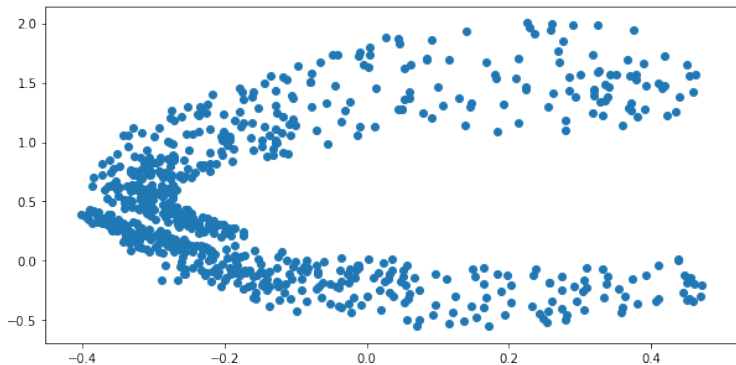
Agrupamiento espectral

Efecto de la K en KNN (generación del grafo)



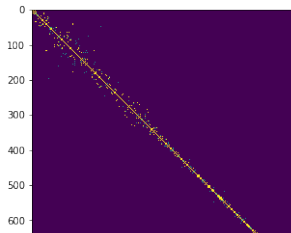
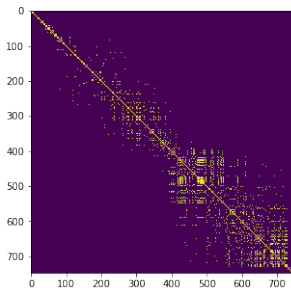
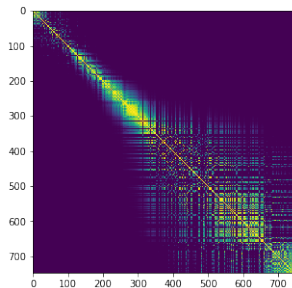
Agrupamiento espectral

Diferentes generaciones del grafo



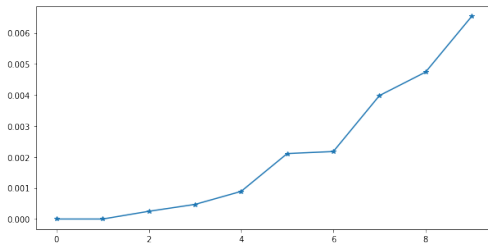
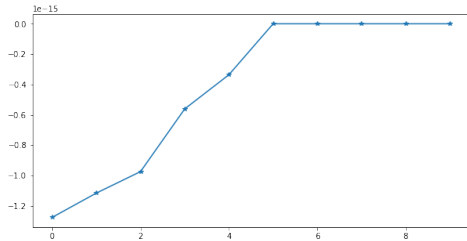
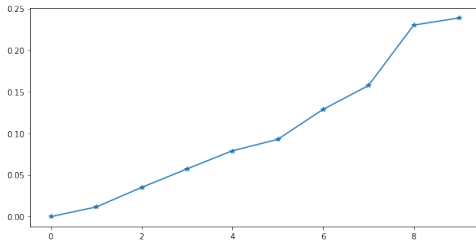
Agrupamiento espectral

Diferentes generaciones del grafo



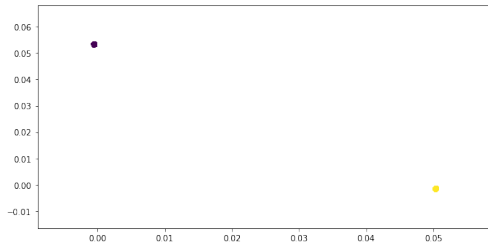
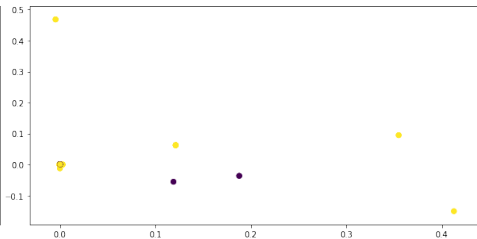
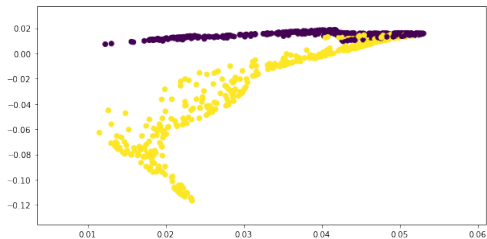
Agupamiento espectral

Diferentes generaciones del grafo



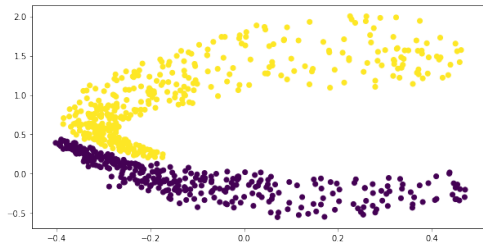
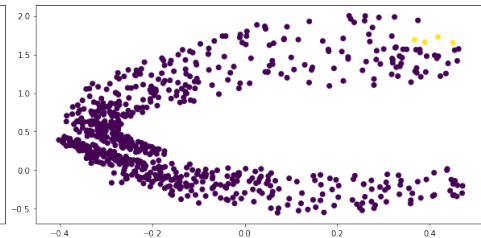
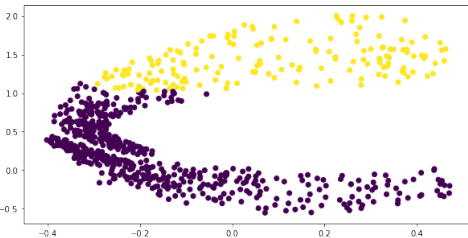
Agrupamiento espectral

Diferentes generaciones del grafo



Agrupamiento espectral

Diferentes generaciones del grafo



Agrupamiento espectral

Puntos básicos

1. Obtener un grafo y su matriz de adyacencias
2. Obtener una representación alternativa de los datos
3. Aplicar un algoritmo de agrupamiento estándar (K -means)

Agrupamiento espectral

Puntos básicos

1. Obtener un grafo y su matriz de adyacencias
2. Obtener una representación alternativa de los datos
3. **Aplicar un algoritmo de agrupamiento estándar (K -means)**

Agrupamiento espectral

Ventajas

- ▶ Sólida base matemática
- ▶ Funciona con clústeres de diversa forma
- ▶ Diferentes criterios y maneras de crear el grafo de similitudes
- ▶ Puede funcionar con diferentes medidas de distancia
- ▶ Se pueden usar diferentes algoritmos sobre la matriz transformada

Agrupamiento espectral

Problemas

Obtener datos transformados

- ▶ Elegir el tipo de grafo (y K en KNN, o ϵ en umbral)
- ▶ Elegir el tipo de matriz Laplaciana
- ▶ Elegir el número de vectores propios

Aggrupamiento espectral

Problemas

Obtener datos transformados

- ▶ Elegir el tipo de grafo (y K en KNN, o ϵ en umbral)
- ▶ Elegir el tipo de matriz Laplaciana
- ▶ Elegir el número de vectores propios
 - ▶ En la práctica, el número de clústeres
 - ▶ Salto máximo entre dos valores propios consecutivos

