



UC | Chile



UC | Chile

Algebra Lineal Aplicada para Ciencia de Datos



Clase 16. Clusterización espectral



- 1 Motivación
- 2 El Laplaciano del grafo
- 3 Transformación espectral
- 4 Clusterización espectral
- 5 Variaciones



UC | Chile

Motivación

Motivación



Suponemos que tenemos datos representados por vectores $y_1, \dots, y_n \in \mathbb{R}^m$ y por la matriz Y de $m \times n$

$$Y = \begin{bmatrix} y_1, \dots, y_n \end{bmatrix}$$

Una forma de agrupar estos vectores es usar ***k*-means**

Si $\bar{y}_1, \dots, \bar{y}_c$ son los **centroides** entonces el **criterio de membresía** es la **distancia Euclidea**

En particular, la clase de y_i es

$$\text{clase}(y_i) = \arg \min \{ \|y_i - \bar{y}_j\|_2 : j \in \{1, \dots, c\} \}.$$

Motivación



Sin embargo, en algunos casos se tiene un **criterio de similitud** entre los datos

En otras palabras, se tiene una matriz S de $n \times n$ donde

$S_{i,j} \geq 0$ y $S_{i,j}$ cuantifica qué tal similar es y_i a y_j

Por ejemplo,

$$S_{i,j} = \exp \left(-\frac{\|y_i - y_j\|_2}{r_s} \right)$$

donde $r_s > 0$ es un factor de escala

Motivación



Podemos representar esta información usando un **grafo no dirigido**

Asociamos a y_1, \dots, y_n el grafo $\mathcal{G} = (V, E)$ donde

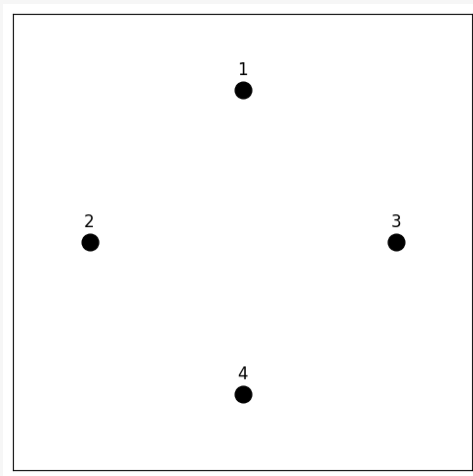
$$V = \{1, \dots, n\} \quad \text{y} \quad E = \{\{i, j\} : i, j \in V, i \neq j\}$$

Además, asociamos a cada **arista** los pesos

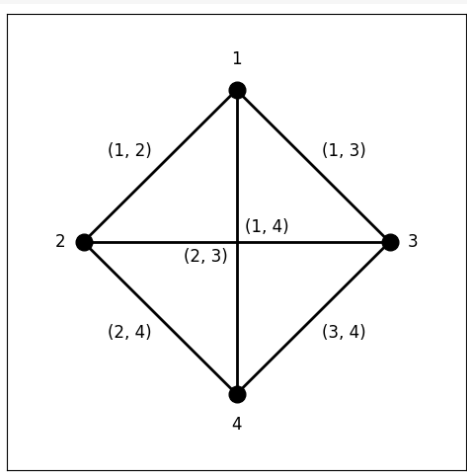
$$\{i, j\} \in E : w(\{i, j\}) = S_{i,j}$$

Llamamos $W \in \mathbb{R}^{n \times n}$ a la matriz de pesos

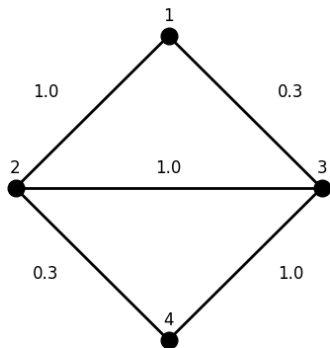
Motivación



Motivación



Motivación



$$W = \begin{bmatrix} 0 & 1 & 0.3 & 0 \\ 1 & 0 & 1 & 0.3 \\ 0.3 & 1 & 0 & 1 \\ 0 & 0.3 & 1 & 0 \end{bmatrix}$$

Por convención, asignar a una arista un **peso cero** es equivalente a **eliminar** la arista

Motivación



En estos casos, nos gustaría usar la información que nos entrega la **matriz de similitud** al momento de agrupar y_1, \dots, y_n

¿Cómo podemos hacer esto?

Motivación



Una forma popular de hacerlo es construir una transformación de los datos

$$y_1, \dots, y_n \xrightarrow{T} z_1, \dots, z_n$$

donde $z_1, \dots, z_n \in \mathbb{R}^r$

Si representamos estos datos como

$$Z = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}$$

entonces escribimos $Z = T(Y)$

¡La transformación $T : \mathbb{R}^m \rightarrow \mathbb{R}^r$ no tiene por qué ser lineal!

Motivación



Los **métodos espectrales** construyen esta transformación calculando la **descomposición de valores propios** de una matriz asociada al grafo

En esta clase, estudiaremos los métodos basados en el **Laplaciano del grafo**



UC | Chile

El Laplaciano del grafo

El Laplaciano del grafo



Vamos a asociar al grafo una matriz llamada **Laplaciano del grafo**

Motivamos esta construcción usando la **difusión del calor**

El Laplaciano del grafo



Suponemos que en un instante 0 cada vértice i está a una **temperatura** $f_i = f_i(0)$

La temperatura del vértice i evoluciona en el tiempo, de modo que $f_i = f_i(t)$

El peso $w(\{i, j\})$ corresponde a la **conductividad térmica** entre el vértice i y el nodo j

El Laplaciano del grafo



La **tasa de cambio** de la temperatura depende de **los gradientes** de temperatura del nodo i en relación a sus vecinos, de modo que

$$\begin{aligned}\dot{f}_i(t) &= \sum_{j=1}^n w(\{i, j\})(f_j(t) - f_i(t)) \\ &= \sum_{j=1}^n W_{i,j}(f_j(t) - f_i(t)) \\ &= -\left(\sum_{j=1}^n W_{i,j}\right) f_i(t) + \sum_{j=1}^n W_{i,j} f_j(t)\end{aligned}$$

El Laplaciano del grafo



Definimos el **grado** del vértice i como

$$d_i = \sum_{j=1}^n W_{i,j}$$

y la matriz diagonal $D \in \mathbb{R}^{n \times n}$ donde $D_{i,i} = d_i$ de modo que

$$\dot{f}_i(t) = -d_i f_i(t) + \sum_{j=1}^n W_{i,j} f_j(t) = - \sum_{j=1}^n (D_{i,j} - W_{i,j}) f_j(t)$$

El Laplaciano del grafo



Definimos el **Laplaciano (no normalizado) del grafo** como

$$L = D - W$$

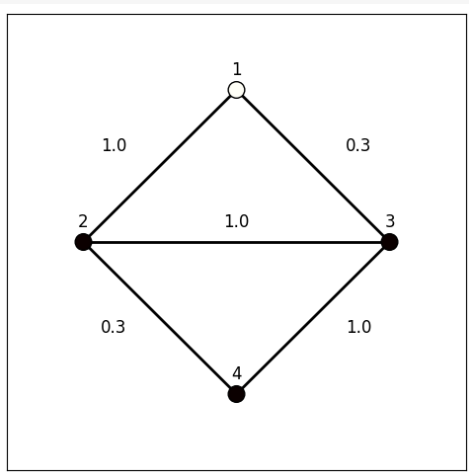
de modo que si

$$f(t) = \begin{bmatrix} f_1(t) \\ \dots \\ f_n(t) \end{bmatrix} \quad \text{y} \quad \dot{f}(t) = \begin{bmatrix} \dot{f}_1(t) \\ \dots \\ \dot{f}_n(t) \end{bmatrix}$$

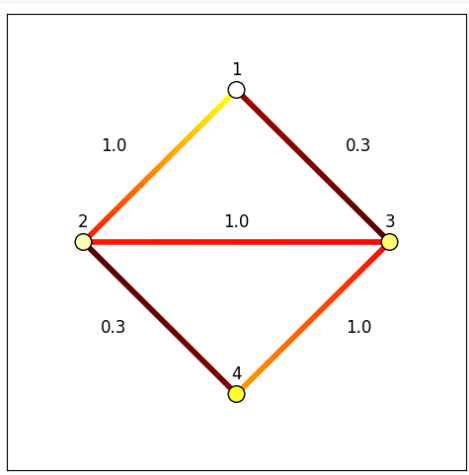
entonces

$$\dot{f}(t) = -Lf(t)$$

El Laplaciano del grafo



El Laplaciano del grafo



El Laplaciano del grafo



La **dinámica** del sistema es consistente con nuestra intuición de **similaridad**

La temperatura cambia rápidamente entre vértices que tienen **mayor** similaridad

Ya que la dinámica del sistema está caracterizada **completamente** por la matriz L
vamos a estudiar la descomposición espectral de esta matriz para definir la
transformación T



UC | Chile

Transformación espectral

Transformación espectral



$$L = D - W$$

Verificamos que el Laplaciano es **simétrico**

Además, se puede verificar que es **semidefinido positivo**

Transformación espectral



Escribimos

$$L = Q\Lambda Q^\top$$

donde

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad \text{con} \quad 0 \leq \lambda_1 \leq \dots \leq \lambda_n$$

¡Observe el cambio en la convención!

¿Qué información acerca del grafo contiene esta descomposición?

Transformación espectral



Los **vectores propios** del Laplaciano representan la **conectividad del grafo**

Por ejemplo, se puede demostrar que **siempre se tiene que** $\lambda_1 = 0$

Además, habrán tantos vectores propios linealmente independientes asociados a λ_1 como **componentes conexas** tenga el grafo

Transformación espectral



Si $V_1, \dots, V_c \subset V$ son las componentes conexas del grafo, entonces existen vectores propios q_1, \dots, q_c asociados al valor propio 0 y tales que

$$q_{i,j} = \begin{cases} 1 & j \in V_i \\ 0 & \text{en otro caso} \end{cases}$$

Transformación espectral



Los vectores propios asociados a los valores propios de mayor magnitud capturan distintas características acerca de la **conectividad del grafo**

Por lo tanto, dado $r \in \{1, \dots, n\}$ definimos la **transformación espectral**

$$T(Y) = \begin{bmatrix} q_1^\top \\ \vdots \\ q_r^\top \end{bmatrix}.$$



UC | Chile

Clusterización espectral

Clusterización espectral



Una vez calculada la transformación espectral, definimos

$$z_i = T(y_i) \quad \text{para } i \in \{1, \dots, n\}$$

Es claro que

$$z_1, \dots, z_n \in \mathbb{R}^r$$

Clusterización espectral



Podemos usar k -means en los datos $z_1, \dots, z_n \in \mathbb{R}^r$

Si los centroides son $\bar{z}_1, \dots, \bar{z}_n$ entonces la clase de y_i es

$$\mathbf{clase}(y_i) = \arg \min \{ \|z_i - \bar{z}_j\|_2 : j \in \{1, \dots, c\} \}.$$

Este método se conoce como **clusterización espectral**



UC | Chile

Variaciones

Variaciones



El Laplaciano

$$L = D - W$$

se conoce como **Laplaciano no normalizado**

Variaciones



Existen dos variaciones: el **Laplaciano normalizado (simétrico)**

$$L_{\text{sim}} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2}$$

y el **Laplaciano normalizado (asimétrico)**

$$L_{\text{asim}} = D^{-1} L = I - D^{-1} W$$

Si bien conceptualmente son equivalentes, hay algunas variaciones en su interpretación y en el escalamiento de los vectores propios que puede tener un impacto en la práctica



UC | Chile