

Análisis de Redes con Autovectores y Autovalores

Métodos Numéricos

2do C 2022

Repaso autovalores y autovectores

Repaso autovalores y autovectores

Definición

Sea $A \in \mathbb{R}^{n \times n}$. Un *autovector* de A es un vector no nulo tal que $Ax = \lambda x$, para algún escalar λ . Un escalar λ es denominado *autovalor* de A si existe una solución no trivial x del sistema $Ax = \lambda x$. En este caso, x es llamado *autovector asociado a λ* .

Consideramos:

$$A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}, u = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Au = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, Av = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2v$$

Gráficamente.... A sólo estira (o encoge) el vector v .

Repaso autovalores y autovectores

Dada una matriz $A \in \mathbb{R}^{n \times m}$, un vector $\mathbf{x} \in \mathbb{R}^m$ no nulo puede ser paralelo a $A\mathbf{x}$, lo que significa que existe una constante λ tal que $A\mathbf{x} = \lambda\mathbf{x}$.

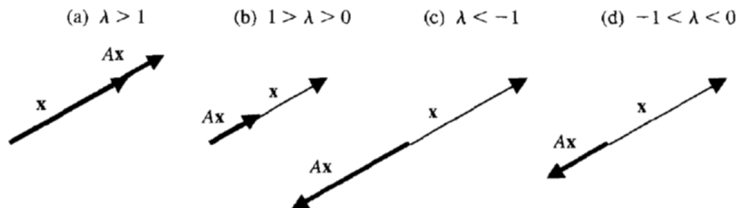
A todos los \mathbf{x} que cumplen esta propiedad se los denomina autovectores de A y λ su autovalor asociado.

Repaso autovalores y autovectores

Dada una matriz $A \in \mathbb{R}^{n \times m}$, un vector $\mathbf{x} \in \mathbb{R}^m$ no nulo puede ser paralelo a $A\mathbf{x}$, lo que significa que existe una constante λ tal que $A\mathbf{x} = \lambda\mathbf{x}$.

A todos los \mathbf{x} que cumplen esta propiedad se los denomina autovectores de A y λ su autovalor asociado.

Representación gráfica:



En muchos casos, la presencia de autovectores-autovalores puede ser utilizada para encontrar una factorización $A = PDP^{-1}$, donde D es una matriz diagonal.

Intuición

Podemos encontrar una base donde la transformación lineal A se comporta como si fuese diagonal. Es decir en esa base, las transformaciones solo expanden o contraen cada dimensión por separado sin afectar a otras.

Observación

No toda matriz $A \in \mathbb{R}^{n \times n}$ es diagonalizable.

Teorema

Una matriz $A \in \mathbb{R}^{n \times n}$ es diagonalizable sí y solo sí A tiene n autovectores linealmente independientes (las columnas de P).

Teorema Espectral

Si $A \in \mathbb{R}^{n \times n}$ es simétrica, entonces existe una base ortonormal de autovectores $\{v_1, \dots, v_n\}$ asociados a $\lambda_1, \dots, \lambda_n$.

Consecuencia: Existe P , y $P^{-1} = P^t$. Luego, $A = PDP^t$.

Redes

Redes

Matriz de adyacencia

Una red G puede representarse como una matriz A , donde se cumple que los elementos toman el valor 1 si el nodo i está conectado con el j ($A_{ij} = 1$), y cero en caso contrario.

Matriz Laplaciana

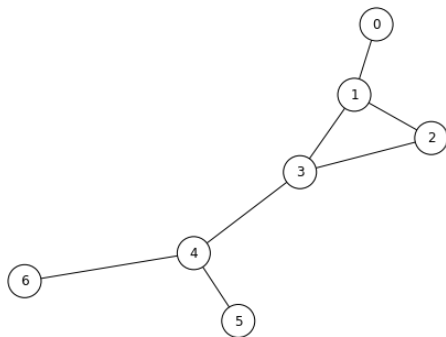
A partir de la matriz de adyacencia A se puede computar la denominada matriz laplaciana L de una red: $L = D - A$

Donde D es una matriz diagonal con valores para cada fila igual a la suma de esa fila. El resultado final tiene la propiedad de que cada fila de la laplaciana suma 0.

Red de ejemplo

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$



Autovalores y Autovectores de una Red

Nos interesa entender qué propiedades de una red se pueden extraer analizando los autovalores y autovectores de las matrices de adyacencia y laplaciana asociadas a la red.

Primero ganemos intuición en la operación de multiplicación de matriz-vector.

Multiplicación Matriz-Vector

Multiplicación Matriz-Vector

La multiplicación de un vector x por una matriz cuadrada A (Ax) se puede pensar como una transformación lineal que genera un nuevo vector con una nueva magnitud y una nueva posición. Tomemos como ejemplo un caso en \mathbb{R}^2

$$A = \begin{bmatrix} 0,5 & 0,8 \\ -0,3 & 0,2 \end{bmatrix}$$

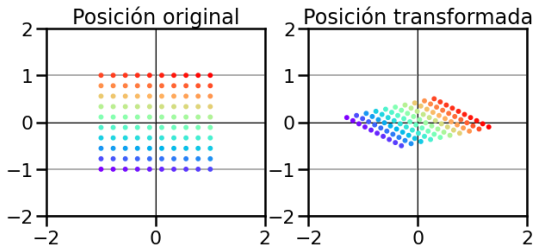


Figura: Puntos en \mathbb{R} transformados al multiplicarlos por una matriz

Multiplicación Matriz-Vector

Si tomamos el resultado y lo volvemos a transformar estamos repitiendo operación y es equivalente a $A^n x$, donde n es la cantidad de veces que se realiza la transformación.

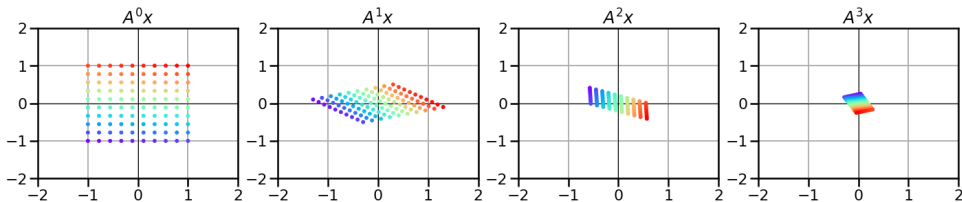


Figura: Puntos en \mathbb{R} transformados al multiplicarlos por una matriz sucesivas veces

Multiplicación Matriz-Vector

Ahora tomemos la matriz de adyacencia de la red de ejemplo, y multipliquemos varias veces por el vector columna

$$x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

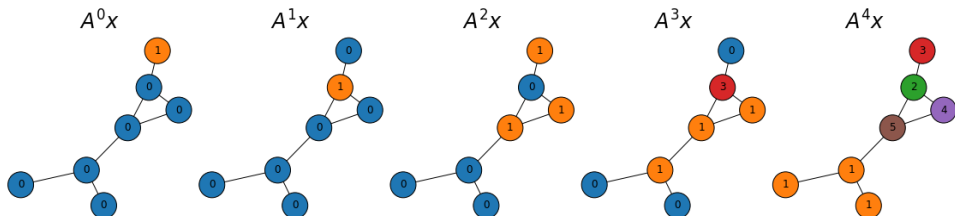


Figura: Transformación de un vector multiplicado por la matriz de adyacencia sucesivas veces

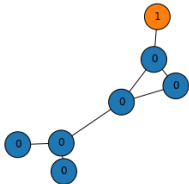
Podemos interpretar los valores como la cantidad de caminos de largo n que empiezan en el nodo inicial y terminan en el nodo que muestra ese valor.

Multiplicación Matriz-Vector

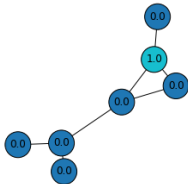
Normalicemos la matriz de adyacencia para que sus columnas sumen 1

$$\begin{bmatrix} 0. & 0,33 & 0. & 0. & 0. & 0. & 0. \\ 1. & 0. & 0,5 & 0,33 & 0. & 0. & 0. \\ 0. & 0,33 & 0. & 0,33 & 0. & 0. & 0. \\ 0. & 0,33 & 0,5 & 0. & 0,33 & 0. & 0. \\ 0. & 0. & 0. & 0,33 & 0. & 1. & 1. \\ 0. & 0. & 0. & 0. & 0,33 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0,33 & 0. & 0. \end{bmatrix}$$

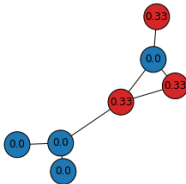
A^0x



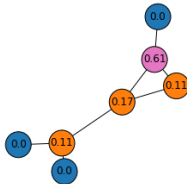
A^1x



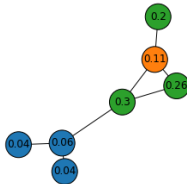
A^2x



A^3x



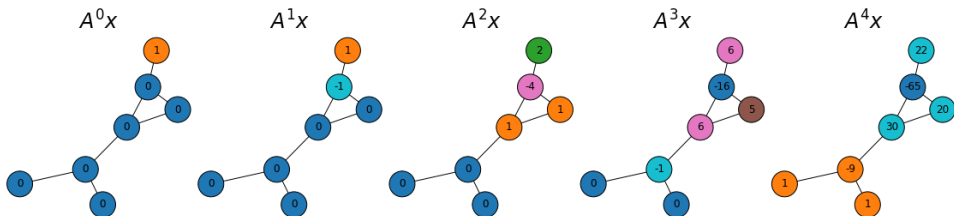
A^4x



Los valores siempre suman 1. En cada paso se distribuye el total entre las conexiones. Para muchas iteraciones, podemos interpretar los valores como la probabilidad de pasar por ese nodo.

Multiplicación Matriz-Vector

Para la matriz laplaciana obtenemos



En este caso los valores siempre suman 0.

Multiplicación Matriz-Vector

Las operaciones matriz vector para las matrices asociadas a una red generan resultados que permiten analizar la estructura de la red.

Los autovectores de las redes serán aquellos cuyos resultados luego de multiplicar por la matriz solo se vean modificados por un escalar.

Aplicaciones Autovectores y Autovalores de Redes

Aplicación 1: centralidad

Centralidad

Las medidas de centralidad buscan ordenar los nodos en base a cual es más “central”, “influyente” o “importante”.

- ▶ Centralidad de grado: la cantidad de conexiones o “grado” de un nodo

$$g_i = \sum_{j=1}^n A_{ij}$$

- ▶ Centralidad de autovector: pesando las conexiones según la importancia de cada vecino

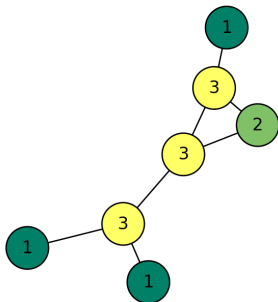
$$x_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} x_j$$
$$x\lambda = Ax$$

- ▶ Page Rank (sin saltos aleatorios y en redes no dirigidas): pesando las conexiones según la importancia de cada vecino normalizado por el grado.

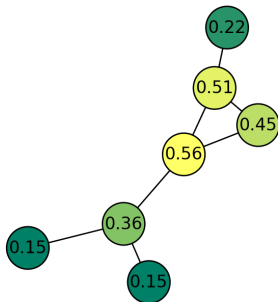
$$x_i = \sum_{j=1}^n (A_{ij}/g_j) x_j$$
$$x = Rx$$

Centralidad

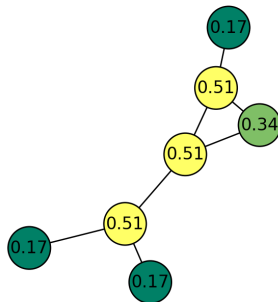
Grado



Autovector



Page Rank

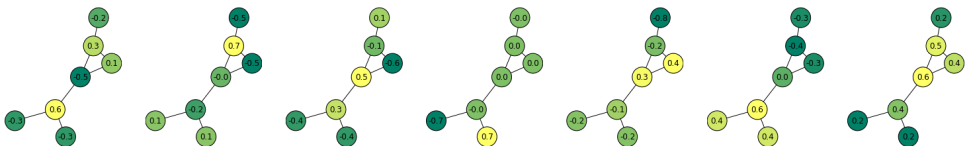


Para la centralidad de autovector, el valor indicado en cada nodo es el valor x_i . Se utiliza el autovector con mayor autovalor.

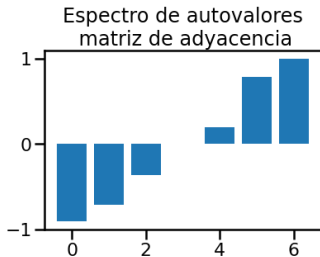
Page Rank sin saltos aleatorios y para una matriz simétrica se comporta igual que el grado, es decir el vector de grados es solución de la ecuación.

Centralidad

Todos los autovectores de la matriz adyacencia



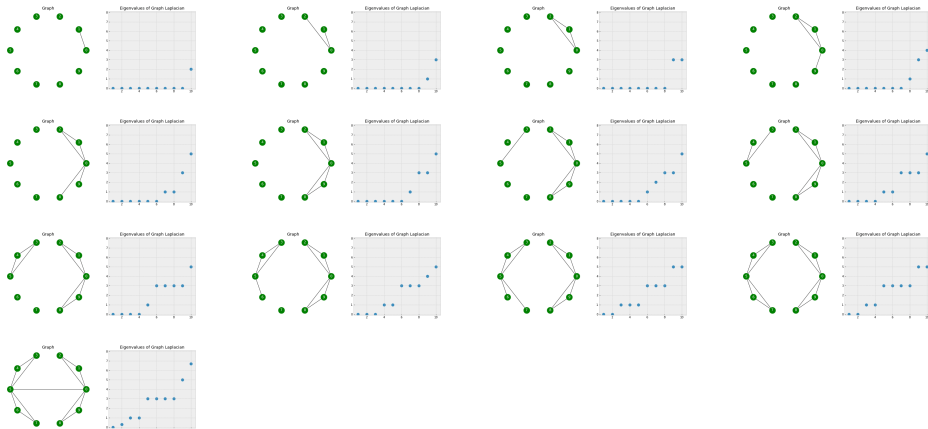
La matriz es simétrica por lo tanto, forman una base y son ortogonales entre si



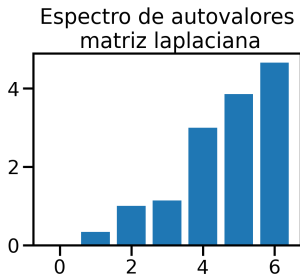
Aplicación 2: Cortar una red y Conectividad Algebraica

Autovalores de Matriz Laplaciana

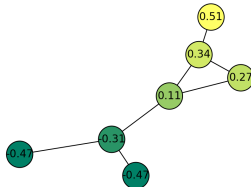
El espectro de la matriz laplaciana también tiene información sobre la estructura de la red



Cortar la red con un Autovector



Autovector con autovalor 0.34



Si se analiza el autovector correspondiente al menor autovalor mayor de cero, se puede utilizar el cambio de signo de los valores para cortar la red en dos partes. Este método aproxima a una solución para el problema de cortar una red en dos partes con misma cantidad de nodos cortando la mínima cantidad de conexiones.

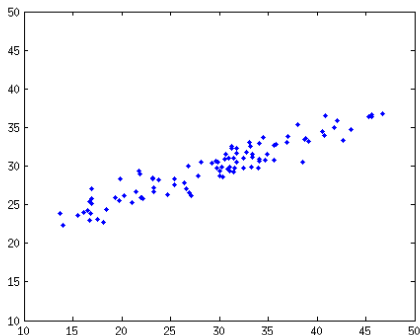
Aplicación 3: PCA

Análisis de Componentes Principales

Ejemplo datos en \mathbb{R}^2

Sean $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ una secuencia de n datos, con $x^{(i)} \in \mathbb{R}^2$.

$$X = \begin{bmatrix} x^{(1)t} \\ x^{(2)t} \\ x^{(3)t} \\ x^{(4)t} \\ x^{(5)t} \\ x^{(6)t} \\ \vdots \\ x^{(n)t} \end{bmatrix} = \begin{bmatrix} 26,4320 & 27,7740 \\ 26,8846 & 26,5631 \\ 23,3309 & 26,6983 \\ 30,6387 & 31,5619 \\ 30,5171 & 30,8993 \\ 45,6364 & 36,6035 \\ \vdots & \vdots \\ 16,0650 & 24,0210 \end{bmatrix}$$



Análisis de Componentes Principales

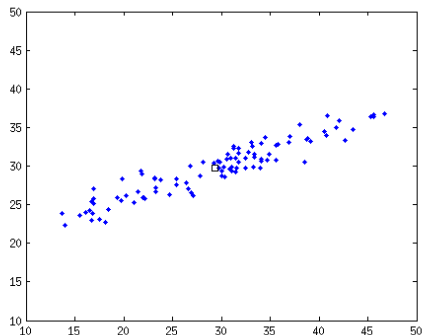
Ejemplo datos en \mathbb{R}^2

$$X = \begin{bmatrix} 26,4320 & 27,7740 \\ 26,8846 & 26,5631 \\ 23,3309 & 26,6983 \\ 30,6387 & 31,5619 \\ 30,5171 & 30,8993 \\ 45,6364 & 36,6035 \\ \vdots & \vdots \\ 16,0650 & 24,0210 \end{bmatrix}$$

Media:

$$\mu = \frac{1}{n}(x^{(1)} + \dots + x^{(n)})$$

$$\mu = (29,3623, 29,7148)$$



Varianza de una variable x_k : Medida para la dispersión de los datos.

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_k^{(i)} - \mu_k)^2$$

$$\sigma_{x_1}^2 = 66,2134, \sigma_{x_2}^2 = 12,5491$$

Análisis de Componentes Principales

Ejemplo datos en \mathbb{R}^2 - Covarianza

Covarianza: Medida de cuánto dos variables varían de forma similar.

Variables con mayor covarianza inducen la presencia de cierta dependencia o relación.

Para dos variables

$$\sigma_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x^{(i)} - \mu_x)(y^{(i)} - \mu_y)$$

Para dos componentes de una variable

$$\sigma_{x_j x_k} = \frac{1}{n-1} \sum_{i=1}^n (x_j^{(i)} - \mu_j)(x_k^{(i)} - \mu_k)$$

Forma matricial (las medias ya deben estar restadas sobre los datos)

$$M_X = \frac{X^t X}{n-1}$$

Análisis de Componentes Principales

Ejemplo datos en \mathbb{R}^2 - Covarianza

Dadas n observaciones de dos variables x_1 , x_2 , y $v = (1, \dots, 1)^t$:

$$\sigma_{x_1 x_2} = \frac{1}{n-1} \sum_{i=1}^n (x_1^{(i)} - \mu_1)(x_2^{(i)} - \mu_2) = \frac{1}{n-1} (x_2 - \mu_2 v)^t (x_1 - \mu_1 v)$$

Matriz de Covarianza:

$$X = \begin{bmatrix} 26,4320 - \mu_1 & 27,7740 - \mu_2 \\ 26,8846 - \mu_1 & 26,5631 - \mu_2 \\ 23,3309 - \mu_1 & 26,6983 - \mu_2 \\ 30,6387 - \mu_1 & 31,5619 - \mu_2 \\ 30,5171 - \mu_1 & 30,8993 - \mu_2 \\ 45,6364 - \mu_1 & 36,6035 - \mu_2 \\ \vdots & \vdots \\ 16,0650 - \mu_1 & 24,0210 - \mu_2 \end{bmatrix}$$

$$M_X = \frac{X^t X}{n-1} = \begin{bmatrix} \sigma_{x_1 x_1} & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma_{x_2 x_2} \end{bmatrix} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 \end{bmatrix}$$

$$M_X = \begin{bmatrix} 66,2134 & 27,1263 \\ 27,1263 & 12,5491 \end{bmatrix}$$

¿Cómo expresar mejor nuestros datos?

Objetivo

Buscamos una transformación de los datos que disminuya la redundancia (es decir, disminuir la covarianza).

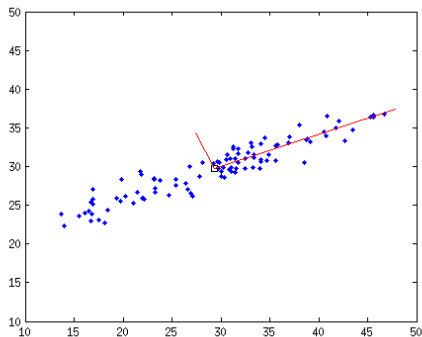
Alternativa y equivalentemente: encontrar direcciones ortogonales que maximicen la varianza de nuestros datos

- ▶ Cambio de base: $\hat{X}^t = PX^t$.
- ▶ Cómo podemos hacerlo? Diagonalizar la matriz de covarianza. Esta matriz tiene la varianza de cada variable en la diagonal, y la covarianza en las restantes posiciones. Luego, al diagonalizar buscamos variables que tengan covarianza cero entre sí y la mayor varianza posible.
- ▶ La matriz de covarianza $M_x = \frac{1}{n-1}X^tX$ es simétrica y semidefinida positiva.
- ▶ Si quieren ver una demo, *Pattern Recognition and Machine Learning* de Christopher Bishop

¿Cómo expresar mejor nuestros datos?

Volvemos al ejemplo

$$\begin{aligned} M_X &= \begin{bmatrix} 66,2134 & 27,1263 \\ 27,1263 & 12,5491 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 0,9228 & -0,3852 \\ 0,3852 & 0,9228 \end{bmatrix}}_V \underbrace{\begin{bmatrix} 77,5362 & 0 \\ 0 & 1,2263 \end{bmatrix}}_{D=M_{\hat{X}}} \underbrace{\begin{bmatrix} 0,9228 & 0,3852 \\ -0,3852 & 0,9228 \end{bmatrix}}_{V^t} \end{aligned}$$



Análisis de Componentes Principales

Resumen hasta acá

- ▶ Tenemos n muestras de m variables.
- ▶ Calculamos el vector μ que contiene la media de cada de una las variables.
- ▶ Construimos la matriz $X \in \mathbb{R}^{n \times m}$ donde cada muestra corresponde a una fila de X y tienen media cero (i.e., $X_i := (x^{(i)} - \mu)$).
- ▶ Diagonalizamos la matriz de covarianzas $M_X = \frac{X^t X}{n-1}$. La matriz V (ortogonal) contiene los autovectores de M_X .

Análisis de Componentes Principales

Propiedades del cambio de base

- ▶ Disminuye redundancias.
- ▶ El cambio de base $\hat{X}^t = PX^t = V^tX^t$ asigna a cada muestra un nuevo *nombre* mediante un cambio de coordenadas.
- ▶ Las columnas de V (autovectores de M_X) son las componentes principales de los datos.
- ▶ En caso de m grande, es posible tomar sólo un subconjunto de las componentes principales para estudiar (i.e., aquellas que capturen mayor proporción de la varianza de los datos).

Matriz de Similaridad

Matriz de similaridad

La matriz $M_x = \frac{1}{n-1}X^tX$ tiene dimensiones $m \times m$.

En cambio la matriz $C = XX^t$ tiene dimensiones $n \times n$ y cada punto C_{ij} computa el producto interno entre el dato $x^{(i)}$ y el dato $x^{(j)}$.

Matriz de similaridad

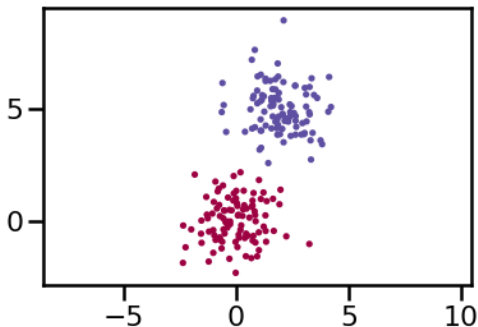


Figura: Dos cúmulos de puntos en dos dimensiones

Vamos a calcular la matriz de similaridad para estos puntos

Matriz de similaridad

Matriz de similaridad producto interno $C = XX^t$ para los datos anteriores.

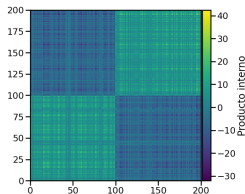


Figura: Matriz de similaridad Producto Interno

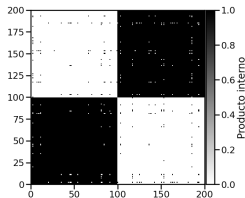
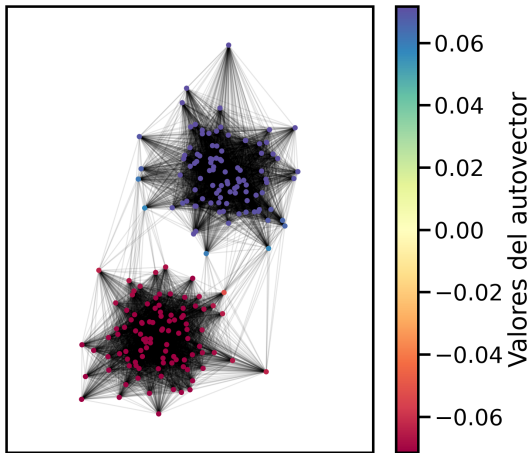


Figura: Binarización: podemos construir una matriz de adyacencia conectando los datos que tienen similaridad mayor que 0

Matriz de similaridad

Construimos una red con la matriz de adyacencia anterior.



Los colores son los valores que toma el autovector de la matriz laplaciana de la red con menor autovalor mayor que cero. Sirve para cortar el grafo e identificar los grupos