

REDUCCIÓN DE DIMENSIONES

LA REDUCCIÓN DE DIMENSIONES ES UNA HERRAMIENTA MUY UTILIZADA A LA HORA DE USAR ALGORITMOS INTELIGENTES PARA PROCESAR DATOS DE N DIMENSIONES. SE ENCARGA DE BUSCAR UNA REPRESENTACIÓN MÁS COMPACTA DE UN SET DE DATOS, NO SOLO PARA OPTIMIZAR EL USO DE RECURSOS SINO TAMBIÉN PARA OTROS USOS COMO VISUALIZACIÓN DE DATOS, COMPRESIÓN DE IMAGENES Y FEATURE ENGINEERING.

FEATURE ENGINEERING

LOS ALGORITMOS DE REDUCCIÓN DE DIMENSIONES SE USAN PARA REDUCIR EL RUIDO (COSAS QUE MOLESTAN) FILTRANDO LOS FEATURES (COLUMNAS) RUIDOSOS Y DEVOLVIENDO UN NUEVO SET DE FEATURES CON MENOR CANTIDAD DE RUIDO.

A VECES TAMBIÉN ES POSIBLE DEDUCIR LA DIM DEL SET DE DATOS Y LUEGO AGREGAR ESAS COLUMNAS AL SET ORIGINAL. ESTE AUMENTO AGREGA FEATURES EN DONDE LA SEÑAL ES MUY FUERTE EN RELACIÓN AL RUIDO Y ES POSIBLE QUE EL ALGORITMO FUNCIONE MEJOR ASÍ.

COMPRESIÓN DE IMAGENES

LAS IMAGENES SE PUEDEN PENSAR COMO DATOS EN UN ESPACIO DE MUCHAS DIM. ESTA DIM SE PUEDE REDUCIR UTILIZANDO ALGUN ALGORITMO TAL COMO **SVD O PCA**. DE ESTA FORMA SE PUEDE AHORRAR LOS BYTES OCUPADOS POR UNA IMAGEN A COSTA'S DE REDUCIR SU CALIDAD DADO QUE PERDEMOS CIERTA INFORMACIÓN.

VISUALIZACIÓN DE DATOS

LA VISUALIZACIÓN DE DATOS NOS AYUDA A ENTENDER UN SET DE DATOS. VISUALIZAR DATOS EN 3 DIM ES DIFÍCIL PARA EL CEREBRO HUMANO PERO EXISTEN CIERTAS TECNICAS DE VISUALIZACION QUE LO PERMITEN.

LOS ALGORITMOS PARA MANIPULAR LA DIMENSIONALIDAD DE LOS DATOS SE DIVIDEN EN:

- METODOS LINEALES → SVD y PCA
- METODOS NO-LINEALES → MANIFOLD LEARNING

SVD

LA SVD ES UNA HERRAMIENTA IMPORTANTE PARA LA REDUCCIÓN DE DIMENSIONES.

SEA A UNA MATRIZ $N \times M$ SE PUEDE DEMOSTRAR

$$A = U \Sigma V^T$$

DONDE U ES UNA MATRIZ UNITARIA EN $N \times N$, V ES OTRA MATRIZ UNITARIA EN $M \times M$ Y Σ ES UNA MATRIZ DIAGONAL EN $N \times M$.

¿COMO CALCULAR LA SVD?

① CALCULAR LOS AUTOVALORES

Y AUTOVECTORES DE $A^T A$. $A = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$$

PARA SACAR LOS AUTOVALORES (λ)

$$\det(A - \lambda I_D) = 0$$

$$\det \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} - \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \det \begin{bmatrix} 9-\lambda_1 & 0 \\ 0 & 4-\lambda_2 \end{bmatrix} = 0$$

$$\det(A - \lambda I_D) = 0 \quad (9 - \lambda_1) \cdot (4 - \lambda_2) = 0$$

$$\lambda_1 = 9 \quad \lambda_2 = 4$$

UNA VEZ QUE TENGO LOS AUTOVALORES, BUSCO LOS AUTOVECTORES

$$AX = \lambda X$$

JUSTO EN ESTE PASO LA MATRIZ ES NULA ENTONCES LOS AUTOVECTORES ASOCIADOS SON:

$$(A - \lambda I_D)X = 0$$

$$V_1 = (1, 0) \quad V_2 = (0, 1)$$

EL CANONICO DE \mathbb{R}^2 .

② LA MATRIZ Σ SE COMPONE CON LOS VALORES SINGULARES ($\sigma = \sqrt{\lambda_i}$) EN LA DIAGONAL DE MAYOR A MENOR

$$\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

③ LA MATRIZ U SE COMPONE CON LOS AUTOVECTORES NORMALIZADOS COMO COLUMNAS

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$v_1 \ v_2$

④ LA MATRIZ V SE PUEDE OBTENER DESOLVIENDO EL SISTEMA $AV = U\Sigma$ SABIENDO QUE VTV ES ID.

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} V = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

PLANTEO
A V COMO
INCÓGNITA

$$\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$3x_1 = 3 \quad -2x_4 = 2$$

$$3x_2 = 0$$

$$-2x_3 = 0$$

$$V = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

COMO DIJIMOS ANTES U Y V SON MATRICES UNITARIAS. ESTO ES ($U^T U = I_D$) Y QUE LAS COLUMNAS DE U FORMAN UNA BASE ORTHONORMAL (VECTORES TIENEN NORMA = 1). TODO ESTO TMB VALE PARA V.

SI A UNA SVD LA MULTIMPLICAMOS A AMBOS LADOS POR V OBTENEMOS $AV = U\Sigma$, ESTO QUIERE DECIR QUE SI A UNA BON LE APlico UNA TRANSFORMACIÓN A \Rightarrow EL RESULTADO ES UNA SEGUNDA BON DONDE LOS VECTORES SON MULTIPLICADOS POR UN ESCALAR.

SVD REDUCIDA

EN DETERMINADAS OCASIONES LOS VALORES DE Σ Y U PUEDEN SER CERO. ENTONCES PUEDES TOMAR DE Σ SOLO AQUELLOS VALORES Q SON POSITIVOS. TENIENDO ESTO EN CUENTA, SERIA NECESARIO SOLO UTILIZAR LAS R COLUMNAS DE U Y LAS PRIMERAS R FILAS DE V^T . R COINCIDE CON EL RANGO DE LA MATRIZ

$\hookrightarrow \text{DIM}(\text{COL}(A)) \rightarrow$ SON C.L
 $\downarrow \text{BASE}$ O I.L.

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \rightarrow A = \begin{pmatrix} -0.44 & -0.89 \\ -0.89 & 0.44 \end{pmatrix} \begin{pmatrix} 3.16 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -0.71 & -0.71 \\ -0.71 & 0.71 \end{pmatrix}$$

$U \quad \Sigma \quad V^T$
 $M \times R \quad R \times R \quad R \times M$

ESTO NOS DA A ENTENDER QUE BAJO CIERTOS ESCENARIOS SE PUEDE ENCONTRAR UNA FORMA MAS COMPACTA DE REPRESENTAR LA MATRIZ A Y ESTO DA ORIGEN A LA SVD PARA EL USO DE COMPRESIÓN DE IMAGENES.

APPROXIMACIÓN DE RANGO K

Dicho que a se puede representar una matriz de manera compacta. Se puede pensar que como los valores de Σ se encuentran de mayor a menor entonces representan un grado de importancia de los datos. Esto da a entender que la matriz V es una base en donde los vectores están ordenados por importancia. Entonces se podría ir más allá de quedarse con los R valores no nulos, es decir, podríamos quedarnos con los k valores ($k < R$) y aproximar bastante bien la matriz A .

$$A_3 = \begin{pmatrix} 0.8138 & 0.2143 & -0.5402 \\ 0.2931 & -0.9540 & 0.0632 \\ 0.5018 & 0.2097 & 0.8392 \end{pmatrix} \begin{pmatrix} 8.1884 & 0 & 0 \\ 0 & 2.1568 & 0 \\ 0 & 0 & 1.1395 \end{pmatrix} \begin{pmatrix} 0.5178 & 0.1475 & 0.8427 \\ 0.2601 & 0.2960 & -0.2116 \\ 0.5201 & 0.5919 & -0.4232 \\ 0.6275 & -0.7350 & -0.2569 \end{pmatrix} \quad A_3 = \begin{pmatrix} 3 & 2 & 4 & 4 \\ 1 & 0 & 0 & 3 \\ 3 & 1 & 2 & 2 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 0.8138 & 0.2143 \\ 0.2931 & -0.9540 \\ 0.5018 & 0.2097 \end{pmatrix} \begin{pmatrix} 8.1884 & 0 \\ 0 & 2.1568 \end{pmatrix} \begin{pmatrix} 0.5178 & 0.1475 \\ 0.2601 & 0.2960 \\ 0.5201 & 0.5919 \\ 0.6275 & -0.7350 \end{pmatrix} \quad A_2 = \begin{pmatrix} 3.51 & 1.87 & 3.73 & 3.84 \\ 0.93 & 0.01 & 0.03 & 3.01 \\ 2.19 & 1.20 & 2.40 & 2.24 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} 0.8138 \\ 0.2931 \\ 0.5018 \end{pmatrix} (8.1884) \begin{pmatrix} 0.5178 \\ 0.2601 \\ 0.5201 \\ 0.6275 \end{pmatrix} \quad A_1 = \begin{pmatrix} 3.45 & 1.73 & 3.46 & 4.18 \\ 1.24 & 0.62 & 1.25 & 1.5 \\ 2.12 & 1.06 & 2.13 & 2.57 \end{pmatrix}$$

Se puede demostrar que la SVD da la mejor Aprox de rango k posible a la matriz original. En donde mejor significa que tiene la menor diferencia con la matriz original de acuerdo a la **NORMA DE FROBENIUS**. La SVD minimiza la suma de cuadrados de la diferencia entre la matriz original y su aproximación.

- Si bien se ha visto que para distintos k , la Aprox es mejor o peor, es necesario en la práctica saber qué k utilizar:

$$\sum_{i=1}^N \sigma_i^2 = \text{ENERGÍA}$$

N es el número de autovalores en cuestión

AL REALIZAR UNA MATRIZ DE APROXIMACIÓN K , ($K < N$) SE SE CAPTURA UN CIERTO PORCENTAJE DE LA ENERGÍA:

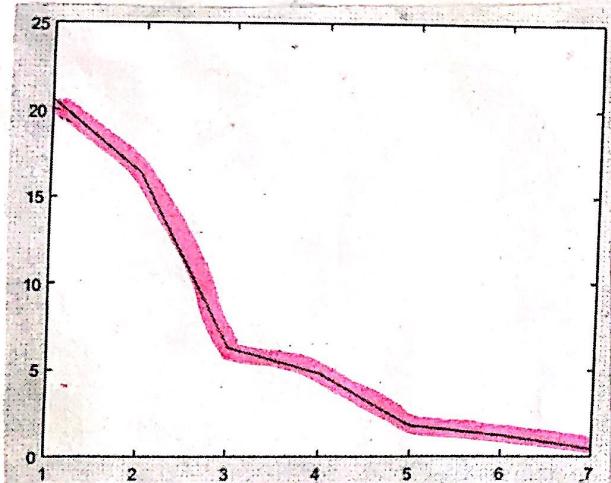
$$\alpha = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^N \sigma_i^2}$$

POR EJEMPLO, PARA LA MATRIZ A:

$$A = \begin{pmatrix} 1 & 9 & 0 & 7 & -2 & 7 & 6 \\ 9 & 6 & 3 & 4 & 1 & 0 & 0 \\ -1 & -1 & -2 & 9 & 0 & 8 & 7 \\ 8 & 2 & 0 & 0 & 0 & -2 & -2 \\ 7 & 6 & 0 & 1 & 0 & 1 & 0 \\ -1 & 0 & 3 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 2 & -1 & 2 & 2 \end{pmatrix}$$

| Cantidad de Valores Singulares | Energía |
|--------------------------------|---------|
| 1 | 55.88 |
| 2 | 91.11 |
| 3 | 96.36 |
| 4 | 99.39 |
| 5 | 99.79 |
| 6 | 99.98 |

LA TABLA MUESTRA COMO VARIA LA ENERGÍA A MEDIDA QUE SE AUMENTA EL VALOR DE K .



CUANDO GRÁFICAMOS LOS AUTOVALORES PODEMOS DETECTAR EN QUÉ PUNTOS SE PRODUCEN 'CODOS' PODEMOS OBSERVAR QUE HAY UN GRAN SALTO A PARTIR DE $K=2$.

COMPRESIÓN DE IMAGENES

PARA VER EL EFECTO QUE TIENE UNA APROXIMACIÓN DE RANGO K A LA HORA DE COMPRESIONAR IMÁGENES CONSIDERA UNA IMAGEN DE 127×350 (44150 PIXELES) DONDE SE REALIZAN APROXIMACIONES CON $K=2$, $K=5$, $K=10$, $K=20$, $K=50$ CUANDO UTILIZAMOS $K=50$, ALMACENAMOS 23850 PIX, ES DECIR UN 53% DE LA IMAGEN ORIGINAL, CON ESTE VALOR ES BASTANTE BUENA LA REPRESENTACIÓN. EN CAMBIO, CON $K=2$ LA IMAGEN ESTÁ MUY DISTORSIONADA. PARA ESTO EXISTE EL ALGORITMO JPG QUE ES MUCHO MEJOR QUE LA SVD PARA IMÁGENES.

OTRAS APLICACIONES DE LA SVD

- REDUCCIÓN DE RUIDO
- EIGENFACES

¿SI TENGO NUEVOS DATOS?

A VECES PUEDE SURGIR LA NECESIDAD DE AGREGAR NUEVAS FILAS A LA MATRIZ A , ES DECIR NUEVOS DATOS EN LA MISMA DIMENSIÓN. PARA ESTO ES NECESARIO OBTENER LA NUEVA SVD. VOLVER A CALCULARLA ES MUY COSTOSO, ENTONCES INTENTAMOS REUTILIZAR LA DESCOMPOSICIÓN QUE TENEMOS Y REALIZAR CERCIOS CÁLCULOS.

PODEMOS PENSAR QUE AGREGAR UNA NUEVA FILA A LA MATRIZ A . ENTONCES SE LA PODRÍA MULTIPLICAR POR V Y LUEGO Σ^{-1} . ASÍ OBTENEMOS SU REPRESENTACIÓN DENTRO DEL NUEVO ESPACIO, ES DECIR, UN NUEVO VECTOR QUE SE AGREGA A LA ÚLTIMA FILA DE U .

A MEDIDA QUE SE VAN AGREGANDO DATOS SE VA PERDIENDO LA PRECISIÓN DE LA SVD ORIGINAL.

PCA

LA IDEA PRINCIPAL DE PCA ES ENCONTRAR LAS 'DIRECCIONES' DE LOS DATOS, ES DECIR, AQUELLAS QUE SOBRE LAS CUALES PODÉMOS PROYECTAR LOS DATOS, RETENIENDO SU VARIABILIDAD.

EN CONCRETO, PCA AJUSTA UNA HIPÉR-ELÍPSIS A LOS DATOS Y DEVUELVE LOS EJES QUE LA COMPONEN, CONOCIDOS COMO **COMPONENTES PRINCIPALES**

LUEGO LOS DATOS SE PODRÁN PROYECTAR SOBRE ESTAS COMPONENTES Y ASÍ REDUCIR LAS DIMENSIONES.

UNA FORMA SENCILLA DE APLICAR ESTE MÉTODO SOBRE A :

- ① PARA CADA ELEMENTO DE LA COL A_i , LE RESTAMOS EL PROMEDIO DE DICHA COLUMNA. ESTO NOS PERMITE CENTRARLOS EN EL ORIGEN.

2

CALCULAR LA MATRIZ DE COVARIANZA

$$\text{COV}(A) = \frac{1}{N-1} A^T A$$

- 3 CALCULAR LOS AUTOVALORES Y AUTOVECTORES DE LA MATRIZ COV(A). LOS AUTOVECTORES CONFORMAN LAS COMPONENTES PRINCIPALES Y LOS AUTOVALORES LA IMPORTANCIA DE CADA COMPONENTE.

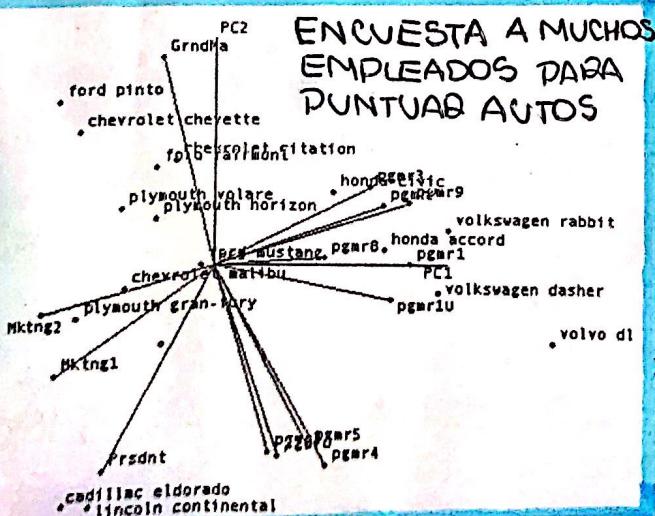
LA MATRIZ DE COVARIANZA ES SIMÉTRICA Y EN CADA CELDA INDICA LA COVARIANZA ENTRE LAS COLUMNAS i, j . CUANDO LA COV = 0 LAS COLS SON INDEPENDIENTES Y CUANDO LA COV = 1 \wedge COV = -1 LAS COLS SON DEPENDIENTES.

→ ES INTERESANTE DEMOSTRAR QUE PARA REDUCIR DIMENSIONES SE PUEDE TOMAR LAS K COMPONENTES PRINCIPALES Y PROYECTAR LOS PUNTOS DE LA MATRIZ SOBRE ESTOS K COMPONENTES REDUCIENDO ASÍ LAS DIMENSIONES.

→ BIPLOTS

CONCLUSIÓN

LA SVD EN GENERAL ES MAS EFICIENTE Y NUMERICAMENTE MAS ESTABLE QUE PCA DADO QUE NO HACE FALTA CONSTRUIR LA MATRIZ DE COVARIANZA Y ES POR ESTO QUE PARA METODOS GENÉRICOS LINEALES SE USA SVD.



LAS DIMENSIONES ORIGINALES COMO VECTORES (ANGULOS Y CERCANIA DE PRESENTAN RELACIÓN)

LAS FILAS (DATOS) SON LOS PUNTOS (AUTOS)
LAS COLUMNAS (VECTORES) SON LOS DISTINTOS EMPLEADOS

MDS

EL ALGORITMO MULTIDIMENSIONAL SCALING BUSCA DESOLVER DADA UNA MATRIZ DE DISTANCIAS ENTRE PUNTOS, QUEREMOS HALLAR LOS PUNTOS EN SI. PARA PODER APLICAR EL ALGORITMO SE DEBE REALIZAR:

- ① ELEVAR LA MATRIZ DE DISTANCIAS AL CUADRADO
- ② CENTRAR LA MATRIZ PARA QUE LAS FILAS Y COLUMNAS TENGAN PROMEDIO CERO
- ③ CALCULAR LA SVD DE LA MATRIZ
- ④ LAS PRIMERAS Q COLUMNAS DE U SON LAS COORDENADAS PARA LOS PUNTOS QUE SE QUIEREN BUSCAR

LAS COORDENADAS QUE DEVUELVE MDS PUEDEN ESTAR ESPEJADAS Y ROTADAS CON RESPECTO A LA REALIDAD YA QUE LAS DISTANCIAS EN TODOS LOS CASOS SON LAS MISMAS.

PERCEPTUAL MAPING

SIRVE PARA APRENDER LOS DATOS A PARTIR DE EVALUACIONES SUBJETIVAS U OBJETIVAS DE LOS MISMOS

UN EJEMPLO PARA UTILIZARLO ES SI A \neq PERSONAS SE LE PIDE QUE CALIFIQUEN VARIOS PRODUCTOS DE LA MISMA CATEGORÍA EN 10 ASPECTOS DISTINTOS (CALIDAD, PRECIO, DISEÑO...)

LAPLACIAN EIGENMAPS

ESTE ALGORITMO CREA UN GRAFO A PARTIR DE LOS DATOS PARA ESTO ES NECESARIO UNA MATRIZ DE ADYACENCIAS EN DONDE CADA NODO VA A ESTAR CONECTADO CON SUS k VECINOS MAS CERCANOS Y EL PESO DE LAS ARISTAS ESTA DADO POR:

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma}}$$

ESTO PERMITE DEGULAR CUANDO SE CONSIDERA QUE DOS PUNTOS SON VECINOS (τ)
CUANTO MAS GRANDE ES EL VALOR w_{ij} MAS CERCAOS SON LOS PUNTOS.

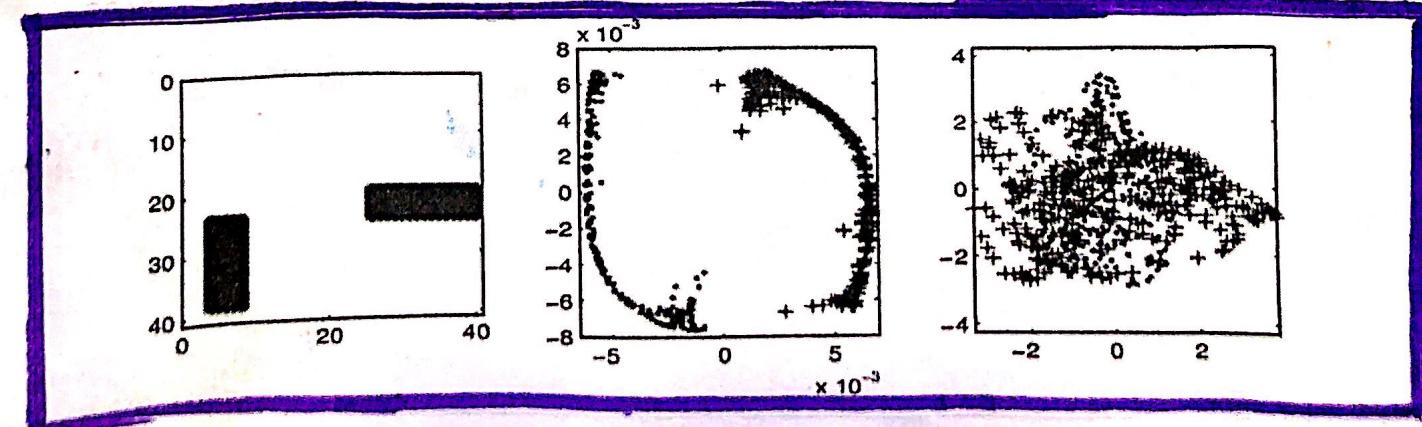
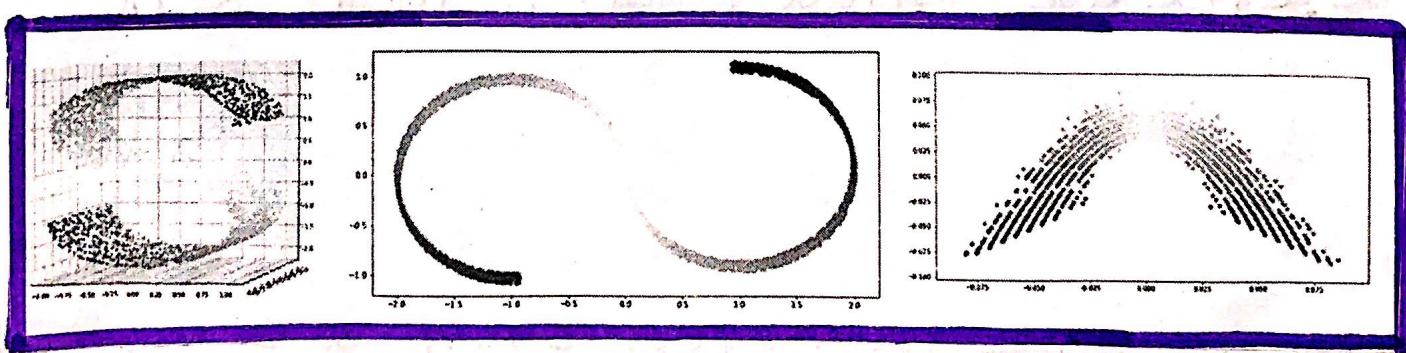
LAPLACIANO

A PARTIR DE UN GRAFO Y SU MATRIZ DE ADYACENCIA SE PUEDE CALCULAR LA MATRIZ LAPLACIANA $L = D - W$.
ESTA MATRIZ L TIENE EN SU DIAGONAL EL VALOR DEL GRADO DE CADA VERTICE.
TODOS LOS AUTOVALORES DE L SON REALES.

AHORA DESCOMPONEMOS L EN AUTOVALORES Y VECTORES
EL AUTOVALOR MAS PEQUEÑO EN VALOR ABS VA A SER CERO Y SU MULTIPLICIDAD VA A SER LA CANTIDAD DE COMPONENTES CONEXOS DEL GRAFO.

CON TODO ESTO SE PUEDE PROBAR QUE SI LOS PUNTOS SON CERCAOS EN EL ESPACIO ORIGINAL \Rightarrow LA DESCOMPOSICION ESPECTRAL LOS MANTIENE CERCAOS EN EL NUEVO ESPACIO DE DIM. REDUCIDAS.

(NO PONE RESTRICCIONES SOBRE PUNTOS QUE NO SON VECINOS).



T-SNE

TSNE ES EL MEJOR ALGORITMO PARA LA REPRESENTACIÓN DE DATOS EN DOS Y TRES DIMENSIONES. ESTE AL IGUAL QUE LAPLACIAN EIGENMAPS VA A INTENTAR QUE LOS PUNTOS QUE ESTABAN CERCA EN N DIMENSIONES SE MANTENGA CERCAOS EN K DIMENSIONES.

ESTE ALGORITMO COMIENZA:

- ① CALCULANDO LA PROBABILIDAD DE QUE j SEA VECINO DE i

ACA CADA PUNTO TIENE ASOCIADO UN PARÁMETRO σ_i^2 QUE INDICA LA DENSIDAD AL REDONDO DEL PUNTO PARA CONSIDERAR A OTRO COMO CERCANO Y ESTA ES LA MAYOR DIFERENCIA FUNDAMENTAL CON LAPLACIAN EIGENMAPS EN DONDE σ ES GLOBAL

- ② ENTONCES SE PUEDE CALCULAR LA PROBABILIDAD DE QUE DOS PUNTOS SEAN VECINOS QUE ES IGUAL A LA PROBABILIDAD DE QUE i SEA VECINO DE j Y VISEVERSA.

- ③ SI LLAMAMOS y A LOS PUNTOS MAPEADOS A 2 O 3 DIM, SE PUEDE CALCULAR LA PROBABILIDAD DE QUE DOS PUNTOS y_i Y y_j SEAN CERCANOS UTILIZANDO UNA DISTRIBUCIÓN T DE STUDENT

EL OBJETIVO DE T-SNE ES LOGRAR QUE SI p_{ij} ES UN VALOR CERCANO A 1 ENTONCES q_{ij} LO SEA TMB Y SI p_{ij} ES CERCANO A CERO q_{ij} QUEDA LIBRE. ESTO SE LOGRA CON LA DIVERGENCIA DE KULLBACK LEIBLER (MINIMIZANDOLA)

UMAP

| | t-SNE | UMAP |
|---------------------------|-----------|------------|
| Soporta nuevos puntos | X | ✓ |
| Reduce a dimensiones | 1~3 | Cualquiera |
| Tiempo | Regular | Más rápido |
| Requiere reducción previa | Sí | No |
| Calidad | Muy buena | Mejor? |
| | Free | Pro |