

ISOMAP

María Angélica Parra Mendoza

Data Science - Minería de Datos basado en Python – Taller 6

Convocatoria Mujeres Digitales 2021 – SENA

ISOMAP: MAPEo de características ISOmétricas

Es un método matemático para visualizar los datos, de escalado multidimensional, MDS, que combinado con una métrica especial llamada *distancia geodésica*, permite reducir la dimensionalidad de los datos muestreados desde una colección.

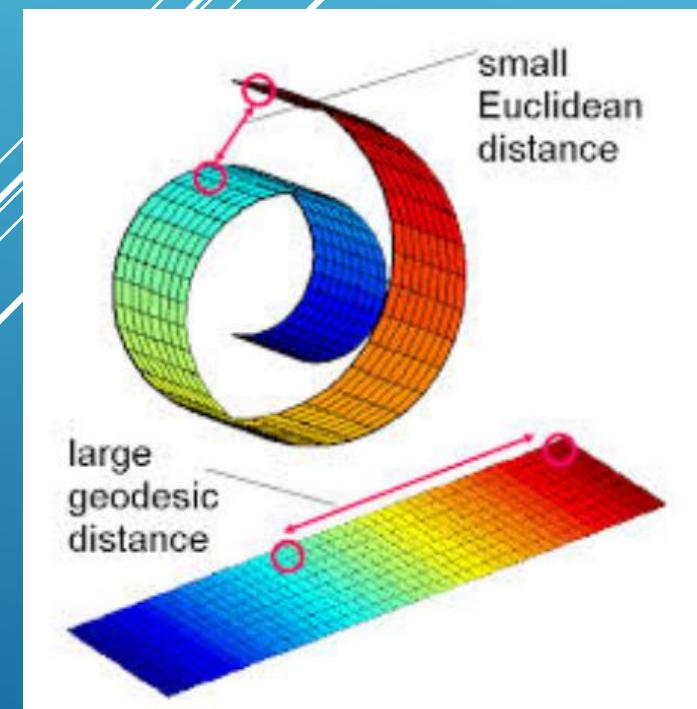
Distancia Euclíadiana

Es un número positivo que indica la separación que tienen dos puntos en un espacio donde se cumplen los axiomas y teoremas de la geometría de Euclides.

La distancia entre dos puntos P y Q de un espacio euclidiano es la longitud del vector PQ perteneciente a la única recta que pasa por dichos puntos.

Distancia Geodésica

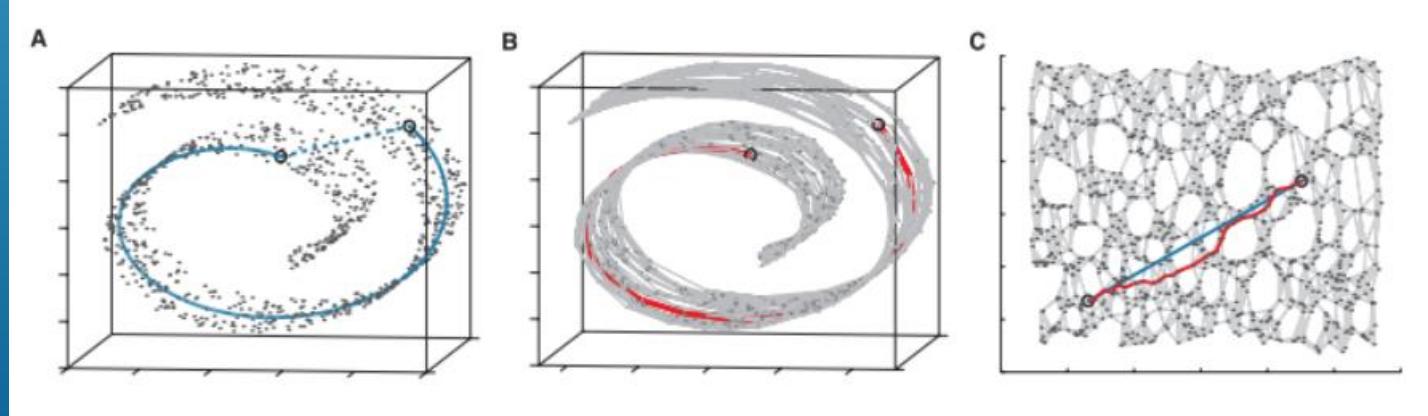
Es la distancia de la longitud de la curva contenida en los datos que une los dos puntos de la distancia Euclíadiana.



Caminos más cortos

De momento, tenemos solo las distancias geodésicas entre las observaciones que son vecinas, para las que no lo son la distancia entre estas es infinito. Pero, podemos calcular la distancia entre dos observaciones que no son vecinas; x_i y x_j como la distancia del camino más corto para ir de x_i a x_j .

Los algoritmos que se usan para encontrar este camino más corto entre dos nodos en el grafo son **Dijkstra** o **Floyd**.



Algoritmo ISOMAP

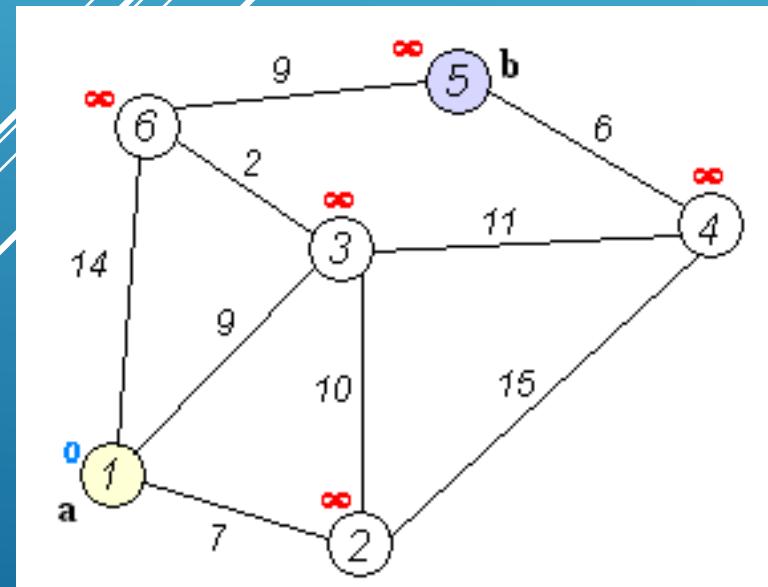
Entrada: k (número de vecinos más cercanos), p (número de dimensiones a obtener en MDS), L (matriz de datos $d \times k$)

Salida: X_p , donde $p < d$

Paso 1 Construir el grafo de los k -vecinos más cercanos

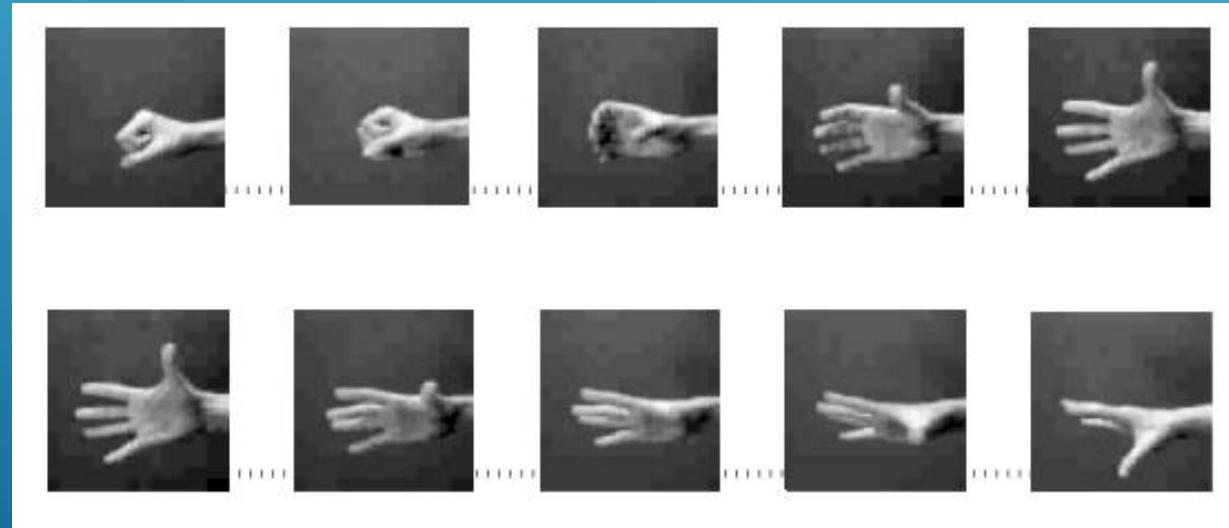
Paso 2 Calcular la matriz G de distancias usando Floyd o Dijkstra

Paso 3 Aplicar MDS a G para reducir la dimensión manteniendo las distancias geodésicas



Aplicación del algoritmo

Para conjunto de imágenes faciales sintéticas, que se sabe que tienen tres grados de libertad, ISOMAP correctamente detecta la dimensionalidad y separa los verdaderos factores subyacentes. El algoritmo también recupera lo conocido de la estructura de baja dimensión de un conjunto de imágenes reales con ruido, generadas por una mano humana que varía en la extensión de los dedos y la rotación de la muñeca.



GRACIAS

