

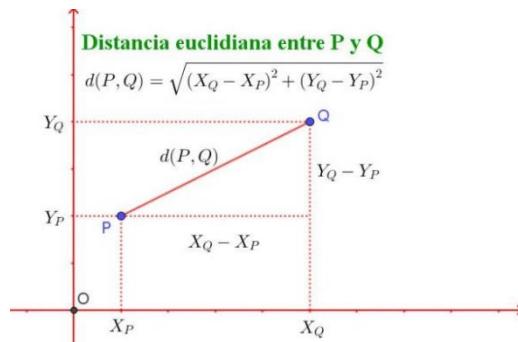
ISOMAP: MAPeo de características ISOmétricas

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.

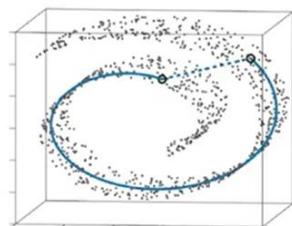
ISOMAP (Tenenbaum et al., 2000)

Se sortean datos x_1, \dots, x_n sobre una variedad \mathcal{M} de \mathbb{R}^r

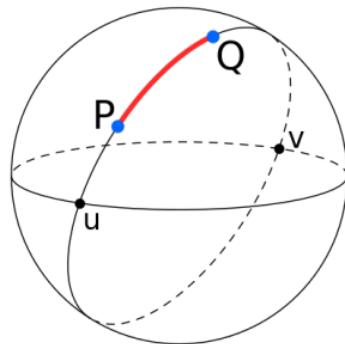
$$\begin{aligned} \mathbb{R}^r &\longrightarrow \mathbb{R}^t \\ x_i, x_j &\longrightarrow y_i, y_j \\ d_{geo}^{\mathcal{M}}(x_i, x_j) &\approx \|y_i - y_j\| \end{aligned}$$

Distancia geodésica sobre \mathcal{M} : Si $p, q \in \mathcal{M}$, definimos:

$$d_{geo}^{\mathcal{M}}(p, q) := \inf_{\substack{C \subseteq \mathcal{M} \text{ curva} \\ \text{de extremos } p \text{ y } q}} \text{longitud}(C)$$



En una esfera, es solo el gran círculo distancia.



Las distancias geodésicas exactas a menudo son imposibles de encontrar (a menos que se conozca la variedad de los datos).

Metodología

Primero: se determina que puntos son vecinos en la colección S. Con este propósito es posible utilizar dos criterios simples: el punto i se considera vecino del punto j si pertenece a los k vecinos más cercanos de j o si se encuentra a una distancia menor de un radio p . A partir de los vecindarios establecidos se construye un grafo G sobre todos los datos de entrada, y se establecen las longitudes de las aristas como $d_x(ij)$ entre puntos vecinos.

Segundo: se realiza el cálculo de los caminos más cortos, estimando las distancias geodésicas $d_s(ij)$ entre todos los pares de puntos en S y calculando las distancias de los caminos o trayectorias más cortas $d_G(ij)$ en el grafo G. Inicialmente $d_G(ij) = d_x(ij)$, si i, j están conectados por una arista. En otro caso, es decir si el punto i y el punto j no se encuentran conectados $d_G = \infty$. La matriz $D_G = d_G(ij)$ contiene las distancias de los caminos más cortos entre todos los pares de puntos en el grafo G.

Hay dos métodos comunes para resolver el camino más corto y la distancia más corta:

El problema de la ruta más corta para determinar el punto de partida, es decir, el problema de encontrar la ruta más corta cuando se conoce el nodo inicial. Adecuado para utilizar el algoritmo de Dijkstra;

Problema global de la ruta más corta: busque todas las rutas más cortas en el gráfico. Es adecuado utilizar el algoritmo Floyd-Warshall.

Tercero: se aplica MDS clásico a la matriz de distancias del grafo $D_G = \{d_G(ij)\}$ construyendo una inmersión de los datos a un espacio de dimensión m que preserva de mejor manera la geometría intrínseca de la variedad original. Los puntos y_i del espacio de salida se eligen de manera que se minimice la función de costo

$$E = \|\tau(\mathbf{D}_G) - \tau(\mathbf{D}_Y)\|_{L^2}$$

donde D_Y es la matriz de distancias Euclídeas

$$(d_Y(i,j) = \|\mathbf{y}_i - \mathbf{y}_j\|) \text{ y } \|A\|_{L^2} \text{ la norma L}^2 \text{ de la matriz } \sqrt{\sum_{i,j} A_{ij}^2}.$$

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.



Bibliografía

https://www.youtube.com/watch?v=23hJDk8Ad9k&ab_channel=AndresFarall

<https://www.lifeder.com/distancia-euclidiana/>

https://www.youtube.com/watch?v=5tOgOzbnpw8&ab_channel=AndresFarall

<https://core.ac.uk/download/pdf/236385169.pdf>

<https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec10ISOMap.pdf>

<https://www.robots.ox.ac.uk/~az/lectures/ml/tenenbaum-isomap-Science2000.pdf>