

Isometric Feature Mapping

NICOLL CRESPO RONCALLO



TEMAS DE LA PRESENTACIÓN

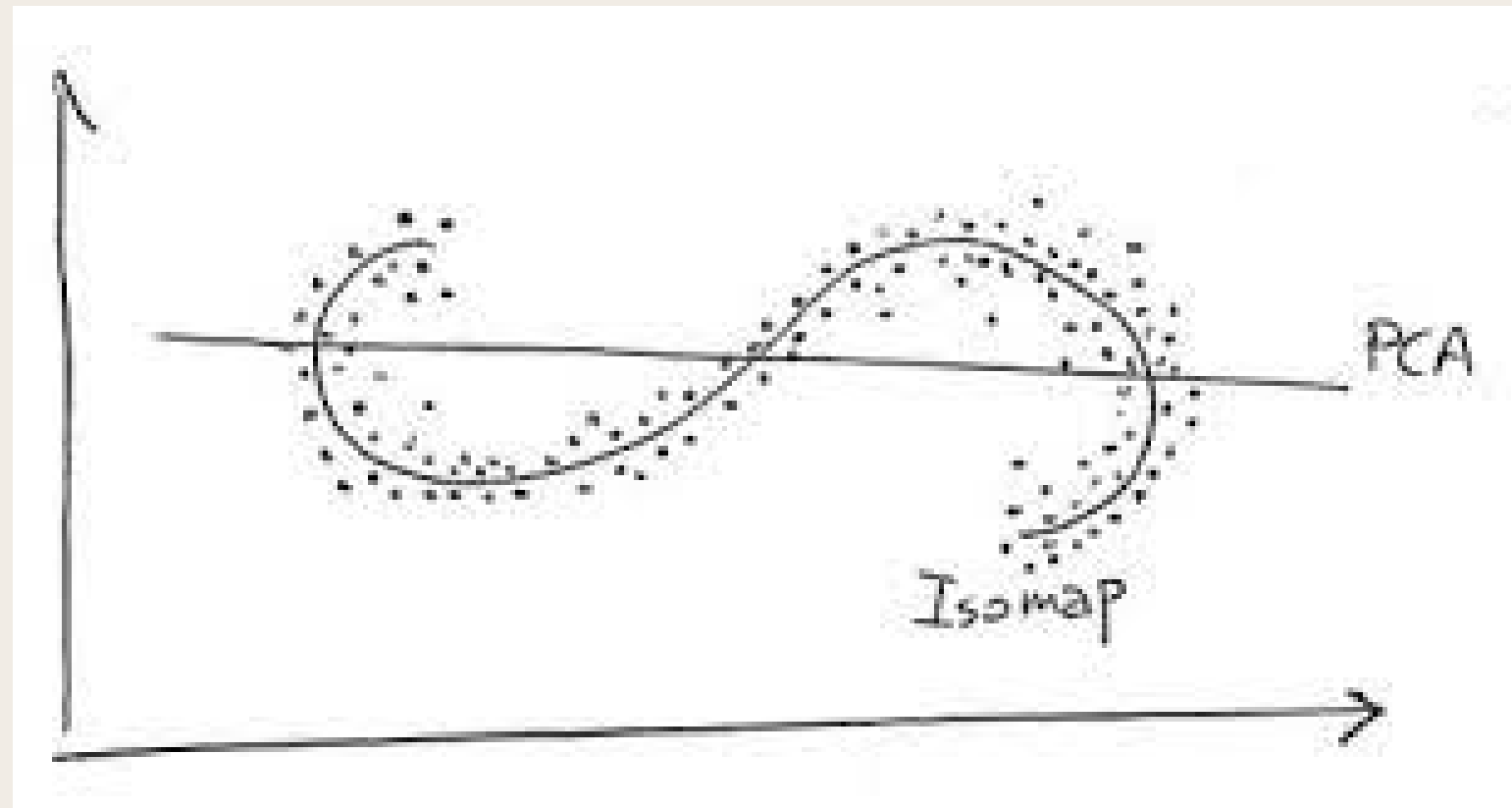
DEFINICIÓN

USOS

METODOLOGÍA

CARACTERÍSTICAS

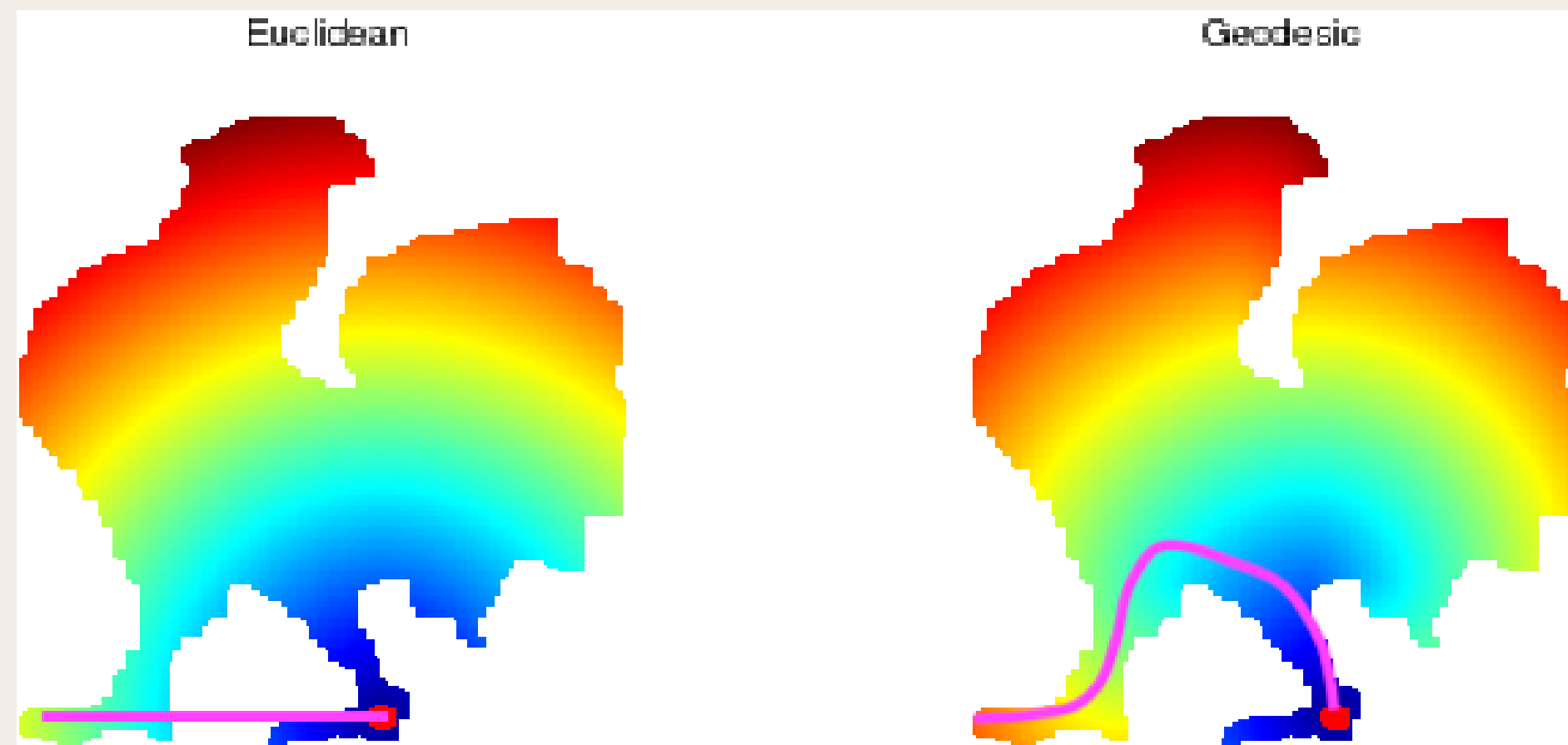
APLICACION



ISOMAP

¿ QUE ES ?

Isomap es una técnica de reducción de dimensión, no lineal basada en MDS (Análisis de escalamiento multidimensional) clásico. Fue introducido por primera vez por Tenenbaum, de Silva, and Langford (1) en el 2000.



¿QUE BUSCA ?

Conservar la geometría propia de los datos, estimando las distancias geodésicas de forma gráfica, dada solamente su distancia Euclídea en el espacio de alta dimensión

¿CUANDO SE USA?

DATOS CONTENIDOS EN UN MANIFOLD NO LINEAR DE D-DIMENSIÓN

No responderán bien a métodos de reducción de dimensión lineal como PCA o MDS clásico. En cambio ISOMAP es un método reducción no lineal que recupera la verdadera dimensión y estructura geométrica de esos datos. Incluso funciona con datos sin una geometría de Manifold clara.

DENSIDAD DE DATOS SUFICIENTE

Debido que usa una aproximación gráfica de la distancia geodésica, esta será más certera en tanto el conjunto de datos sea más denso.

¿PORQUE SE USA?

CONSERVA LO MEJOR DE AMBOS MUNDOS

Gracias a la estimación de la distancia geodésica, conserva la geometría intrínseca de los datos y hereda las cualidades de optimalidad global de los métodos lineales

EFICIENCIA Y ADAPTABILIDAD

ISOMAP es altamente eficiente debido a su simplicidad y se puede usar en varios tipos de datos y dimensiones



METODOLOGÍA



1

Construcción del gráfico de vecindad

Generalmente a través de los k-vecinos más cercanos o ϵ -vecinos (menos usado)



2

Calculo de las distancias gráficas más cortas

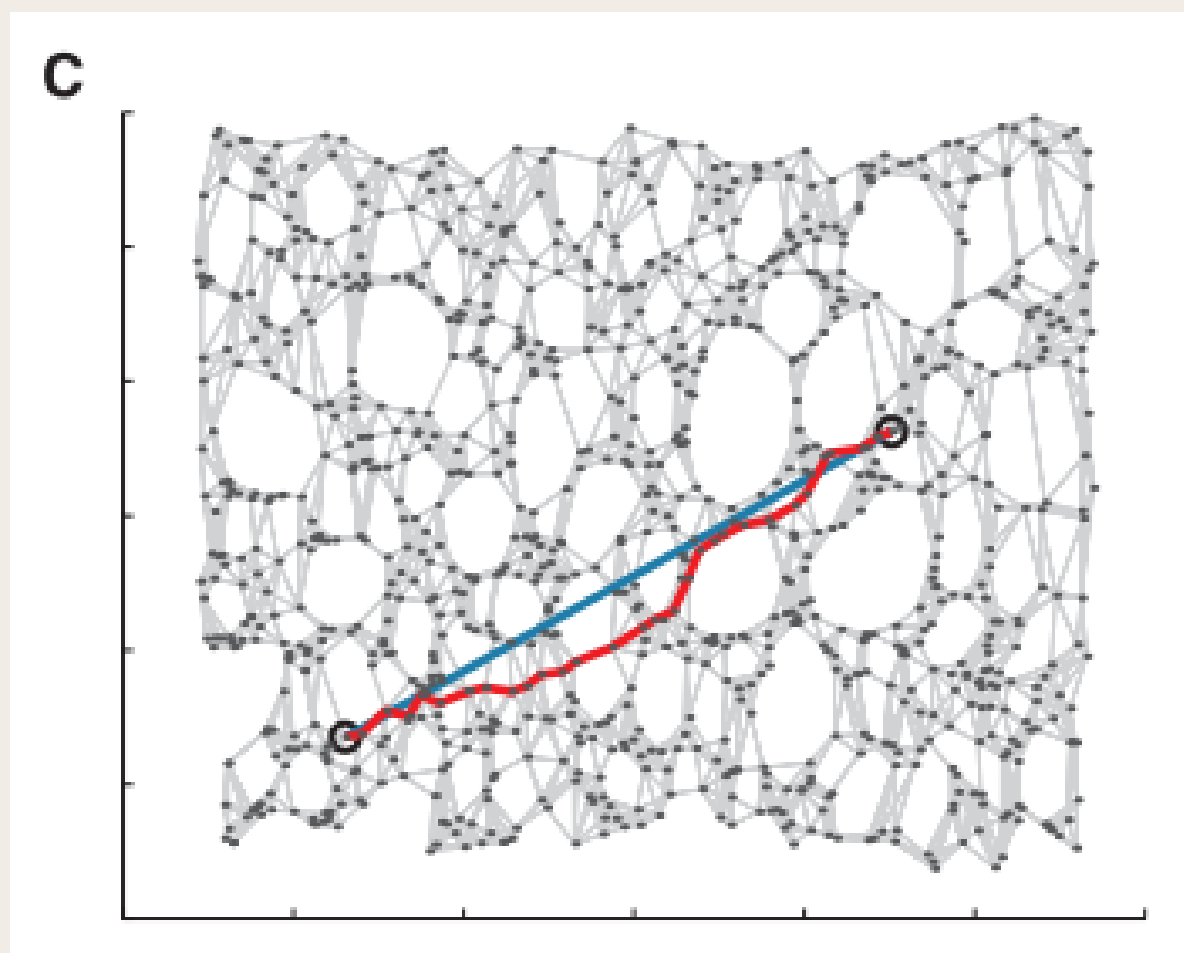
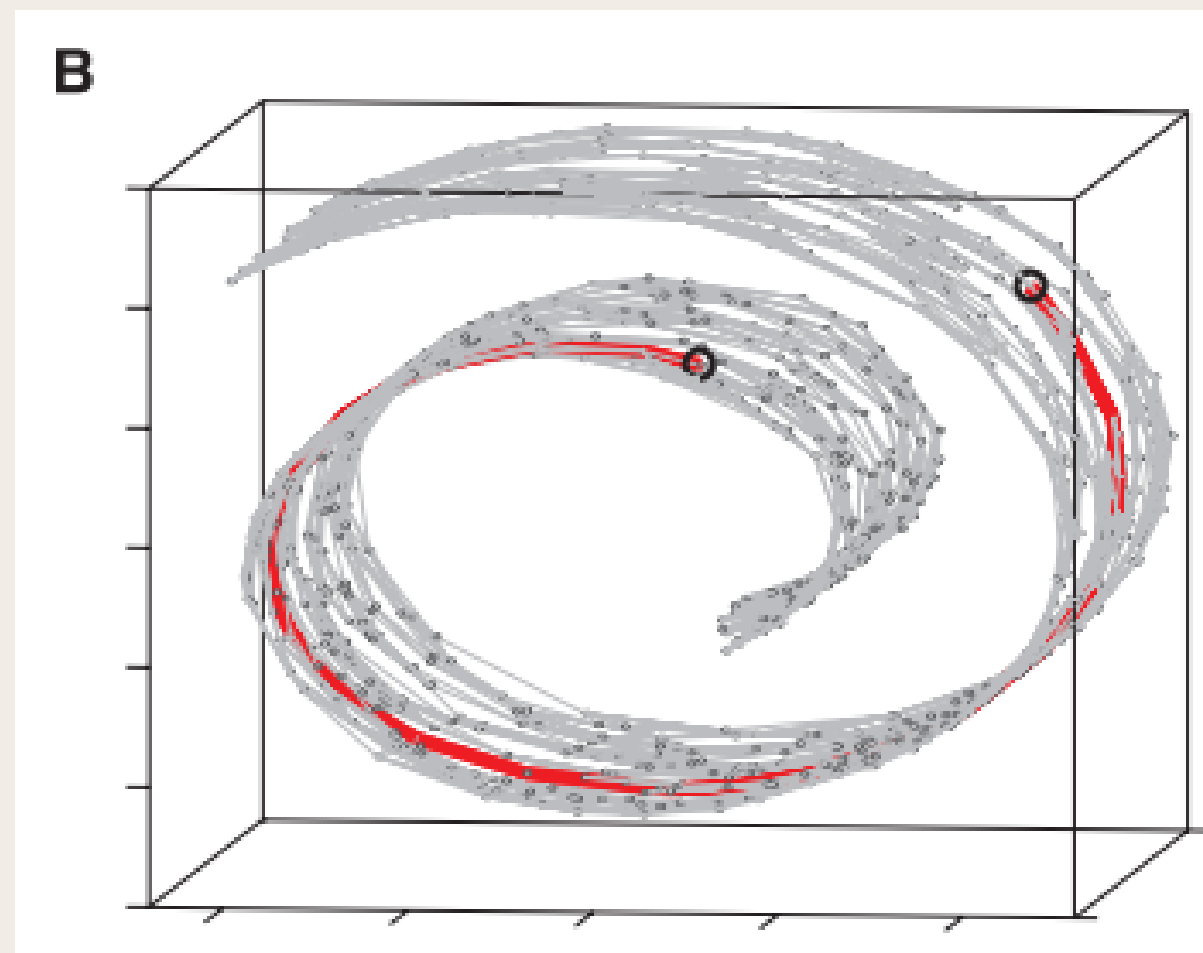
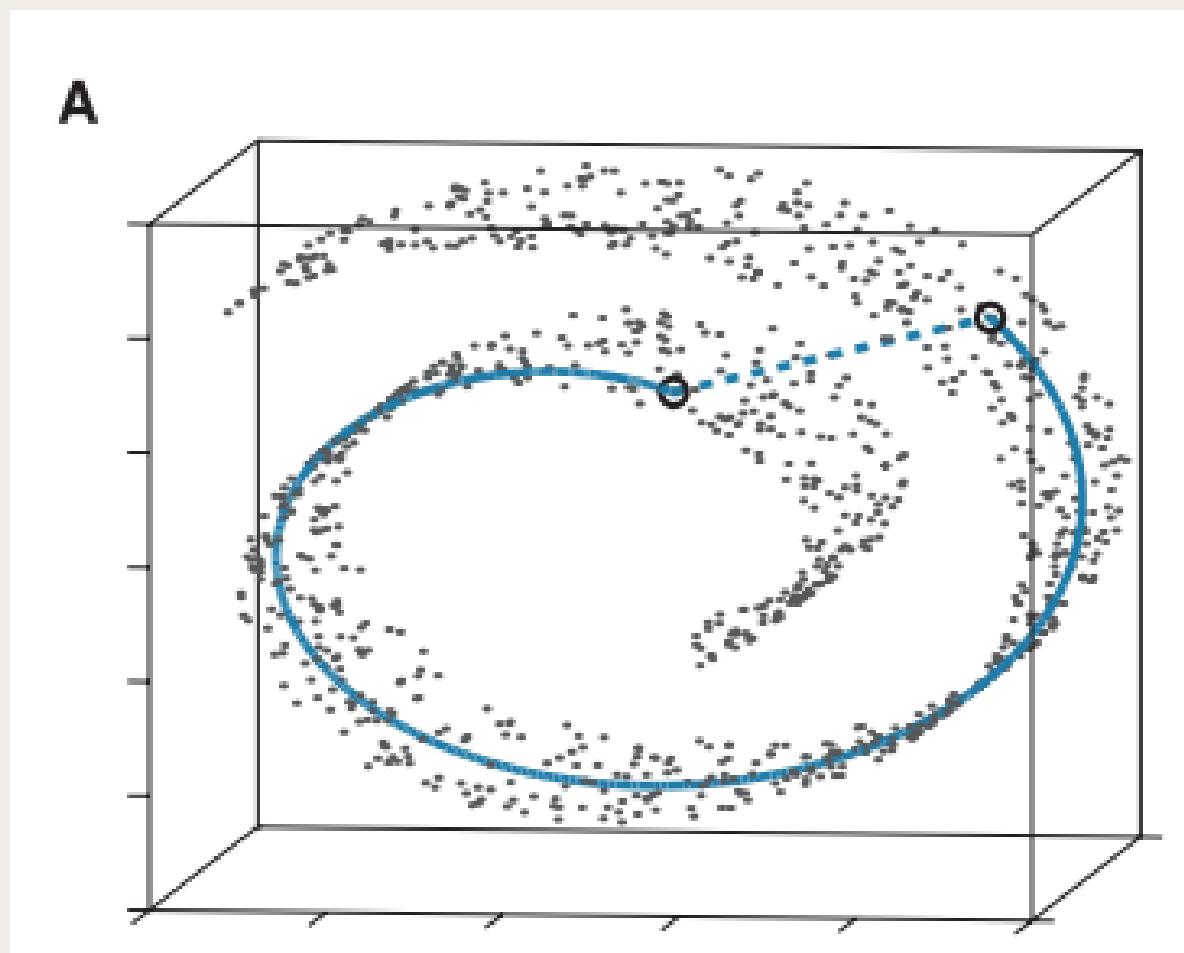
Los algoritmos conocidos más eficientes para esto son el algoritmo de Dijkstra (menos tiempo computacional) y el algoritmo Floyd-Warshall



3

Aplicar el MDS clásico.

A la matriz de distancias gráficas. Mediante una descomposición parcial de valores propios.



Caso del "Swiss Roll".

A) Diferencia entre distancia euclidiana y geodesica

B) Construcción del gráfico de vecindad con $k=7$ y 1000 puntos de datos, mostrando la aproximación de la distancia con las distancias geodesicas

C) Grafico de dos dimensiones luego de aplicar el algoritmo

CARACTERISTICAS 04

EFICIENCIA

Gracias a su simpleza comparado con otros métodos especializados de aplanamiento

03

AMPLIO RANGO DE APLICACIÓN

Debido que es más acertado que PCA Y MDS al reconocer la verdadera dimensión y se adapta incluso a casos sin forma geométrica clara

02

PRESERVACIÓN DE LA ESTRUCTURA GEOMETRICA

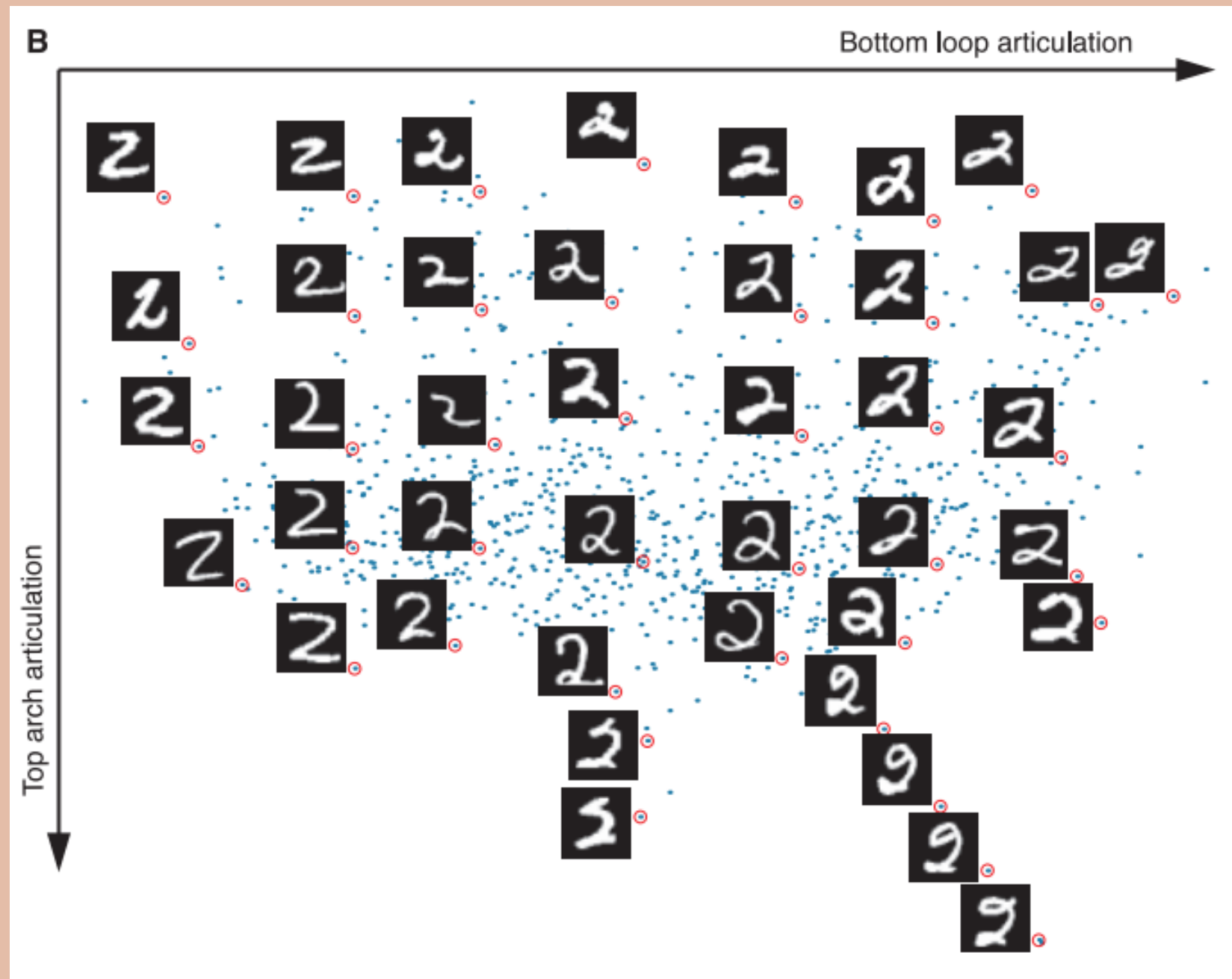
Gracias a el uso de la estimación de distancias geódesicas

01

REDUCCION DE DIMENSION NO LINEAR

Aplicable a datos con formas geométricas no lineales

APLICACIÓN



- Reducción a 2 dimensiones: x-articulacion del bucle inferior, y-articulación del arco superior
- Las distancias de espacio de entrada, se midieron mediante la distancia tangente, una métrica diseñada para capturar las invariancias relevantes en el reconocimiento de escritura a mano
- Uso de ϵ -ISOMAP (con $\epsilon = 4.2$), debido que no se esperaba una misma dimensión en todo el conjunto de datos
- De hecho, ISOMAP encuentra varios casos que sobresalen desde el conjunto de datos de dimensión más alta y que representan exageraciones sucesivas de un trazo o adorno adicional en el dígito

REFERENCIAS

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science (New York, N.Y.)*, 290(5500), 2319–2323.

Wang, J. (2012). *Geometric structure of high-dimensional data and dimensionality reduction (Vol. 5)*. Berlin Heidelberg: Springer.