

Métodos locales de reducción de dimensión

Parte 1: ISOMAP

Parte 2: T-SNE

Parte 3: Local Linear Embedding (LLE)

Parte 4: Self Organizing Maps

Parte 1: ISOMAP

J. B. Tenenbaum, et al.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319-2323
<http://www-clmc.usc.edu/publications/T/tenenbaum-Science2000.pdf>

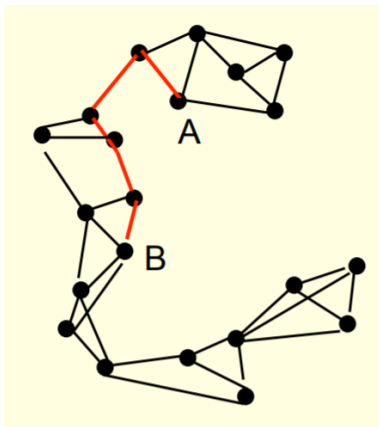
Idea: Hacer CDS con distancias entre puntos calculadas a partir de un grafo que refleja la estructura local de los datos.

Parte 1: ISOMAP

J. B. Tenenbaum, et al.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319-2323
<http://www-clmc.usc.edu/publications/T/tenenbaum-Science2000.pdf>

Idea: Hacer CDS con distancias entre puntos calculadas a partir de un grafo que refleja la estructura local de los datos.

1. Construye un grafo ponderado basado en estructura local: cada dato es un vértice; conecta un dato con sus k -vecinos más cercanos (simetrizar); pesos son distancias
2. Calcula para cada par de datos $d(x_i, x_j)$ la distancia del camino más corto entre x_i y x_j sobre el grafo.
3. Aplica escalamiento multidimensional (CDS) a partir de $\{d(x_i, x_j)\}$

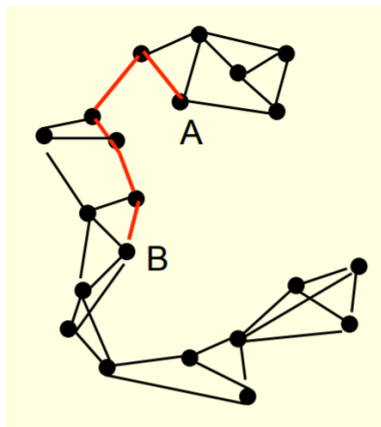


Parte 1: ISOMAP

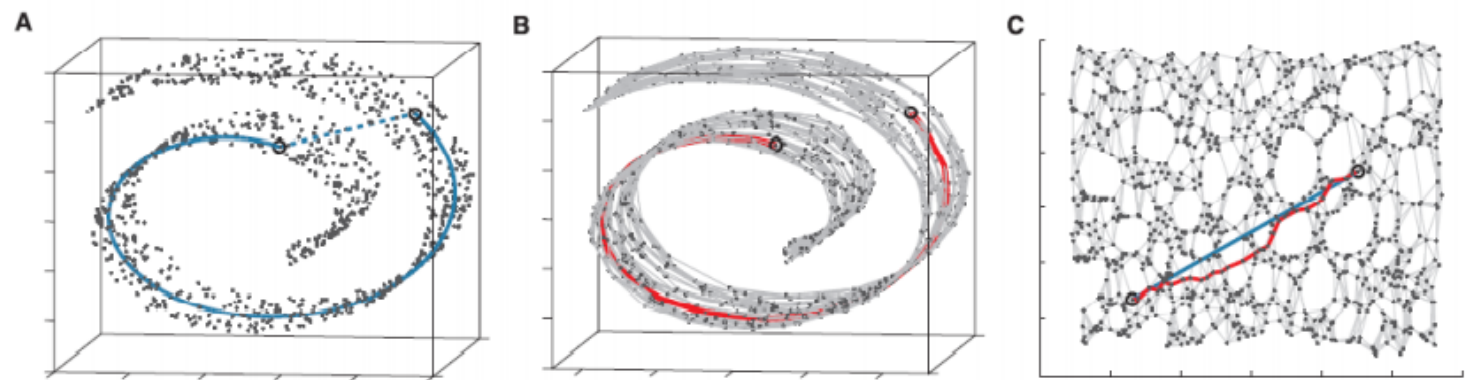
J. B. Tenenbaum, et al.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319-2323
<http://www-clmc.usc.edu/publications/T/tenenbaum-Science2000.pdf>

Idea: Hacer CDS con distancias entre puntos calculadas a partir de un grafo que refleja la estructura local de los datos.

1. Construye un grafo ponderado basado en estructura local: cada dato es un vértice; conecta un dato con sus k-vecinos más cercanos (simetrizar); pesos son distancias
2. Calcula para cada par de datos $d(x_i, x_j)$ la distancia del camino más corto entre x_i y x_j sobre el grafo.
3. Aplica escalamiento multidimensional (CDS) a partir de $\{d(x_i, x_j)\}$



Hinton et al.

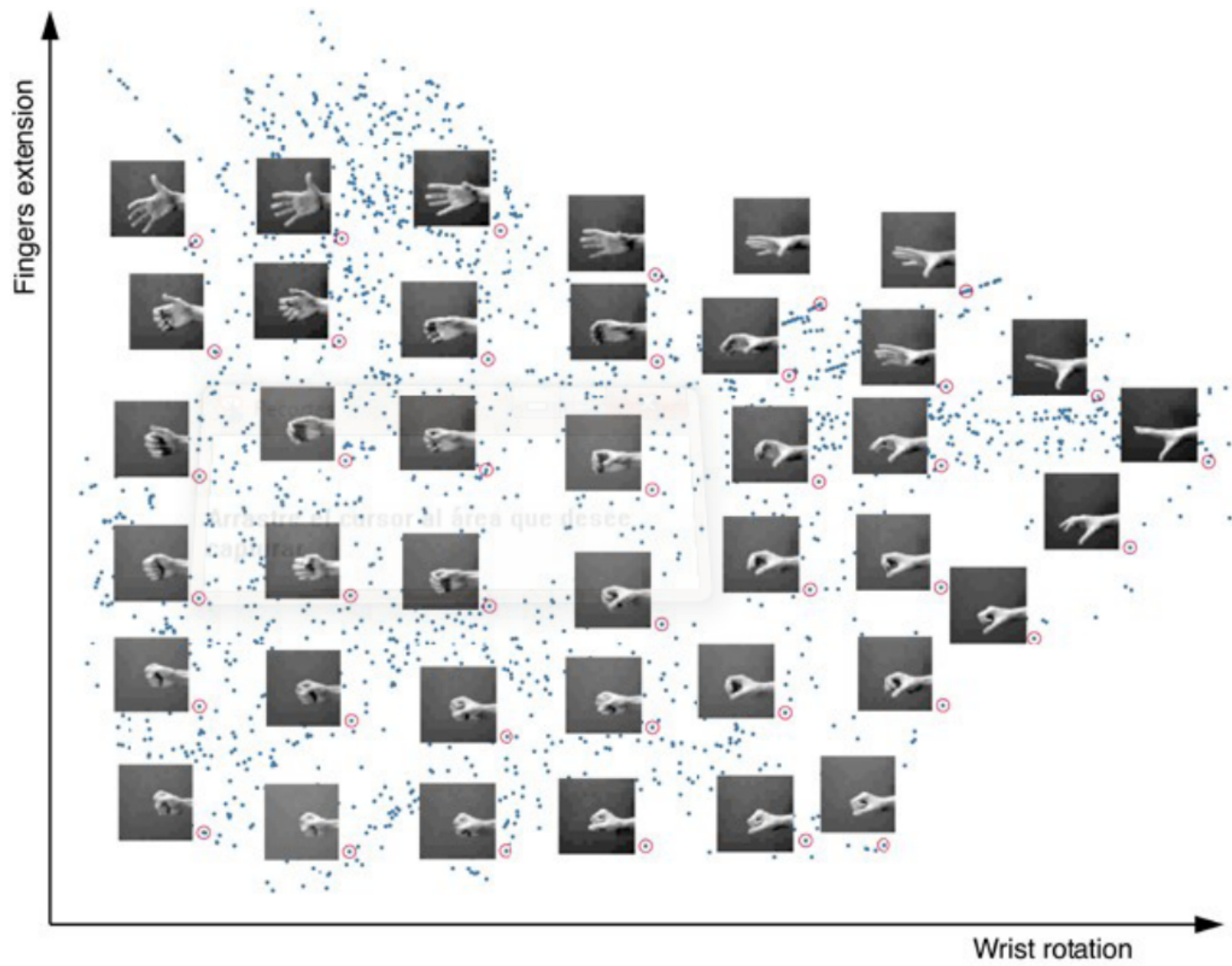


(Tenenbaum et al.)

A

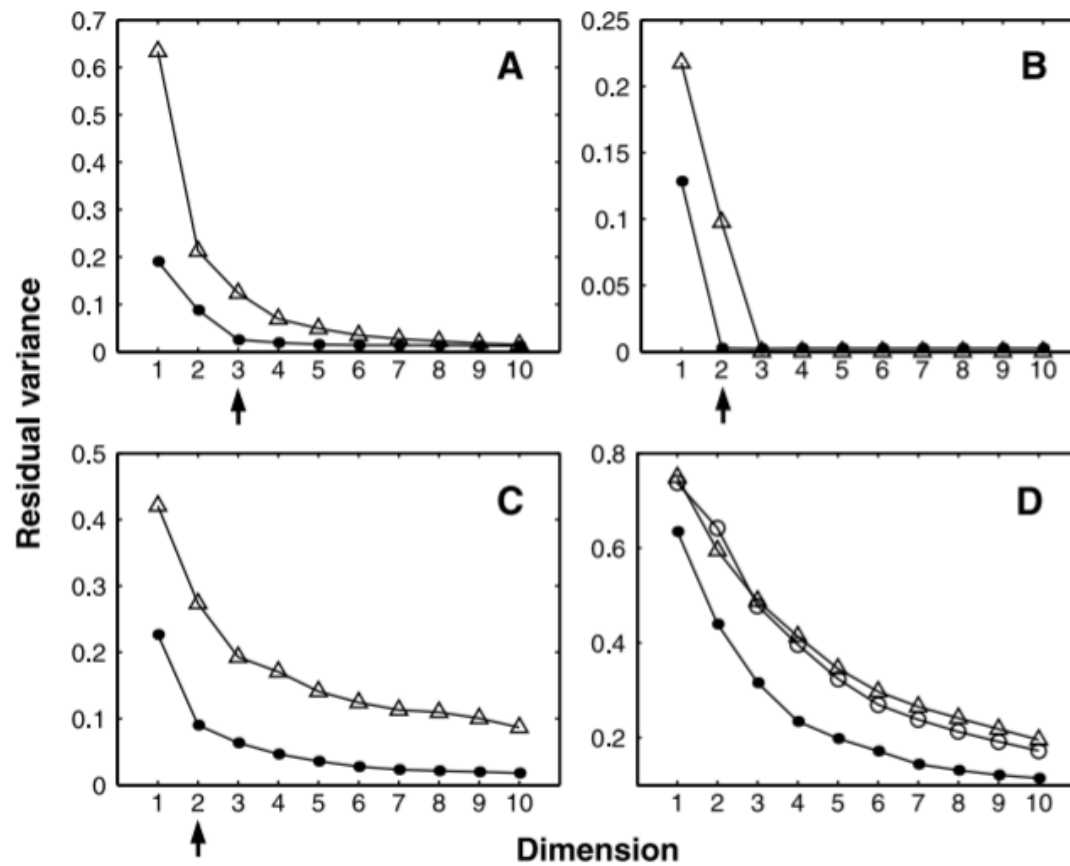


(Tenenbaum et al.)



(Tenenbaum et al.)

Fig. 2. The residual variance of PCA (open triangles), MDS [open triangles in (A) through (C); open circles in (D)], and Isomap (filled circles) on four data sets (42). (A) Face images varying in pose and illumination (Fig. 1A). (B) Swiss roll data (Fig. 3). (C) Hand images varying in finger extension and wrist rotation (20). (D) Handwritten "2"s (Fig. 1B). In all cases, residual variance decreases as the dimensionality d is increased. The intrinsic dimensionality of the data can be estimated by looking for the "elbow"



at which this curve ceases to decrease significantly with added dimensions. Arrows mark the true or approximate dimensionality, when known. Note the tendency of PCA and MDS to overestimate the dimensionality, in contrast to Isomap.

(Tenenbaum et al.)

En R: `library(vegan) isomap()`

En Python:

```
from sklearn import manifold  
iso = manifold.Isomap(n_neighbors=6, n_components=2)  
iso.fit(datos)
```