

Taller Semana 6

ISOMETRIC FEATURE MAPPING

Nicoll Crespo Roncallo

1. Definición

Isomap o Isometric Feature Mapping fue introducido por primera vez por Tenenbaum, de Silva, and Langford (1) en el 2000. Es una técnica de reducción de dimensión, no lineal basada en MDS (Análisis de escalamiento multidimensional) clásico, usada en largos volúmenes de datos y que es capaz de mantener la estructura original de los datos.

2. Objetivo

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.

3. Uso

Este algoritmo es usado cuando:

- **Datos contenidos en un manifold no linear de d-dimensión**
Este tipo de estructura de datos no responderán bien a métodos de reducción de dimensión lineal como PCA o MDS clásico. En cambio ISOMAP es un método de 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 a que usa una aproximación gráfica de la distancia geodésica, esta será más certera en tanto el conjunto de datos sea más denso.

Se usa porque:

- **Conserva lo mejor de ambos tipos de algoritmos**
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.

4. Metodología

El algoritmo se compone de tres pasos principales:

- **Construcción del gráfico de vecindad**
Generalmente a través de los k-vecinos más cercanos o ϵ -vecinos (menos usado)
- **Cálculo 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

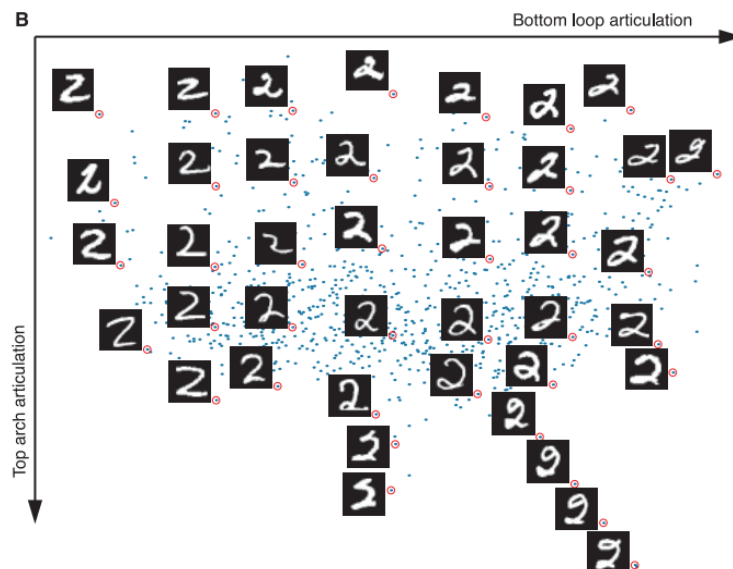
- **Aplicar el MDS clásico**

Se le aplica a la matriz de distancias gráficas. Mediante una descomposición parcial de valores propios.

5. Características

- **Reducción de dimensión no lineal:** Aplicable a datos con formas geométricas no lineales
- **Preservación de la estructura geométrica:** Gracias a el uso de la estimación de distancias geodésicas
- **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
- **Eficiencia:** Gracias a su simpleza, comparado con otros métodos especializados de aplanamiento. Sin embargo, en la medida en que un conjunto de datos presenta valores extremos o se desvíe de una densidad uniforme, la convergencia asintótica todavía se mantiene en general, pero el tamaño de muestra requerido para estimar la distancia geodésica con precisión puede ser demasiado grande.

6. Aplicación



- En este caso el algoritmo se aplica a un conjunto de 1000 datos de números “2” escritos a mano. El resultado obtenido es una reducción de dimensiones a 2: en “x” la articulación del bucle inferior y en “y” la articulación del arco superior.
- Las distancias de espacio de entrada, se midieron mediante la distancia tangente, una métrica diseñada para capturar las variaciones 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
- ISOMAP encuentra además, 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.

7. 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.