UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Métodos locales I

Alan Reyes-Figueroa

Aprendizaje Estadístico

(Aula 11) 12.febrero.2024

Métodos locales

Recordemos la idea subyacente en el escalamiento multidimensional: mapear los datos $\mathbf{x}_i \in \mathbb{R}^d$ a un espacio de menor dimensión $\mathbf{x}_i^* \in \mathbb{R}^p$, con $p < d$

$$\min_{\mathbf{x}_i^*, \mathbf{x}_j^*} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(d(\mathbf{x}_i, \mathbf{x}_j)^2 - d(\mathbf{x}_i^*, \mathbf{x}_j^*)^2 \right)^2. \tag{1}$$

Los métodos locales tienen el mismo propósito, queremos reducir la dimensión de los datos $\mathbf{x}_i$. De igual forma, mapeamos los datos via una función (no lineal) $f : \mathbb{R}^d \to \mathbb{R}^p$, $f(\mathbf{x}_i) = \mathbf{x}_i^*$.
de modo que f preserve la estructura de los datos originales $\mathbf{x}_i$.

Obs! La diferencia con los métodos globales (PCA, MDS) es que no utilizan todos los datos, y usualmente no son lineales.

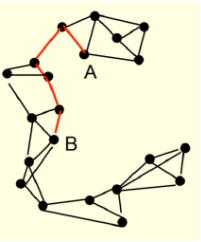
Isomap

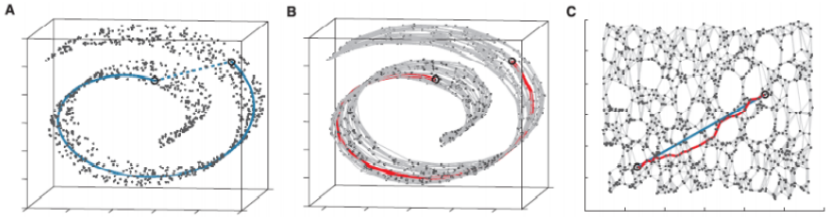
Ref: J. B. Tenenbaum *et al.* A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319-2323.
`http://www-clmc.usc.edu/publications/T/tenenbaum-Science2000.pdf`

Idea: Hacer MDS (escalamiento multidimensional) con distancias entre puntos calculadas a partir de un grafo que refleja la estructura local de los datos.

- Construye un grafo ponderado G basado en estructura local: cada dato $\mathbf{x}_i$ es un vértice; conecta un dato con sus k-vecinos más cercanos (simetrizar); pesos son distancias.
- Calcula para cada par de datos $d(\mathbf{x}_i, \mathbf{x}_j)$ la distancia del camino más corto entre $\mathbf{x}_i$ y $\mathbf{x}_j$ sobre el grafo G (algoritmo de Dijkstra).
- Aplicar escalamiento multidimensional a partir de $\{d(\mathbf{x}_i, \mathbf{x}_j)\}$

Isomap



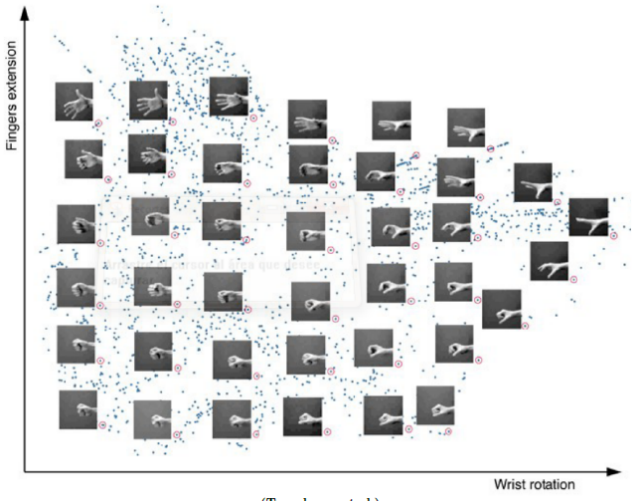


(Tenenbaum et al.)

Isomap

(Tenenbaum et al.)

Isomap



(Tenenbaum et al.)

Isomap

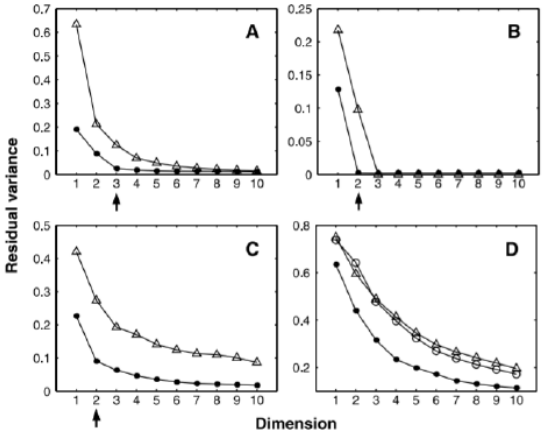


Fig. 2. The residual variance of PCA (open triangles), MDS [open triangles in (A) through (C); open circles in (D)], and Isomap (filled circles) on four data sets (*42*). (**A**) Face images varying in pose and illumination (Fig. 1A). (**B**) Swiss roll data (Fig. 3). (**C**) Hand images varying in finger 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.

t-SNE

Refs: SNE: Roweis, Sam; Hinton, G. (2002). Stochastic neighbor embedding. Neural Information Processing Systems.
T-SNE: van der Maaten, L.J.P.; Hinton, G.E. (2008). Visualizing Data Using t-SNE. Journal of Machine Learning Research.

Idea: convierte similitudes entre datos en probabilidades de un experimento aleatorio. Trata de conservar estas distribuciones en el nuevo espacio.

- Para un dato $\mathbf{x}_i$ define $P_i: \ p_{j|i} =$ probabilidad de elegir $\mathbf{x}_j$ como vecino: entre más similar, mayor probabilidad.
- Buscamos datos $\{\mathbf{x}_i^*\}$ con $Q_i: \ q_{j|i} =$ probabilidad de elegir $\mathbf{x}_j^*$ como vecino de $\mathbf{x}_i^*$, tal que las distribuciones $p_{j|i}$ y $q_{j|i}$ se parecen.

¿Cómo medir distancias entre distribuciones? Divergencia Kullback-Leibler: $D_{KL}(P||Q) = \sum_i P_i \log \frac{P_i}{Q_i}$.

SNE

SNE: *stochastic neigbourhood embedding*.
Definir:

$$P_i: \quad p_{j|i} = \frac{1}{c_i}\exp(-||x_j - x_i||^2/\sigma_i), \quad c_i = \sum_{k\neq i}\exp(-||x_k - x_i||^2/\sigma_i);$$

$$Q_i: \quad q_{j|i} = \frac{1}{c_i^*}\exp(-||x_j^* - x_i^*||^2), \quad c_i^* = \sum_{k\neq i}\exp(-||x_k^* - x_i^*||^2).$$

Función de costo: $J = \sum_i d(P_i, Q_i)$.

La derivada de la funcion de costo en $\frac{\partial J}{\partial x_i}$ es

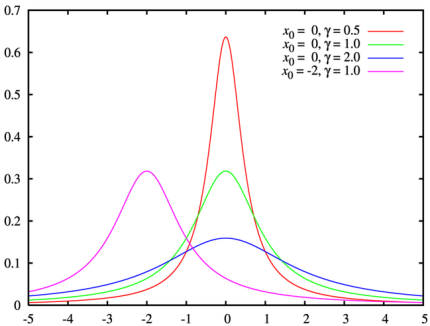
$$\frac{\partial J}{\partial x_i} = 2\sum_j (\mathbf{x}_j^* - \mathbf{x}_i^*)^2 (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}).$$

Está relacionada con atracción / repulsión.

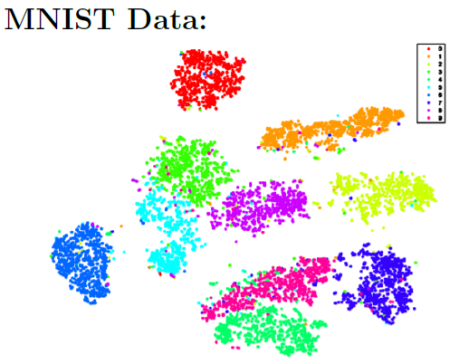
t-SNE

t-SNE: *t-distributed stochastic neigbourhood embedding*.
En el espacio de $\{\mathbf{x}_i^*\}$, cambiamos la gausiana por una distribucion t_1 (distribución Cauchy): $f(t) = \frac{1}{\pi(1+t^2)}$, tiene colas más pesadas.

$\Rightarrow$ se castiga menos distancias grandes.

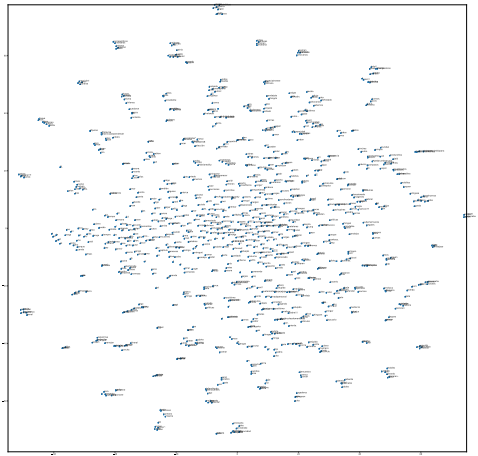


t-SNE



Explorar `https://projector.tensorflow.org/`

t-SNE



t-SNE aplicado a palabras en *tweets*.