UFM UNIVERSIDAD FRANCISCO MARROQUÍN
VERITAS • LIBERTAS • JVSTITIA

FACULTAD de CIENCIAS ECONÓMICAS

Métodos locales I (*Manifold Learning*)

Alan Reyes-Figueroa

Elements of Machine Learning (Aula 10) 21.febrero.2023

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, pero ahora usando 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, Escalamiento) es que los métodos locales no utilizan todos los datos, y usualmente no son lineales.

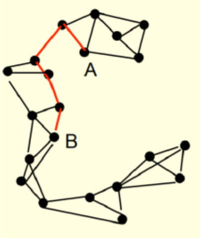
Isomap

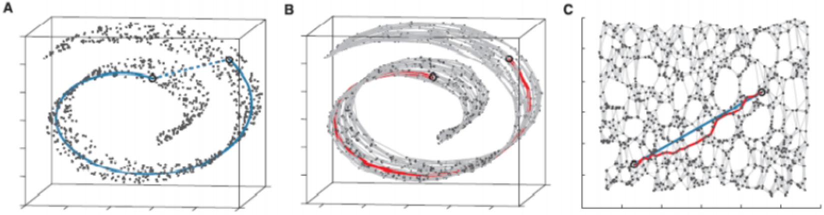
Ref: J. B. Tenenbaum *et al.* A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319-2323.
`http://www-clmc.usc.edu/publications/T/tenenbaum-Science2000.pdf`

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



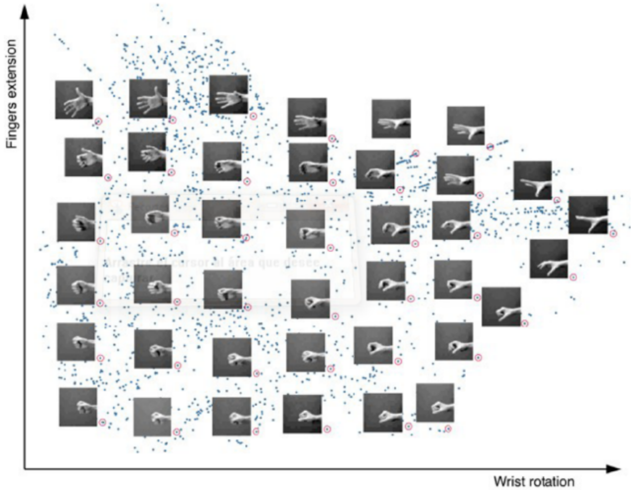


(Tenenbaum et al.)

Isomap

(Tenenbaum et al.)

Isomap



(Tenenbaum et al.)

UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Isomap

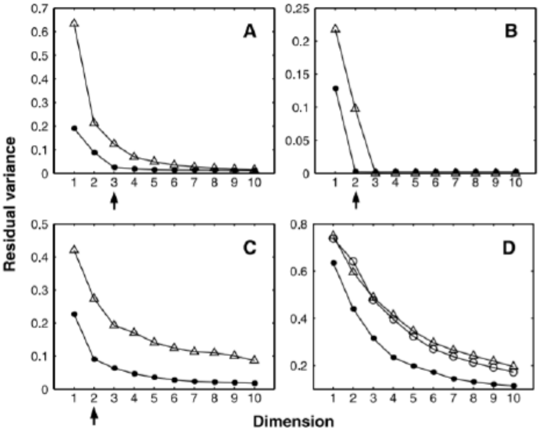


Fig. 2. The residual variance of PCA (open triangles), MDS [open triangles in (A) through (C); open circles in (D)], and Isomap (filled circles) on four data sets (*42*). (**A**) Face images varying in pose and illumination (Fig. 1A). (**B**) Swiss roll data (Fig. 3). (**C**) Hand images varying in finger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

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

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

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

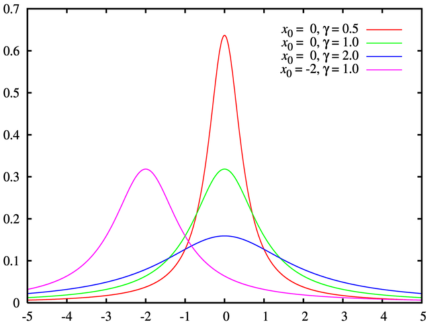
$$\frac{\partial J}{\partial x_i} = 2\sum_j (\mathbf{x}_j^* - \mathbf{x}_i^*)^2 (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}).$$

Está relacionada con atracción / repulsión.

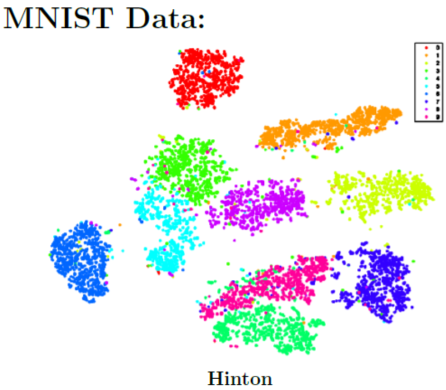
Parámetro *perplexity* (para calcular las σ_i): número de vecinos efectivos de un dato (se base en la entropía de la distribución de las distancias).

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.

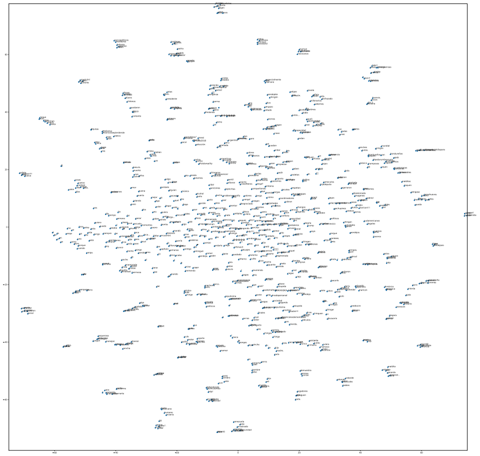


t-SNE



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

t-SNE



t-SNE aplicado a palabras en *tweets*.

LLE

LLE: *Local Linear Embedding*
Refs: Roweis ST, Lawrence LK (2000) Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science 290(5500): 2323-2326.
`https://cs.nyu.edu/ roweis/lle/publications.html`

Idea: Caracterizar la estructura local en el espacio original, y tratamos de conservar esta estructura local en el nuevo espacio.

- Si conozco los k-vecinos más cercanos a $\mathbf{x}_i$, denotados por $\{\mathbf{x}_j : j \in vec(i)\}$. Vamos a tratar de escribir $\mathbf{x}_i$ como combinación lineal de sus k-vecinos más cercanos ($k < d$)

$$\mathbf{x}_i = \sum_{j \in vec(i)} w_{ij}\mathbf{x}_j, \quad \text{com} \sum_{j \in vec(i)} w_{ij} = 1.$$

LLE

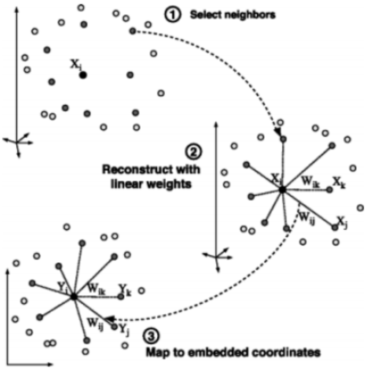
- Para cada $\mathbf{x}_i$ buscamos los k-vecinos más cercanos $\{\mathbf{x}_j : j \in vec(i)\}$.
- Resolvemos

$$\min_{w_{ij}} \sum_i ||\mathbf{x}_i - \sum_{j \in vec(i)} w_{ij}\mathbf{x}_j||^2,$$

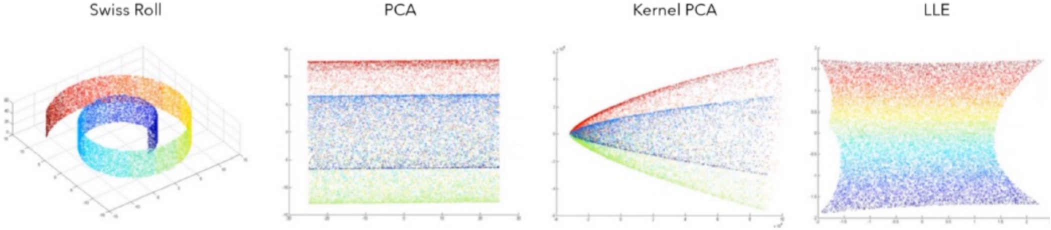
  sujeto a $\sum_j w_{ij} = 1$.
- Resolvemos

$$\min_{\mathbf{x}_j^*} \sum_i ||\mathbf{x}_i^* - \sum_{j \in vec(i)} w_{ij}\mathbf{x}_j^*||^2,$$

  sujeto a restricciones de norma y promedio de $\mathbf{x}_j^*$.

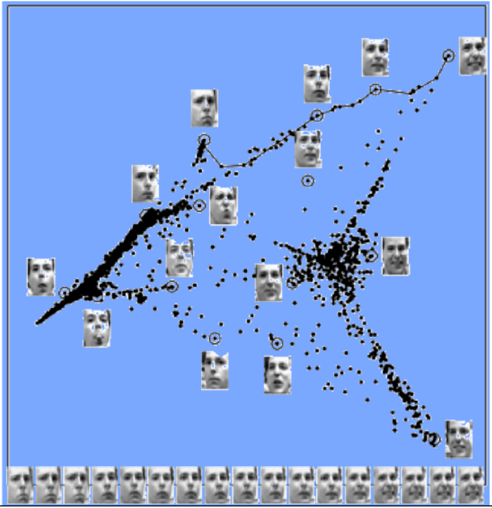


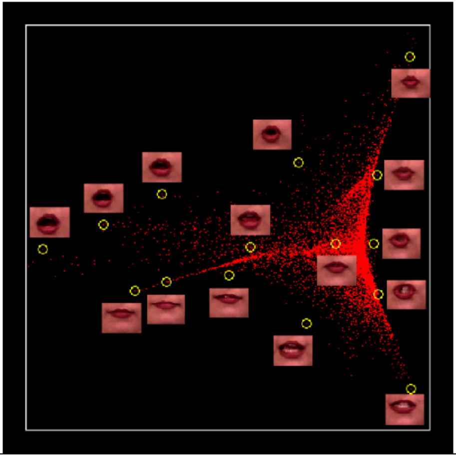
LLE



Source: Jennifer Chu. Image free to share

LLE



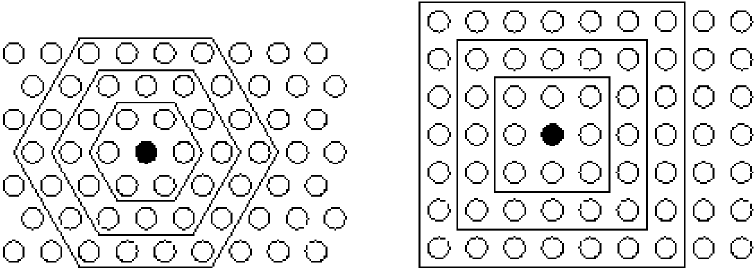


UNIVERSIDAD FRANCISCO MARROQUÍN

SOM

SOM: *Self organizing maps*
Ref: Kohonen, Teuvo (1982). Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics 43 (1): 59-69.

Idea: Colocar cada dato $\mathbf{x}_i$ en una celda $c_{\ell(i)}$ de una retícula o *grid*. Asociamos con cada celda c_ℓ un representante $\mathbf{m}_\ell \in \mathbb{R}^d$.



SOM

Imponemos la condición que

- los representantes a celdas cercanas sean similares,
- los datos son similares al representante de su celda.

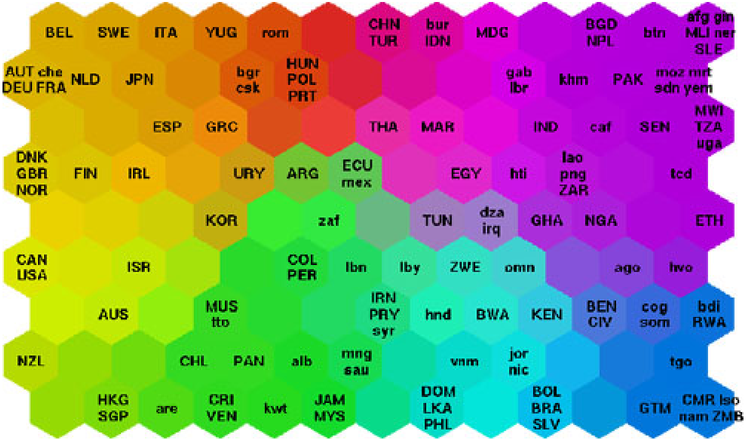
Repetir para cada $\mathbf{x}_i$:

1. Buscar el repreentante más cercano a $\mathbf{x}_i$, denotado como $\mathbf{m}_{\ell(i)}$
2. Para todas las celdas c_k,, actualizamos
$$\mathbf{m}_k = \mathbf{m}_k + \alpha h\big(d(c_k, c_\ell(i))^2\big)||\mathbf{x}_i - \mathbf{m}_k||^2.$$
(*h* es positiva y decreciente, *d* es la distancia en el grid, α es un tamaño de paso decreciente en el tiempo.)

El método minimiza la función de costo
$$J(\{\mathbf{m}_k\}, \{\ell(i)\}) = \sum_{\ell}\sum_{k} h\big(d(c_k, c_\ell(i))^2\big)||\mathbf{x}_i - \mathbf{m}_k||^2.$$

UFM UNIVERSIDAD FRANCISCO MARROQUÍN

SOM



SOM de países sobre 39 indicadores: salud, educación, economía, servicios, ... (Kohonen)

SOM

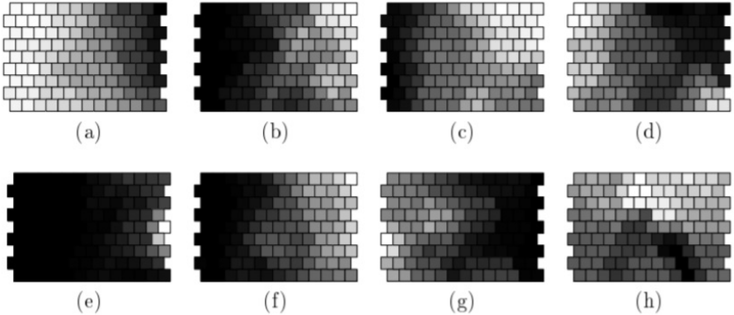
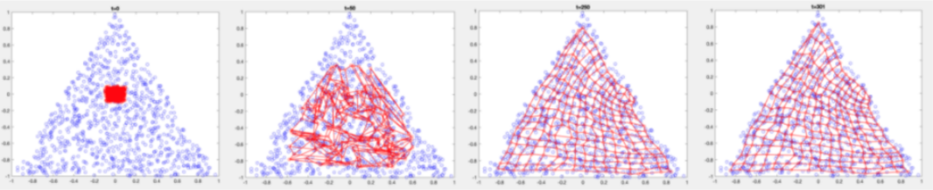
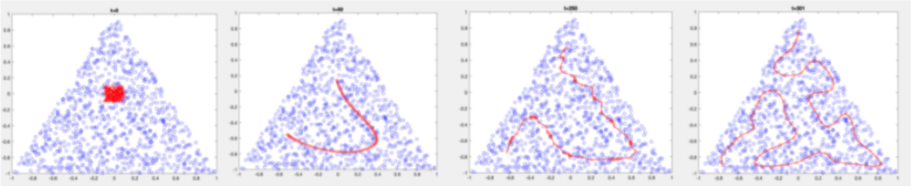


Figure 2: The values of some of the indicators visualized on the SOM groundwork: (a) Life expectancy at birth (years); (b) Adult illiteracy (%); (c) Share of food in household consumption (%); (d) Share of medical care in household consumption (%); (e) Population per physician; (f) Infant mortality rate (per thousand live births); (g) Tertiary education enrollment (% of age group); and (h) Share of the lowest-earning 20 percent in the total household income. In each display, white indicates the largest value and black the smallest, respectively.

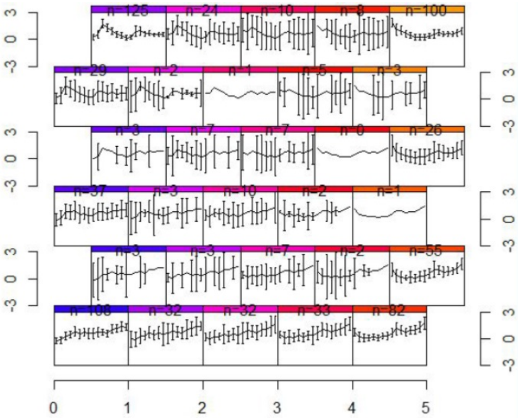
UFM UNIVERSIDAD FRANCISCO MARROQUÍN

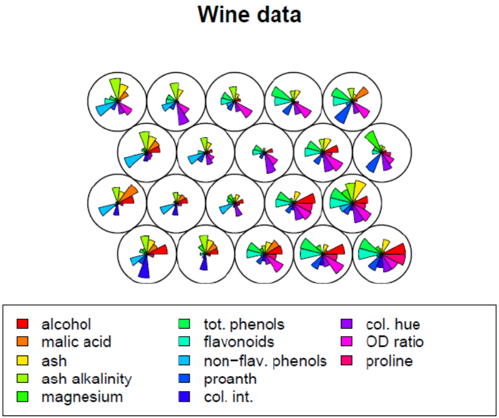
Ejemplo 2 https://towardsdatascience.com/how-to-implement-kohonens-self-organizing-maps-989c4da05f19





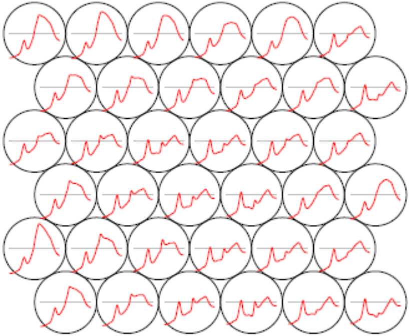
UFM UNIVERSIDAD FRANCISCO MARROQUÍN




Wine data
alcohol
malic acid
ash
ash alkalinity
magnesium
tot. phenols
flavonoids
non−flav. phenols
proanth
col. int.
col. hue
OD ratio
proline

SOM

X



Y