ESCALAMIENTO MULTIDIMENSIONAL (MDS)

Contenido

1.1 Definición

¿Qué es escalamiento multidimensional (MDS)?

MDS es una técnica que trabaja con proximidades entre persona, objetos o estímulos usados para producir una representación espacial de estos ítems.

La **proximidad** expresa la *similaridad* o *disimilaridad* entre los valores que representan a los objetos.

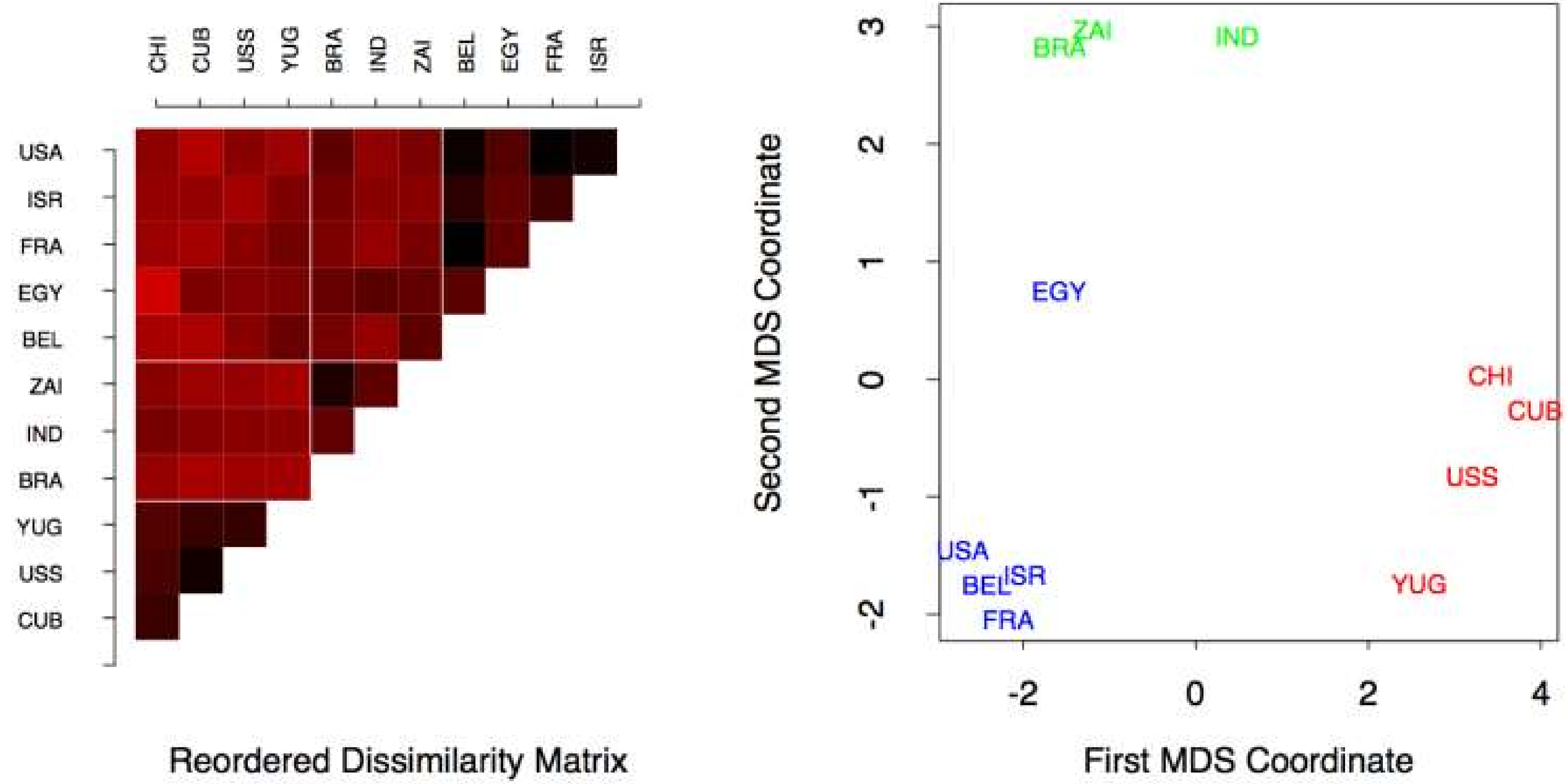
Su objetivo

Es la reducción de la dimensionalidad porque su interés radica en encontrar un conjunto de puntos en un espacio de menor dimensión, por lo general euclidiano, que represente la configuración de los datos de una dimensión superior pero desconocida.

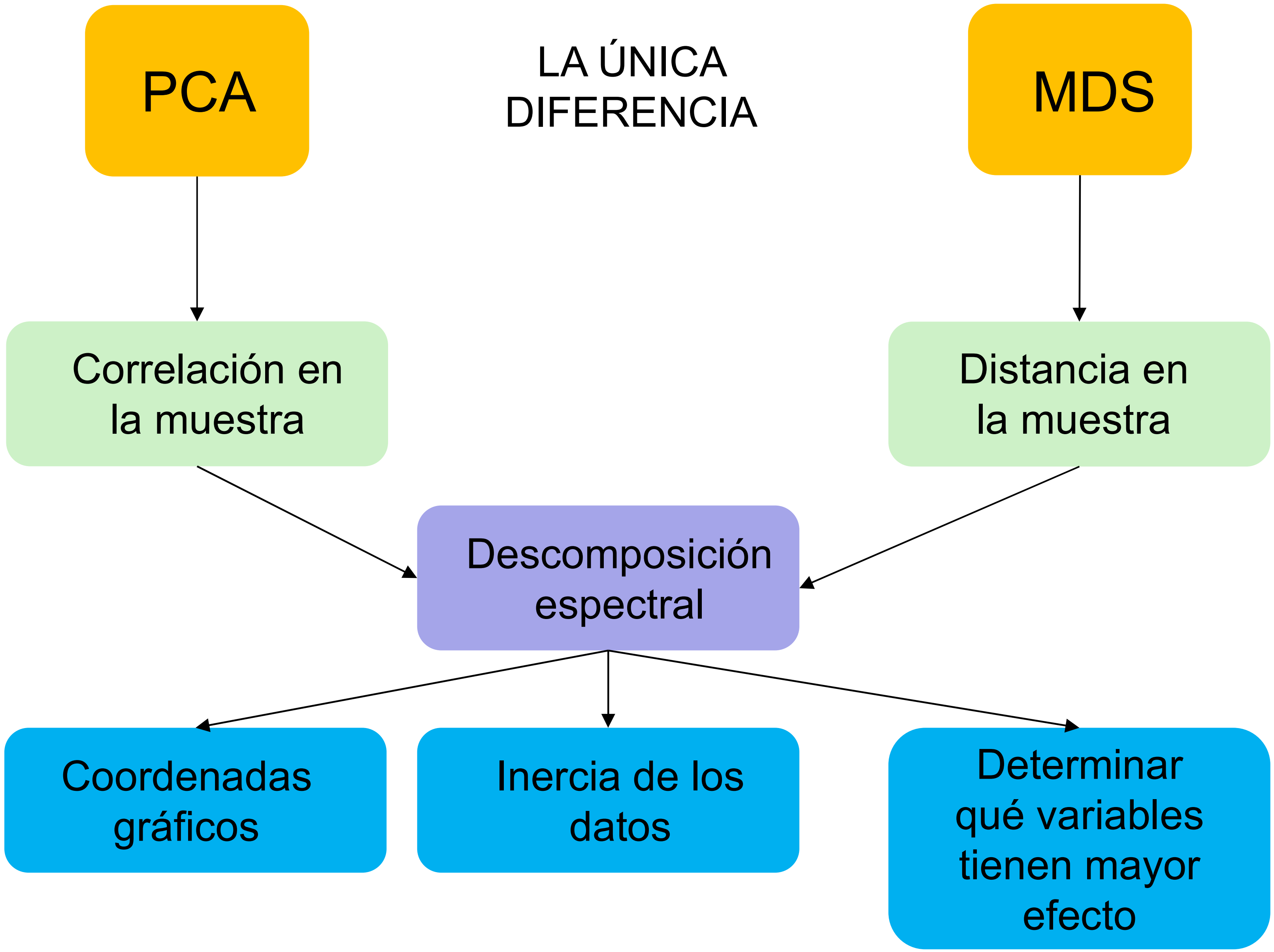
La configuración en alta dimensión es representada por la distancia (d) o matriz de disimilaridad Δ

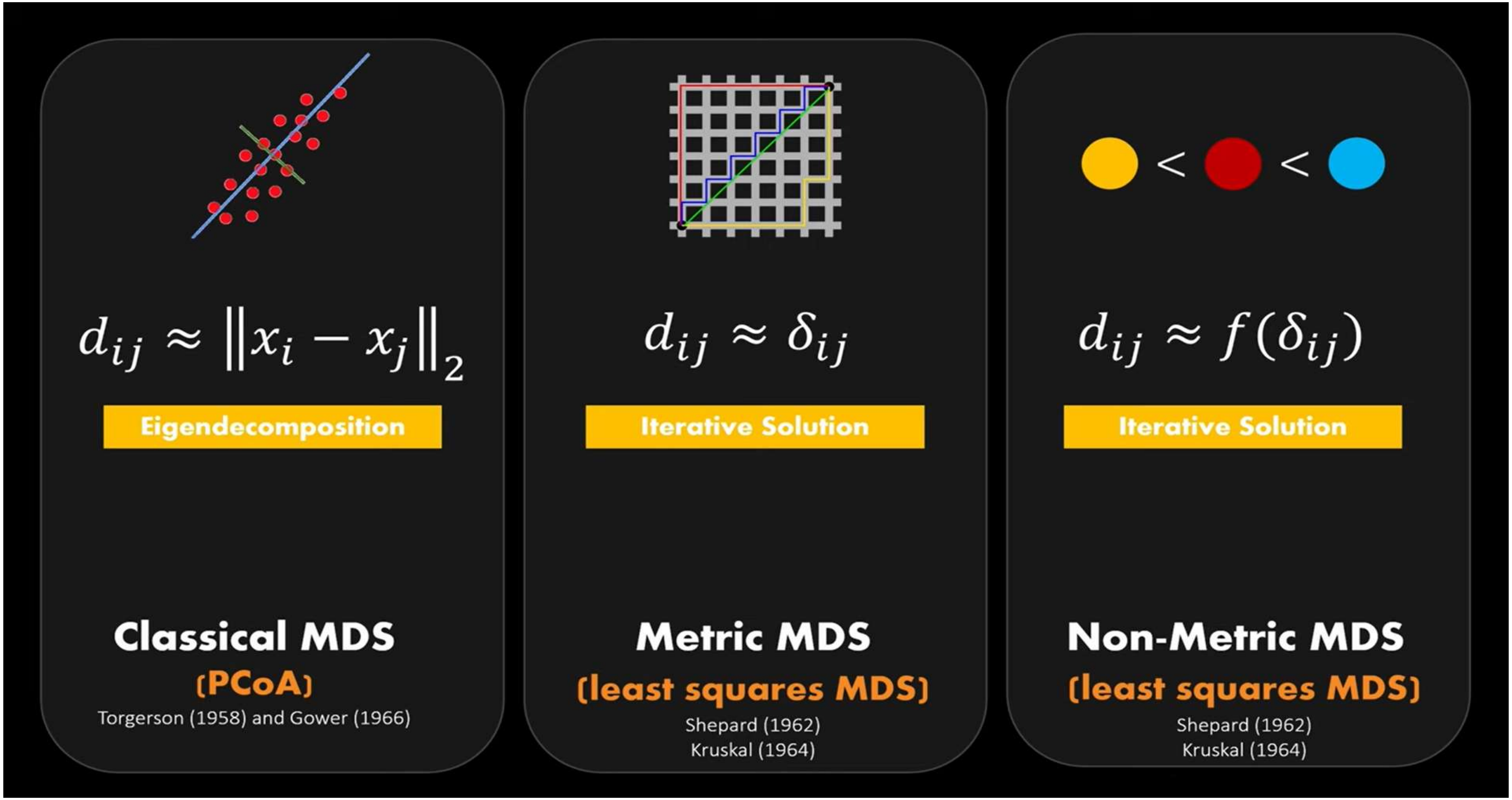
"Dado una disimilaridad (no necesariamente una métrica), reconstruir un mapa que preserve distancias"

Su objetivo



“Dado una disimilaridad (no necesariamente una métrica), reconstruir un mapa que preserve distancias”


PCA
LA ÚNICA DIFERENCIA
MDS
Correlación en la muestra
Distancia en la muestra
Descomposición espectral
Coordenadas gráficos
Inercia de los datos
Determinar qué variables tienen mayor efecto



Principal Coordinates Analysis
Distancias euclideas

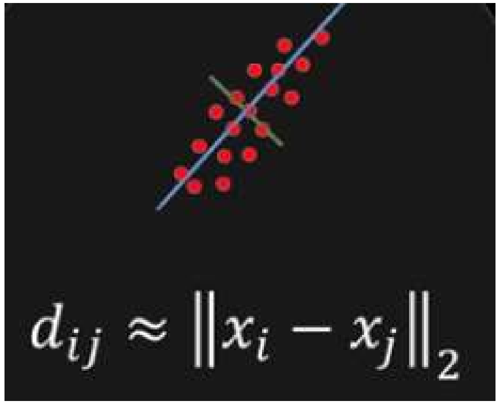
$d_{r1,s1} < d_{r2,s2} < \ldots < d_{rm,sm} \leftrightarrow f(d_{r1,s1}) < f(d_{r2,s2}) < \ldots < f(d_{rm,sm})$

Definiciones básicas del MSD

δ_{ij} $\longrightarrow$ Distancia observada / disimilaridad (en el espacio original)

d_{ij} $\longrightarrow$ Distancia en una configuración de menor dimensión (en el espacio reducido)

$\hat{d}_{ij} = f(\delta_{ij})$ $\longrightarrow$ Disparidad



Implementación del MDS clásico (cMDS)

1. **Matriz B**: Calcular la matriz de doble centralidad o matriz producto interno
2. **Descomposición espectral**: Calcular los m valores estadísticos más grandes
3. **Proyectar**: Construir la matriz de coordenadas finales

1. **Matriz B**: Calcular la matriz de doble centralidad o matriz producto interno

$$\mathbf{B} = \mathbf{HAH}$$ es la matriz de doble centralización *nxn* o matriz producto interno

Donde: $$\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$$ es la matriz de centralización *nxn*

y $\mathbf{A} = \left[a_{rs} \right]$ con $a_{rs} = -\frac{1}{2} d_{rs}^2,$

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)^T (\mathbf{x}_r - \mathbf{x}_s).$$

2. **Descomposición espectral**: Calcular los m valores característicos más grandes

$$B = V \Lambda V'$$

3. **Proyección**: Construir la matriz de coordenadas finales (en un espacio de menor dimensión “m”)

$$X_{(m)} = V_m \sqrt{\Lambda_m}$$

$X_{(m)}$: matriz de puntos (coordenadas)

V_m : matriz construida con las m primeros vectores característicos de V

Λ_m : matriz m-dimensional de valores característicos asociados a V

Definición (Mardia, pp.397)

Una matriz de distancia **D** se le dice euclideana si existe una configuración de puntos en algún espacio euclideo cuya distancia interpuntos están dadas por **D**; esto es, si para algún **p**, existen puntos $x_1, \ldots, x_n \in R^p$ tal que:

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)^T(\mathbf{x}_r - \mathbf{x}_s).$$

Teorema 14.2.1 (Mardia)

Dado una matriz de distancia **D** y de doble centralización **B**, entonces, **D** es euclideana si y solo si **B**≥0. En particular, los siguientes resultados se confirman:

Nota.-

Si **D** es la matriz de distancias euclideanas interpuntos para una configuración **z**= $(x_1, x_2, \ldots, x_n)$, entonces,

$$\mathbf{b}_{rs} = (x_r - \overline{x})^t (x_s - \overline{x})$$

En notación matricial $\mathbf{B} = (\mathbf{H}X)(\mathbf{H}X)^t$ tal que $\mathbf{B} \geq 0$.

Prueba:

Los elementos de la matriz **D=** d_{rs} son distancias euclideas, si y solo si, la matriz de doble centralidad **B=HAH** es definida no negativa.

⇨) Hipótesis d^2_{rs} es una distancia euclidea

$$\mathbf{A} = \left(a_{rs} = -\tfrac{1}{2} d^2_{rs}, \right) \quad \text{y} \quad \mathbf{B=HAH}$$

Tesis **B** es semidefinida positiva

⇨ Como $\mathbf{B} = \mathbf{HAH} = (\mathbf{I}_n - 1/n\, \mathbf{J}_n)\mathbf{A}(\mathbf{I}_n - 1/n\, \mathbf{J}_n)$

$$b_{rs} = a_{rs} - \bar{a}_{.s} - \bar{a}_{r.} + \bar{a}_{..} \quad \text{.................. (i)}$$

Pero p.h. $a_{rs} = -\frac{1}{2} d^2_{rs} = -1/2\, (x_r - x_s)^t(x_r - x_s)$

$$= -\,(1/2)x_r^t x_r - (1/2)x_s^t x_s + x_r^t x_s \quad \text{.................. en (i)}$$

⇨ $b_{rs} = (x_r - \overline{x})^t(x_s - \overline{x})$

⇨ $\mathbf{B} = \left(b_{rs} \right) = (\mathbf{H}x)(\mathbf{H}x)^t = \mathbf{ZZ}^t \geq 0,\ \forall Z$

⇦) Hipótesis **B** es semidefinida positiva

$\mathbf{A} = \left(a_{rs} = -\frac{1}{2} d_{rs}^2, \right)$ y **B=HAH**

Tesis d^2_{rs} es una distancia euclidea

⇨ a) simetría

p.h. $\mathbf{B} = \mathbf{ZZ}^t$ ⇨ $\mathbf{B}^t = \mathbf{ZZ}^t = \mathbf{B}$ luego, **B** es simétrica

⇨ p.h. **B** = **HAH**, entonces, **A** es simétrica.

Además, $a_{rs} = -\frac{1}{2}d_{rs}^2,$ ⇨ $d^2_{rs} = -2a_{rs} = -2a_{sr} = d^2_{sr}$

b) no negatividad

$d^2_{rs} = \mathbf{d}^t\mathbf{d} > 0$ solo si $d \neq 0$. En particular, $d = xr - xs \neq 0$

⇨ $x_r \neq x_s$

c) identidad

$d^2_{rs} = 0$ pero $d^2_{rs} = \mathbf{d}^t\mathbf{d}$ es igual a cero solo si $\mathbf{d}=0$

⇨ $x_r = x_s$

De a), b) y c) concluimos que $\mathbf{d}^2_{rs}$ es una distancia euclidea (…)

Resultado a) $\mathbf{b}_{rs} = (x_r - \overline{x})^t(x_s - \overline{x})$

Aquí asumimos que la configuración es **X**

$\Rightarrow$ $\mathbf{B} = (\mathbf{X} - 1_n\overline{x})^t(\mathbf{X} - 1_n\overline{x})$

$\mathbf{B} = (\mathbf{HX})(\mathbf{HX})^T = \mathbf{WW}^t \geq 0$

Resultado b) $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_p),\quad \mathbf{V}_1 = [\mathbf{v}_1, \ldots, \mathbf{v}_p]$ dado que: r(**B**)=**p**

$\Rightarrow$ $\boldsymbol{\Lambda}_1^{\frac{1}{2}} = \operatorname{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_p^{\frac{1}{2}})$

$\Rightarrow$ $\mathbf{X} = \mathbf{V}_1 \boldsymbol{\Lambda}_1^{\frac{1}{2}}$

$\mathbf{B} = \mathbf{V}_1 \boldsymbol{\Lambda}_1 \mathbf{V}_1^T$ es la descomposición espectral de **B**

$\Rightarrow$ $\mathbf{B} = \mathbf{X}\mathbf{X}^T$

O también: $[\mathbf{B}]_{rs} = b_{rs} = \mathbf{x}_r^T \mathbf{x}_s.$

Matriz producto interno de la configuración **X**

Se sabe que **d** es una distancia euclidea:

$$d_{rs}^2 = \mathbf{x}_r^T \mathbf{x}_r + \mathbf{x}_s^T \mathbf{x}_s - 2\mathbf{x}_r^T \mathbf{x}_s$$

Aplicando sumatorias:

$$\Rightarrow \quad \frac{1}{n}\sum_{r=1}^{n} d_{rs}^2 = \frac{1}{n}\sum_{r=1}^{n} \mathbf{x}_r^T \mathbf{x}_r + \mathbf{x}_s^T \mathbf{x}_s,$$

$$\frac{1}{n}\sum_{s=1}^{n} d_{rs}^2 = \mathbf{x}_r^T \mathbf{x}_r + \frac{1}{n}\sum_{s=1}^{n} \mathbf{x}_s^T \mathbf{x}_s,$$

$$\frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n} d_{rs}^2 = \frac{2}{n}\sum_{r=1}^{n} \mathbf{x}_r^T \mathbf{x}_r.$$

Asumiendo que: $$\sum_{r=1}^{n} x_{ri} = 0 \qquad (i = 1, \ldots, p)$$

Porque la configuración **X** está centrada en **0**

Se sabe que:

$$\begin{aligned} b_{rs} &= \mathbf{x}_r^T \mathbf{x}_s, \\ &= -\frac{1}{2}\left(d_{rs}^2 - \frac{1}{n}\sum_{r=1}^{n} d_{rs}^2 - \frac{1}{n}\sum_{s=1}^{n} d_{rs}^2 + \frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n} d_{rs}^2 \right) \\ &= a_{rs} - \overline{a}_{r.} - \overline{a}_{.s} + \overline{a}_{..}, \end{aligned}$$

Donde:

$$a_{rs} = -\tfrac{1}{2} d_{rs}^2,$$

$$\overline{a}_{r.} = n^{-1}\sum_s a_{rs}, \quad \overline{a}_{.s} = n^{-1}\sum_r a_{rs}, \quad \overline{a}_{..} = n^{-2}\sum_r\sum_s a_{rs}$$

Expresando **B** en términos de **A**:

$$
\begin{aligned}
(\mathbf{x}_r - \mathbf{x}_s)^T(\mathbf{x}_r - \mathbf{x}_s) &= \mathbf{x}_r^T\mathbf{x}_r + \mathbf{x}_s^T\mathbf{x}_s - 2\mathbf{x}_r^T\mathbf{x}_s \\
&= b_{rr} + b_{ss} - 2b_{rs} \\
&= a_{rr} + a_{ss} - 2a_{rs} \\
&= -2a_{rs} = \delta_{rs}^2,
\end{aligned}
$$

Dado que : $\mathbf{B1} = \mathbf{HAH1} = \mathbf{0}$. ,entonces, **1** es un vector característico cuyo valor característico asociado es **0**.

⇨ El vector **1** es ortogonal a las columnas de **X** dado que $\mathbf{B} = (\mathbf{HX})(\mathbf{HX})^T$

(…)

EJEMPLO.-
Disparidad ($\hat{d}$) entre 5 objetos

$$
\begin{array}{c} \\ O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \end{array}
\begin{array}{c}
\begin{array}{ccccc} O_1 & O_2 & O_3 & O_4 & O_5 \end{array} \\
\begin{pmatrix}
0.00 & 2.37 & 1.05 & 2.34 & 2.34 \\
2.37 & 0.00 & 2.34 & 2.81 & 2.81 \\
1.05 & 2.34 & 0.00 & 2.42 & 2.42 \\
3.54 & 3.63 & 3.50 & 0.00 & 4.24 \\
2.34 & 2.81 & 2.42 & 4.24 & 0.00
\end{pmatrix}
\end{array}
$$

- ✓ No se conocen los datos originales
- ✓ No se sabe cómo se determinaron las disparidades (δ_{ij})
- ✓ Aplicamos el teorema 14.2.1 (calculamos **B**)
- ✓ Si son datos métricos, entonces,
- ✓ Aplicamos algún procedimiento MDS

Para calcular la matriz **B** hacemos δ^2_{ij} $\forall$i,j y se obtiene la siguiente matriz:

$$\begin{pmatrix} 0.00 & 5.62 & 1.11 & 12.59 & 5.48 \\ 5.62 & 0.00 & 5.50 & 13.20 & 7.94 \\ 1.10 & 5.50 & 0.00 & 12.31 & 5.87 \\ 12.59 & 13.20 & 12.31 & 0.00 & 18.03 \\ 5.48 & 7.94 & 5.87 & 18.03 & 0.00 \end{pmatrix}$$

Dado que $a_{ij} = -\frac{1}{2}\delta^2_{ij}$

$$A = -\frac{1}{2}\begin{pmatrix} 0.00 & 5.62 & 1.11 & 12.59 & 5.48 \\ 5.62 & 0.00 & 5.50 & 13.20 & 7.94 \\ 1.10 & 5.50 & 0.00 & 12.31 & 5.87 \\ 12.59 & 13.20 & 12.31 & 0.00 & 18.03 \\ 5.48 & 7.94 & 5.87 & 18.03 & 0.00 \end{pmatrix}$$

Y dado que: $b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}$

Obtenemos:

$$B = \begin{pmatrix} 1.454 & -0.610 & 0.898 & -1.708 & -0.034 \\ -0.610 & 2.946 & -0.551 & -1.267 & -0.518 \\ 0.898 & -0.551 & 1.452 & -1.569 & -0.230 \\ -1.708 & -1.267 & -1.569 & 7.720 & -3.176 \\ -0.034 & -0.518 & -0.230 & -3.176 & 3.958 \end{pmatrix}$$

Cuyos valores característicos son:

10.0895   3.6366   3.2546   0.5493   0.0000

Entonces, **B** ≥ 0, lo que significa que la matriz original de disimilaridades es euclideana

Datos

$$\mathbf{X} = \begin{pmatrix} 3 & 4 & 4 & 6 & 1 \\ 5 & 1 & 1 & 7 & 3 \\ 6 & 2 & 0 & 2 & 6 \\ 1 & 1 & 1 & 0 & 3 \\ 4 & 7 & 3 & 6 & 2 \\ 2 & 2 & 5 & 1 & 0 \\ 0 & 4 & 1 & 1 & 1 \\ 0 & 6 & 4 & 3 & 5 \\ 7 & 6 & 5 & 1 & 4 \\ 2 & 1 & 4 & 3 & 1 \end{pmatrix}$$

```
# Cargar datos
> X = read.delim("clipboard")
> dim(X)
[1] 10 5
```

```
> D <-dist(X)
> D
          1        2        3        4        5        6        7        8        9
2  5.196152
3  8.366600 6.082763
4  7.874008 8.062258 6.324555
5  3.464102 6.557439 8.366600 9.273618
6  5.656854 8.426150 8.831761 5.291503 7.874008
7  6.557439 8.602325 8.185353 3.872983 7.416198 5.000000
8  6.164414 8.888194 8.366600 6.928203 6.000000 7.071068 5.744563
9  7.416198 9.055385 6.855655 8.888194 6.557439 7.549834 8.831761 7.416198
10 4.358899 6.164414 7.681146 4.795832 7.141428 2.645751 5.099020 6.708204 8.000000
```

cmdscale()
cmdscale(dist(X), k=5)
cmdscale(D,k=5)

k = 2, 3

```
> cmdscale(dist(X),k=5)
            [,1]        [,2]       [,3]       [,4]
 [1,] -1.6038325  2.38060903 -2.2301092 -0.3656856
 [2,] -2.8246377 -2.30937202 -3.9523782  0.3419185
 [3,] -1.6908272 -5.13970089  1.2880306  0.6503227
 [4,]  3.9527719 -2.43233961  0.3833746  0.6863995
 [5,] -3.5984894  2.75538195 -0.2551393  1.0783741
 [6,]  2.9520356  1.35475175 -0.1899027 -2.8211220
 [7,]  3.4689928  0.76411068  0.3016531  1.6369166
 [8,]  0.3545235  2.31408566  2.2161772  2.9240116
 [9,] -2.9362323 -0.01279597  4.3117385 -2.5122743
[10,]  1.9256952  0.32526941 -1.8734445 -1.6188611
             [,5]
 [1,]  0.11536476
 [2,]  0.33169405
 [3,] -0.05133897
 [4,] -0.03460933
 [5,] -1.26125237
 [6,]  0.12385813
 [7,] -1.94209512
 [8,]  2.00450379
 [9,] -0.18911558
[10,]  0.90299062
```

distancia MDS

minimizar $\phi = \sum_{r,s}^{n} (d_{rs}^2 - \hat{d}_{rs}^2)$

distancia euclidea

```
> dist(X)-dist(cmdscale(dist(X),k=5))
```

O también

```
> D-dist(cmdscale(D,k=5))
```

donde **D** es la matriz de distancias euclideas

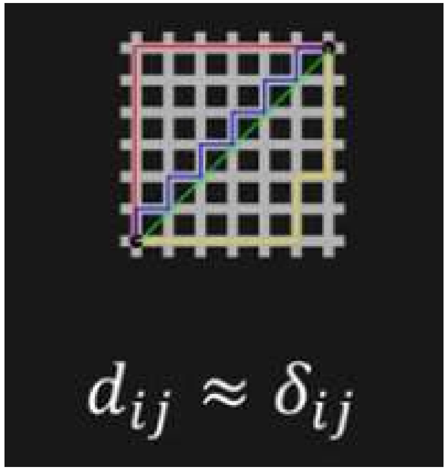
⇨ $\Phi \approx 0$

Comprobación haciendo r=1 y s=2 en la configuración de MDS $\widehat{X}$:

$\widehat{O}_1$	1.6	2.38	2.23	-0.37	0.12
$\widehat{O}_2$	2.82	-2.31	3.95	0.34	0.33
$\widehat{x}_{1j} - \widehat{x}_{2j} \rightarrow$	-1.22	4.69	-1.72	-0.71	-0.21
$(\widehat{x}_{1j} - \widehat{x}_{2j})^2 \rightarrow$	1.4884	21.9961	2.9584	0.5041	0.0441

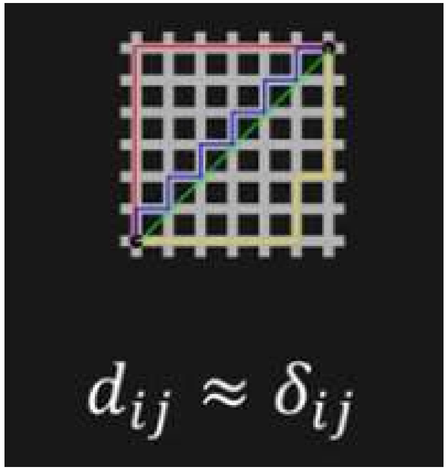
$\Rightarrow$ $\widehat{d}^2_{12} =$ 5.19529595 distancia euclidea entre $\mathbf{O_1}$ y $\mathbf{O_2}$ en el espacio de menor dimensión

$d^2_{12} =$ 5.196152 distancia euclidea real entre $\mathbf{O_1}$ y $\mathbf{O_2}$



Implementación del MDS métrico (mMDS)

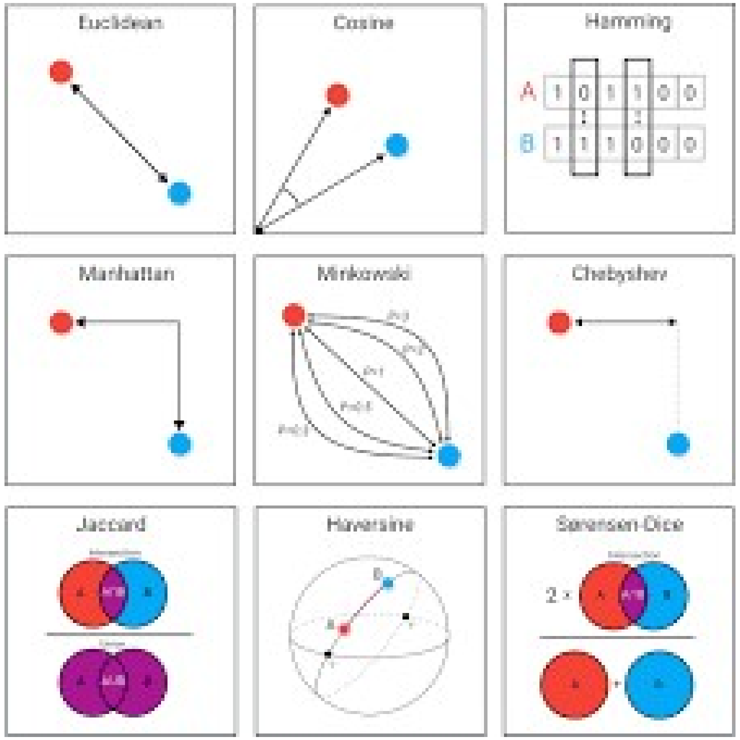
Si existen *n* objetos con disimilaridades { δ_{ij} } el MDS métrico (*mMDS*) intenta encontrar un conjunto de puntos en un espacio donde cada punto representa a un objeto y las distancias entre puntos $\{d_{rs}\}$ son tal que: $d_{rs} \approx f(\delta_{ij})$ donde *f* es una función paramétrica monotónica continua.

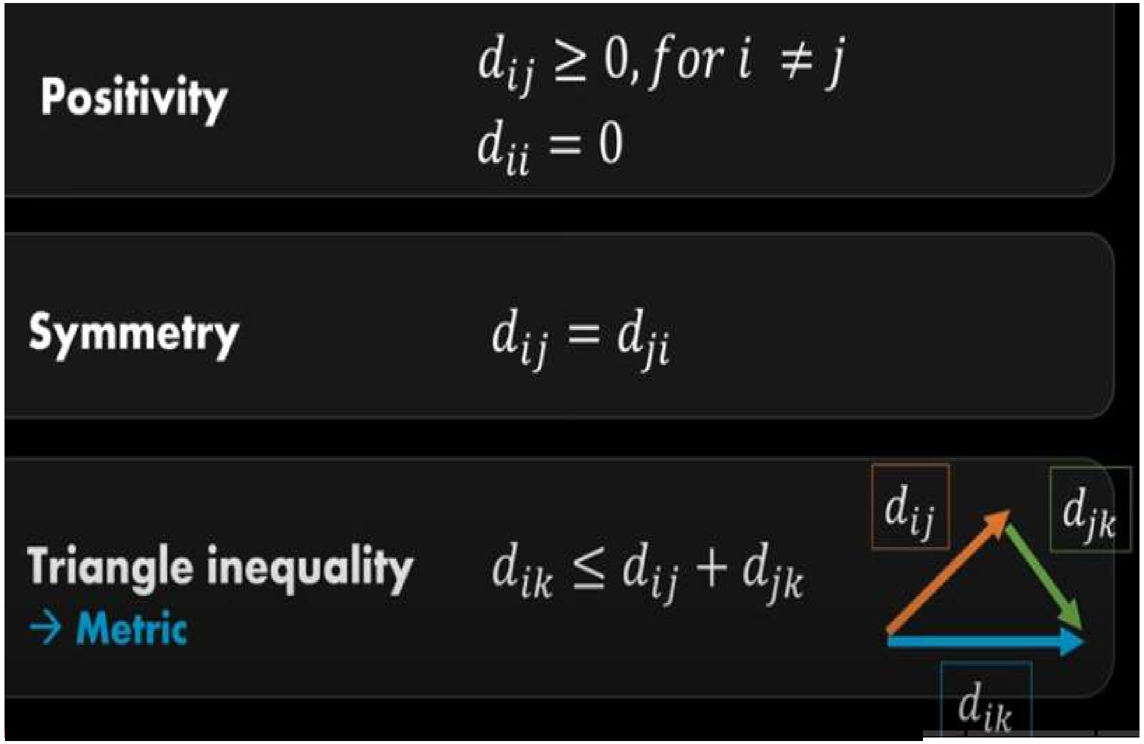


Implementación del MSD métrico (mMDS)

1. **Matriz de distancias**: La d_{ij} es una métrica (escala de proporción)
2. **Solución iterativa**: inicialización, cálculo de distancias, calculo de la perdida y optimización para minimizar *stress*
3. **Pérdida**: **S**tandardized **R**esidual **S**um of **S**quares (*STRESS*)

1. **Matriz de distancias**: Tratamiento más general de distancias. La métrica.





Medidas de disimilaridad para datos cuantitativos

1. $d(x, y) \geq 0,$
2. $d(x, y) = 0$ if and only if $x = y,$
3. $d(x, y) = d(y, x),$
4. $d(x, z) \leq d(x, y) + d(y, z).$

Euclidean distance	$\delta_{rs} = \left\{ \sum_i (x_{ri} - x_{si})^2 \right\}^{\frac{1}{2}}$
Weighted Euclidean	$\delta_{rs} = \left\{ \sum_i w_i (x_{ri} - x_{si})^2 \right\}^{\frac{1}{2}}$
Mahalanobis distance	$\delta_{rs} = \{(\mathbf{x}_r - \mathbf{x}_s)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_r - \mathbf{x}_s)\}^{\frac{1}{2}}$
City block metric	$\delta_{rs} = \sum_i \lvert x_{ri} - x_{si} \rvert$
Minkowski metric	$\delta_{rs} = \left\{ \sum_i w_i \lvert x_{ri} - x_{si} \rvert^{\lambda} \right\}^{\frac{1}{\lambda}} \quad \lambda \geq 1$
Canberra metric	$\delta_{rs} = \sum_i \lvert x_{ri} - x_{si} \rvert / (x_{ri} + x_{si})$
Divergence	$\delta_{rs} = \frac{1}{p} \sum_i (x_{ri} - x_{si})^2 / (x_{ri} + x_{si})^2$
Bray-Curtis	$\delta_{rs} = \frac{1}{p} \frac{\sum_i \lvert x_{ri} - x_{si} \rvert}{\sum_i (x_{ri} + x_{si})}$
Soergel	$\delta_{rs} = \frac{\sum_i \lvert x_{ri} - x_{si} \rvert}{\sum_i \max(x_{ri}, x_{si})}$
Bhattacharyya distance	$\delta_{rs} = \left\{ \sum_i (x_{ri}^{\frac{1}{2}} - x_{si}^{\frac{1}{2}})^2 \right\}^{\frac{1}{2}}$
Wave-Hedges	$\delta_{rs} = \frac{1}{p} \sum_i \left(1 - \frac{\min(x_{ri}, x_{si})}{\max(x_{ri}, x_{si})}\right)$
Angular separation	$\delta_{rs} = 1 - \frac{\sum_i x_{ri} x_{si}}{[\sum_i x_{ri}^2 \sum_i x_{si}^2]^{\frac{1}{2}}}$
Correlation	$\delta_{rs} = 1 - \frac{\sum_i (x_{ri} - \bar{x}_r)(x_{si} - \bar{x}_s)}{\left\{ \sum_i (x_{ri} - \bar{x}_r)^2 \sum_i (x_{si} - \bar{x}_s)^2 \right\}^{\frac{1}{2}}}$

Datos binarios

		Object s 1	0	
Object r	1	a	b	$a+b$
	0	c	d	$c+d$
		$a+c$	$b+d$	$p=a+b+c+d$

Coeficientes de similaridad para datos binarios

Braun, Blanque	$s_{rs} = \dfrac{a}{\max\{(a+b),(a+c)\}}$
Czekanowski, Sørensen, Dice	$s_{rs} = \dfrac{2a}{2a+b+c}$
Hamman	$s_{rs} = \dfrac{a-(b+c)+d}{a+b+c+d}$
Jaccard coefficient	$s_{rs} = \dfrac{a}{a+b+c}$
Kulczynski	$s_{rs} = \dfrac{a}{b+c}$
Kulczynski	$s_{rs} = \dfrac{1}{2}\left(\dfrac{a}{a+b}+\dfrac{a}{a+c}\right)$
Michael	$s_{rs} = \dfrac{4(ad-bc)}{\{(a+d)^2+(b+c)^2\}}$
Mountford	$s_{rs} = \dfrac{2a}{a(b+c)+2bc}$
Mozley, Margalef	$s_{rs} = \dfrac{a(a+b+c+d)}{(a+b)(a+c)}$
Ochiai	$s_{rs} = \dfrac{a}{[(a+b)(a+c)]^{\frac{1}{2}}}$
Phi	$s_{rs} = \dfrac{ad-bc}{[(a+b)(a+c)(b+d)(c+d)]^{\frac{1}{2}}}$
Rogers, Tanimoto	$s_{rs} = \dfrac{a+d}{a+2b+2c+d}$
Russell, Rao	$s_{rs} = \dfrac{a}{a+b+c+d}$
Simple matching coefficient	$s_{rs} = \dfrac{a+d}{a+b+c+d}$
Simpson	$s_{rs} = \dfrac{a}{\min\{(a+b),(a+c)\}}$
Sokal, Sneath, Anderberg	$s_{rs} = \dfrac{a}{a+2(b+c)}$
Yule	$s_{rs} = \dfrac{ad-bc}{ad+bc}$

2. **Solución iterativa**: familia de algoritmos diseñados para alcanzar una configuración óptima de baja dimensión.

Pasos

1. **Inicialización**: Se inicializa con puntos en posiciones aleatorias
2. **Cálculo de distancias**: Se obtiene la matriz de distancias para la configuración
3. **Cálculo de la pérdida** (*loss*): Evaluar la función *stress*
4. **Optimizar**: “descenso por gradiente” para actualizar el *stress* minimizado

2. **Cálculo de distancias**: Se obtiene la matriz de distancias para la configuración

Una matriz **D**(nxn) se llama matriz distancia si es simétrica y

$$d_{rr} = 0, \quad d_{rs}>0, \quad \text{si } r \neq s$$

tal que si $\mathbf{d}_{rs}$ denota una distancia euclidea entre $\mathbf{P}_r$ y $\mathbf{P}_s$,

3. **Cálculo de la pérdida** (*loss*): Evaluar la función *stress*

Standardized **R**esidual **S**um of **S**quares (*STRESS*)

Dado una dimensión p y una función monótona f(), entonces, mMDS trata de encontrar una configuración óptima $X \subset R^p$ tal que:

$$f(d_{ij}) \approx \hat{d}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$$

y

$$f(d_{ij}) = Ad_{ij} + b$$

De esta forma, **P**$_1$, …, **P**$_n$ con coordenadas $x_i' = (x_{i1}, \dots, x_{ip})$, i=1, …, n representan una solución *mMDS* p-dimensional

Es óptima en la medida que explícitamente se cumple:

$$\text{stress} = \mathcal{L}(\hat{d}_{ij}) = \left(\sum_{i<j} (\hat{d}_{ij} - f(d_{ij}))^2 / \sum d_{ij}^2 \right)^{\frac{1}{2}}$$

y la métrica MDS minimiza $\mathcal{L}(\hat{d}_{ij})$ sobre todos los $\hat{d}_{ij}$ y ***A***, ***b***.

La métrica usual del *mMDS* es el caso especial donde $f(d_{ij}) = d_{ij}$

"La solución usual *mMDS* (por optimización) ≠ a la solución *cMDS*"

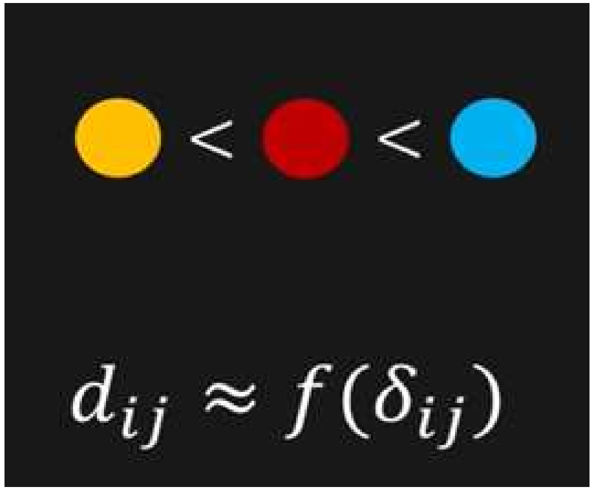
Diagrama de Shepard

Stress	Goodness of fit
20%	poor
10%	fair
5%	good
$2\frac{1}{2}$%	excellent
0%	"perfect"

4. **Optimizar**: “descenso por gradiente” o para actualizar el *stress* minimizado

Gradient Descent

SMACOF
Scaling by Majorizing a Complicated Fuction



Implementación del MSD no métrico (nMDS)

Ordinal – función monotónica – disparidad – preserva distancia – regresión isotónica – algoritmo iterativo

No métrico

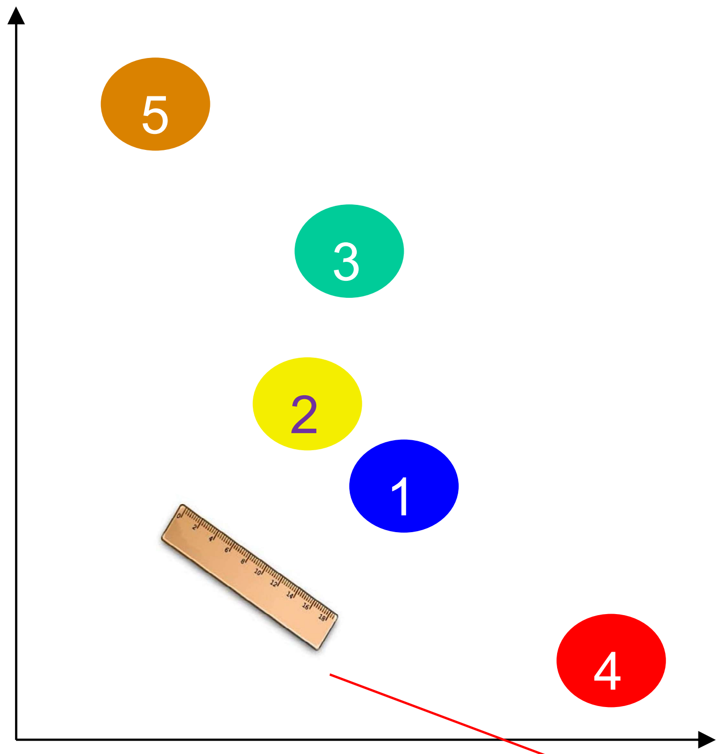
$d_{ij}=0$

$d_{jk}=0$

$d_{ik}=1$

$$d_{ik} \leq d_{ij} + d_{jk}$$

Ordinal MDS

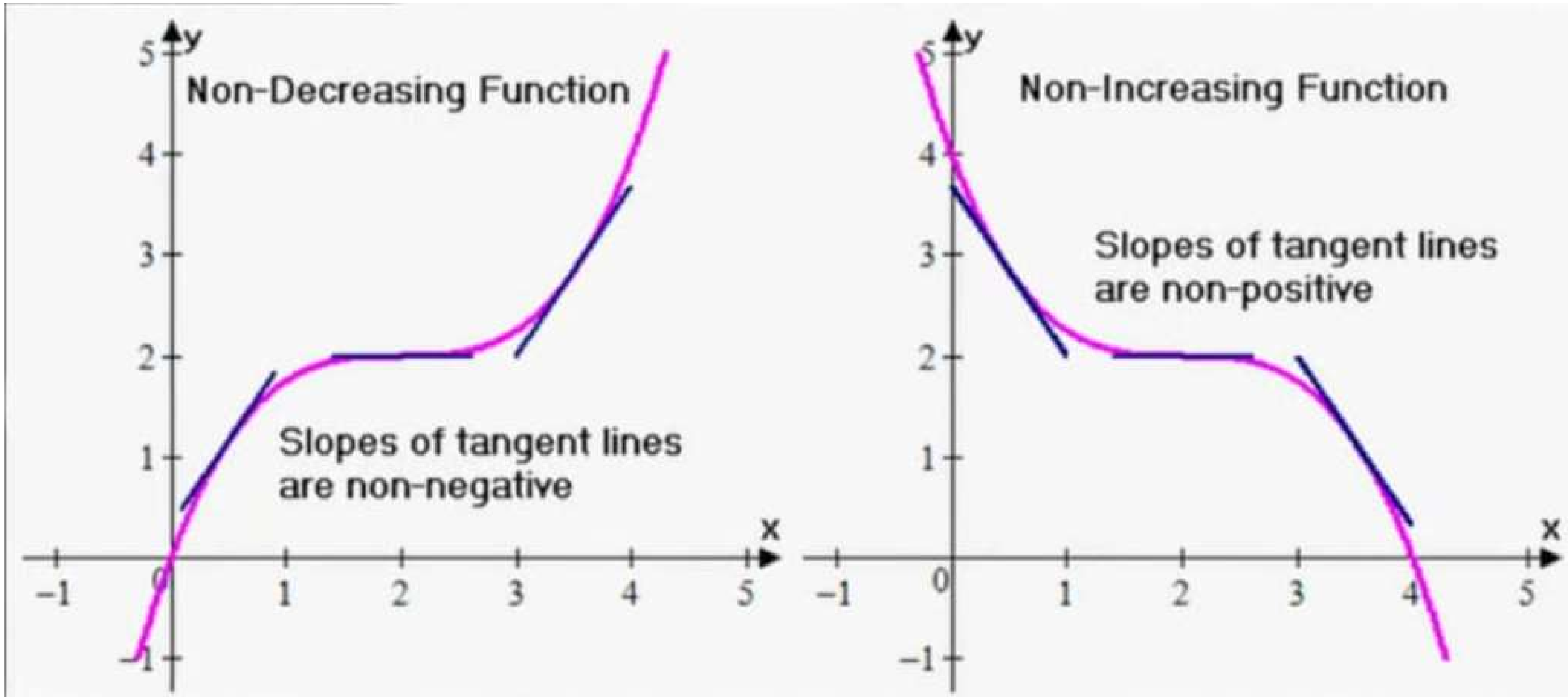


	695	760	6264	10160
		545	5563	10538
			5567	9997
				7420

$$\delta_{12} < \delta_{23} < \delta_{13} < \ldots$$

¿Cómo aseguramos que en el espacio de menos dimensión se preservan las distancias reales?

Función monotónica creciente
o decreciente



$$\mathbf{d_{ij} \approx f(\delta_{12}) = \hat{d}_{ij}}$$

f: función monotónica que mapea las disparidades

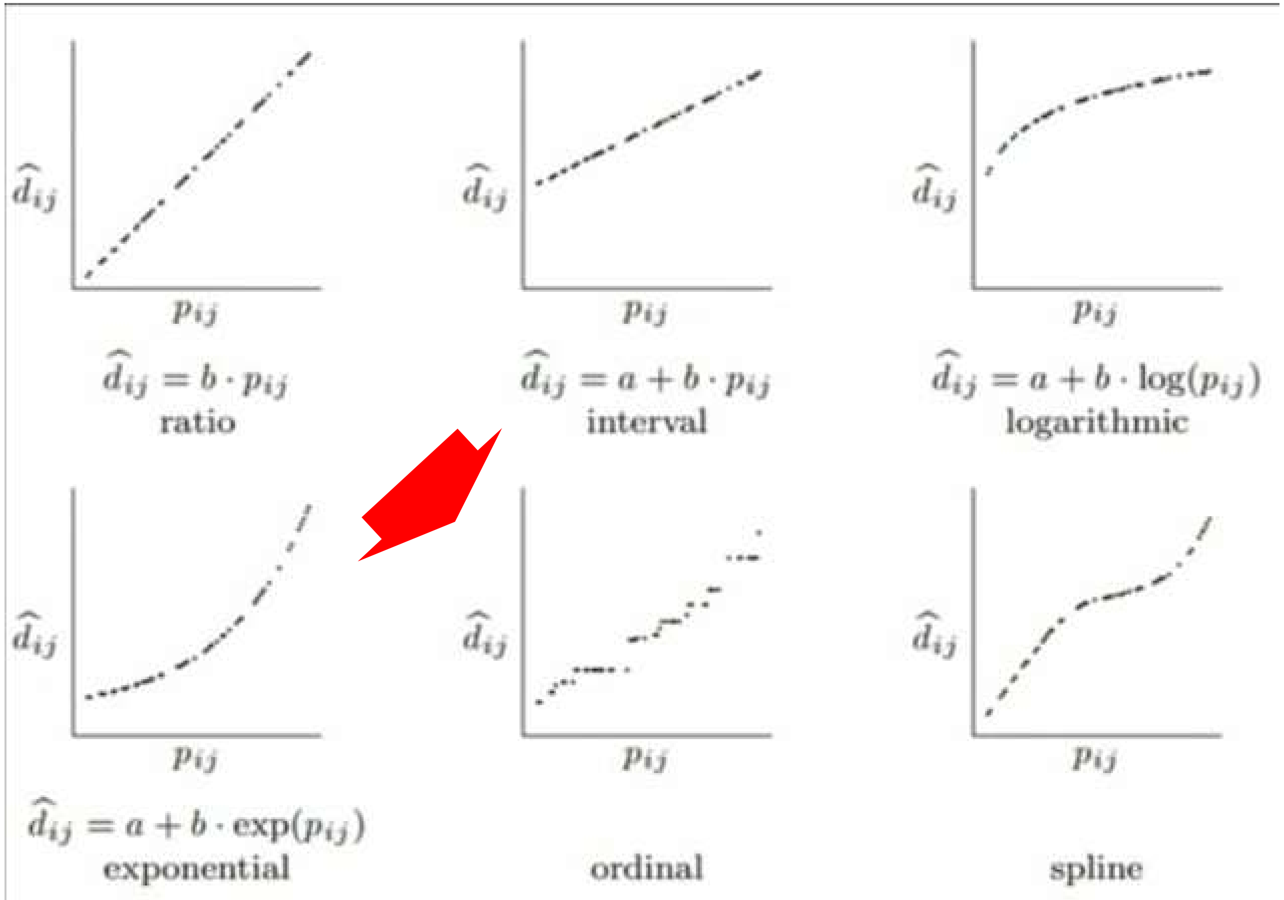
$\hat{d}$: disparidades

$$\mathbf{\delta_{ij} < \delta_{kl}}$$

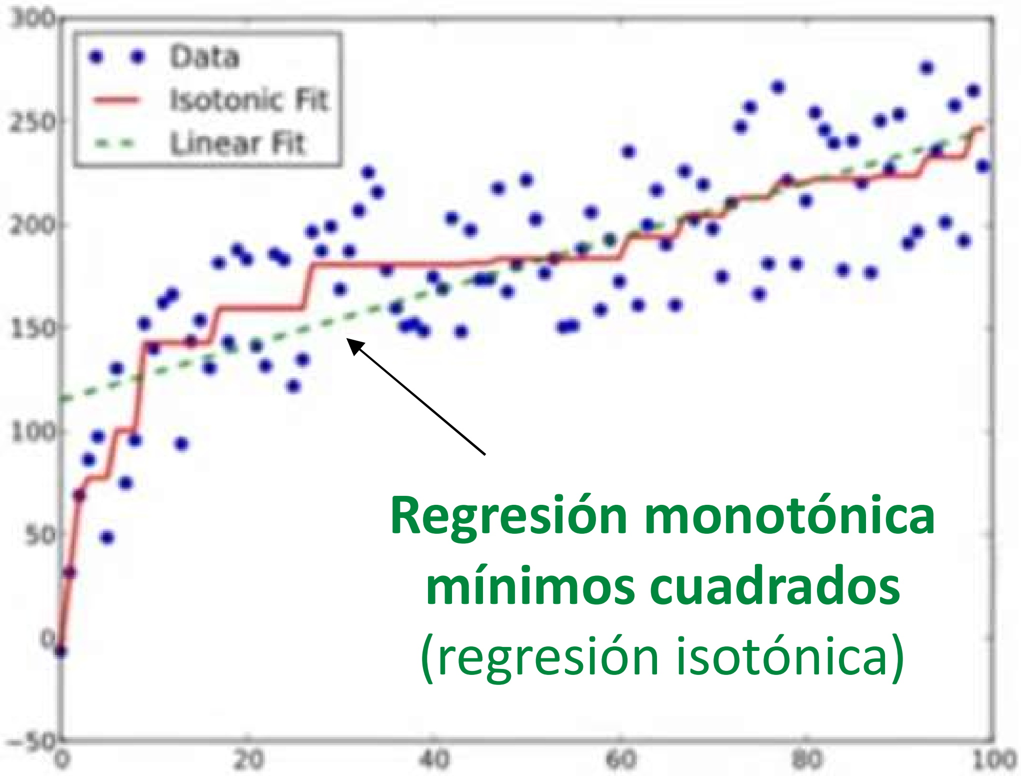
Preserva el orden

$$\mathbf{f(\delta_{ij}) < f(\delta_{kl})}$$

Funciones monotónicas



Función monotónica



Algoritmo iterativo: algoritmo de infractores adyacentes al grupo

Ajuste isotónico / ajuste lineal

Predicho por la regresión monotónica

$$Stress = \left(\frac{\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2} \right)^{1/2}$$

Caso: disimilaridades ordinales

A un panel de personas que tienen cierto conocimiento sobre la distribución física de las principales ciudades de una región se les nombra un par de ciudades y deben indicar si este par de ciudades ´que tan cerca o lejos están. Para ello, se les proporciona una escala de calificación del **1** (muy cercanos) al **9** (muy lejanos).

La distancia definida no cumple las propiedades de distancia ***(requiere un tratamiento ordinal)***

Disimilaridades entre ciudades

0	1	5	2	7	2	2	7	8	1	ATLANTA
1	0	3	3	7	5	2	7	6	1	CHICAGO
5	3	0	3	3	6	6	4	4	6	DENVER
2	3	3	0	5	4	5	6	7	5	HOUSTON
7	7	3	5	0	8	9	1	4	8	LOS ANGELES
2	5	6	4	8	0	4	9	9	3	MIAMI
2	2	6	5	9	4	0	9	8	1	NEW YORK
7	7	4	6	1	9	9	0	2	9	SAN FRANCISCO
8	6	4	7	4	9	8	2	0	8	SEATTLE
1	1	6	5	8	3	1	9	8	0	WASHINGTON, DC

Cantidades mayores, mayor disimilaridad y viceversa

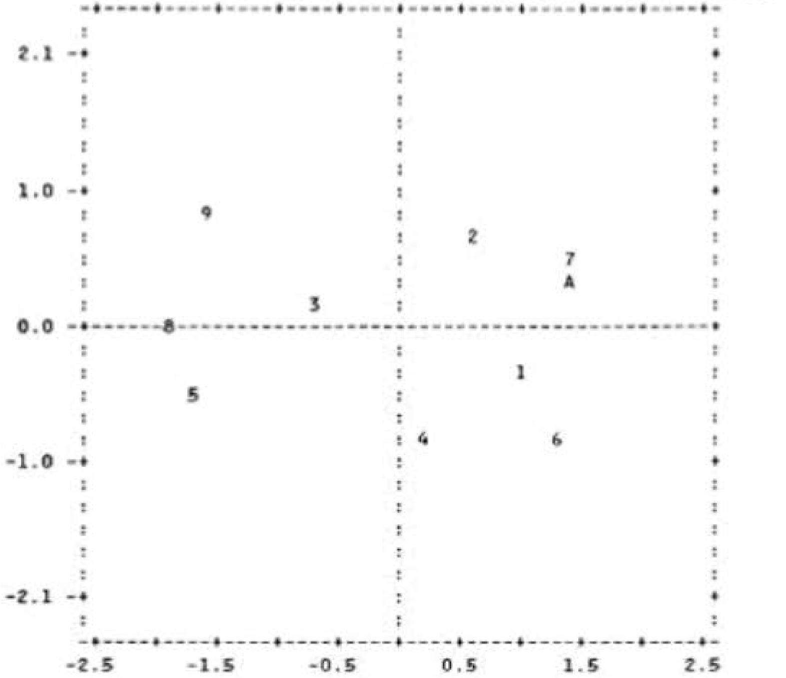
nMDS

STIMULUS COORDINATES

STIMULUS NUMBER	STIMULUS NAME	PLOT SYMBOL	DIMENSION 1	DIMENSION 2
1	ATLANTA	1	0.9586	-0.3385
2	CHICAGO	2	0.6336	0.6347
3	DENVER	3	-0.7085	0.1229
4	HOUSTON	4	0.1955	-0.8345
5	LA	5	-1.6803	-0.5261
6	MIAMI	6	1.3276	-0.8150
7	NEWYORK	7	1.4289	0.5066
8	SANFRAN	8	-1.8769	-0.0782
9	SEATTLE	9	-1.6377	0.9331
10	WASHDC	A	1.3592	0.3948

S-STRESS=0.07
(Muy bueno)

DERIVED STIMULUS CONFIGURATION:
DIMENSION 1 (HORIZONTAL) VS DIMENSION 2 (VERTICAL)



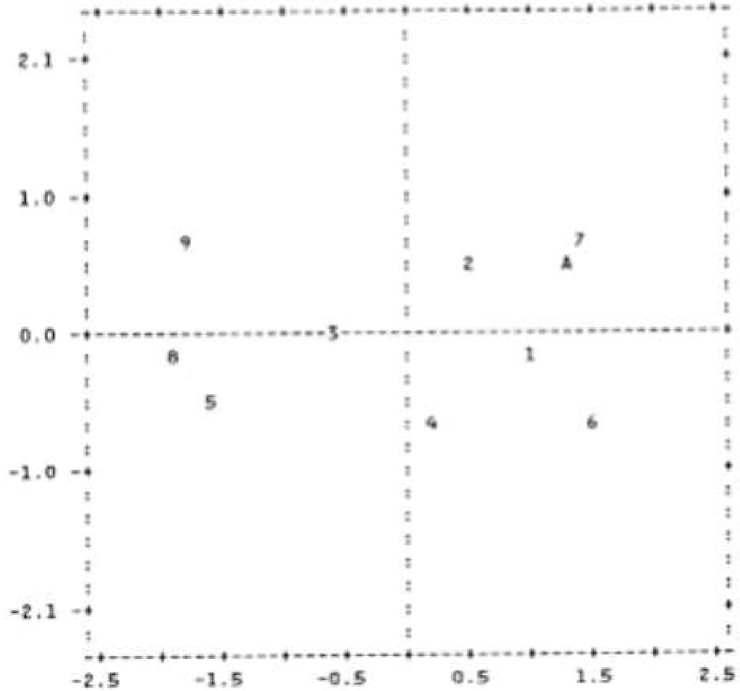
mMDS

STIMULUS COORDINATES

STIMULUS NUMBER	STIMULUS NAME	PLOT SYMBOL	DIMENSION 1	DIMENSION 2
1	ATLANTA	1	0.9575	-0.1905
2	CHICAGO	2	0.5090	0.4541
3	DENVER	3	-0.6416	0.0337
4	HOUSTON	4	0.2151	-0.7631
5	LA	5	-1.6036	-0.5197
6	MIAMI	6	1.5101	-0.7752
7	NEWYORK	7	1.4284	0.6915
8	SANFRAN	8	-1.8925	-0.1500
9	SEATTLE	9	-1.7875	0.7723
10	WASHDC	A	1.3051	0.4469

S-STRESS=0.003
(Muy bueno)

DERIVED STIMULUS CONFIGURATION:
DIMENSION 1 (HORIZONTAL) VS DIMENSION 2 (VERTICAL)



Caso: gaseosas

A un panel de personas que tienen cierto conocimiento sobre gaseosas van a responder preguntas respecto a Diet Pepsi, Royal Crown, Yukon , Dr. Pepper, Shasta, Coca Cola, Diet Dr. Pepper, Tab, Pepsi y Diet Rite Cola. Las disimilaridades se mide en una escala desde 0 hasta 100 en donde puntuaciones bajas significa gaseosas similares y altas calificaciones significa grandes diferencias.

Disimilaridades promedio entre encuestados

0										DIET PEPSI
34.8	0									RC
79.2	54.4	0								YUKON
86	56	70.5	0							DR PEPPER
76.3	30.5	51.2	66.3	0						SHASTA
63.3	40.7	37.8	90	35.4	0					COCA COLA
57.9	86	77.7	50.1	76	77.1	0				DIET DR PEPPER
62.5	80.7	71.6	88.6	67.5	54.1	66.1	0			TAB
65.6	23.6	69.4	66.2	22.6	35.1	76.8	71.3	0		PEPSI
26	60.9	70	89.4	63.1	67.9	59.3	33.6	59.3	0	DIET RITE

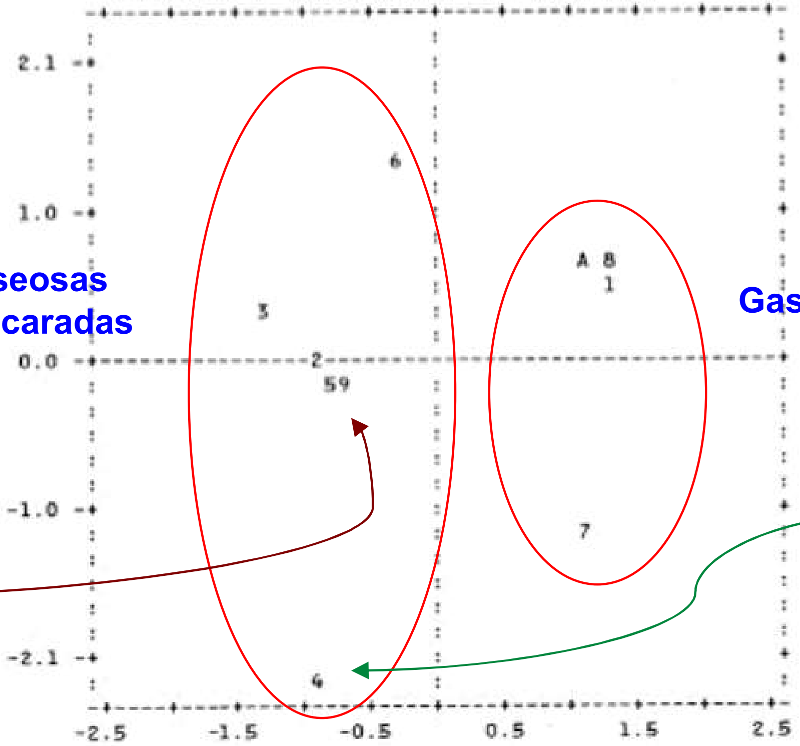
Dr. Pepper y Coca Cola se percibieron como muy disimilares

Shasta y Pepsi se percibieron como similares

STIMULUS COORDINATES

STIMULUS NUMBER	STIMULUS NAME	PLOT SYMBOL	DIMENSION 1	DIMENSION 2
1	DIETPEPS	1	1.2975	0.6838
2	RC	2	-0.8878	-0.0077
3	YUKON	3	-1.2968	0.3561
4	DRPEPPER	4	-0.8796	-2.1726
5	SHASTA	5	-0.7476	-0.1650
6	COCACOLA	6	-0.2746	1.3555
7	DIETDRPR	7	1.0675	-1.2826
8	TAB	8	1.2870	0.7553
9	PEPSI	9	-0.6862	-0.2130
10	DIETRITE	A	1.1205	0.6902

DERIVED STIMULUS CONFIGURATION:
DIMENSION 1 (HORIZONTAL) VS DIMENSION 2 (VERTICAL)



Gaseosas azucaradas

Gaseosas "ligth"

Pepsi Cola *tiene que actuar para diferenciarse*

Tomar acciones respecto a la diferenciación de **Dr. Pepper**

Gracias!!!

Tomado de: https://www.quirks.com/articles/data-use-multidimensional-scaling-for-market-research