

Métodos de Reducción Dimensión

Máster Universitario en
Ingeniería Computacional y Sistemas Inteligentes
UPV/EHU

Métodos - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

- Análisis de coordenadas principales

- MDS métrico

- MDS no métrico

Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

- Análisis de coordenadas principales

- MDS métrico

- MDS no métrico

Calculo Matricial - Repaso

- Un conjunto de p números reales \mathbf{x} se puede representar como un vector en \mathbb{R}^p .

- Producto escalar entre dos vectores $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$.

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = x_1y_1 + \dots + x_py_p.$$

Diremos que son **ortogonales** si $\mathbf{x}'\mathbf{y} = 0$.

- La norma de un vector $\mathbf{x} \in \mathbb{R}^p$.

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{x_1^2 + \dots + x_p^2}.$$

- La distancia euclídea entre \mathbf{x} y \mathbf{y} :

$$\begin{aligned}d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| \\d^2(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}'\mathbf{y}\end{aligned}$$

Calculo Matricial - Repaso

Consideremos x_1, \dots, x_n y y_1, \dots, y_n las observaciones recogidas de dos variables cuantitativas sobre un conjunto de n objetos.

Sean $\mathbf{x} = (x_1, \dots, x_n)'$ y $\mathbf{y} = (y_1, \dots, y_n)'$.

- La medida de dependencia lineal entre dos variables cuantitativas \mathbf{x}, \mathbf{y} es la covarianza.

$$\begin{aligned}s_{xy} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n-1} (\mathbf{x} - \bar{x}\mathbf{1})'(\mathbf{y} - \bar{y}\mathbf{1})\end{aligned}$$

donde $\mathbf{1} = \overset{(n)}{(1, \dots, 1)'}.$

- Cuando las variables están centradas, i.e., $\bar{x} = \bar{y} = 0$, $s_{xy} = \mathbf{x}'\mathbf{y}/(n-1)$ es básicamente el producto escalar.
 \Rightarrow ortogonales \equiv incorrelacionadas

Calculo Matricial - Repaso

- ▶ Llamaremos matriz A de dimensiones $n \times p$ a un conjunto de números reales en n filas y p columnas.
- ▶ Si medimos p variables sobre n individuos, podemos representar cada variable por un vector columna de dimensión n . El conjunto de datos muestrales es una matriz $n \times p$.

Traza

Definición

La traza de una matriz cuadrada $C = (c_{ij})_{i,j}$ $p \times p$ es la suma de todos los elementos de la diagonal principal.

$$\text{tr}(C) = c_{11} + c_{22} + \dots + c_{pp} = \sum_{i=1}^p c_{ii}.$$

Propiedades

- ▶ $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ $A, B \in \mathbb{R}^p \times \mathbb{R}^p$
- ▶ $\text{tr}(\lambda A) = \lambda \text{tr}(A)$ $\lambda \in \mathbb{R}$
- ▶ $\text{tr}(A \cdot B \cdot C) = \text{tr}(B \cdot C \cdot A) = \text{tr}(C \cdot A \cdot B)$ $A, B, C \in \mathbb{R}^p \times \mathbb{R}^p$
- ▶ Si C es simétrica ($C = C'$), $\text{tr}(C^2) = \sum_{i=1}^p \sum_{j=1}^p c_{ij}^2$

Rango

Definición

El rango de una matriz A ($n \times p$) es el número columnas (o filas) linealmente independientes que tiene,

$$rg(A) \leq \min\{n, p\}.$$

Cuando C es una matriz cuadrada ($p \times p$), y $rg(C) < p$, una fila o columna de C es combinación lineal de las demás y decimos que la matriz es **singular**.

Rango

Propiedades

Consideremos C ($p \times p$).

► $rg(C) = p \Leftrightarrow \det(C) \neq 0$

Si además C es no-singular ($rg(C) = p$).

► $rg(C) = rg(C')$

► $rg(C' C) = rg(CC') = rg(C)$

► $rg(CA) = rg(BC) = rg(C)$ si A, B son no-singulares.

Matrices ortogonales

Definición

Una matriz C ($p \times p$) se dice que es ortogonal cuando $C^{-1} = C'$.

Propiedades

- ▶ Dado un vector $\mathbf{x} \in \mathbb{R}^p$, $C\mathbf{x} \in \mathbb{R}^p$ es una rotación o una reflexión de \mathbf{x} .
- ▶ Los vectores columna (o fila) de C son ortogonales.
- ▶ $\det(C) = \pm 1$.

Matriz definida positiva

Definición

Consideremos C ($p \times p$) simétrica ($C = C'$).

- Diremos que C es definida positiva si

$$\mathbf{x}' C \mathbf{x} > 0, \quad \mathbf{x} \neq \mathbf{0}$$

- Diremos que C es semi-definida positiva si

$$\mathbf{x}' C \mathbf{x} \geq 0, \quad \mathbf{x} \neq \mathbf{0}$$

Descomposición espectral

Valores propios

Sea C una matrix cuadrada $(p \times p)$.

Llamaremos **vectores propios** a aquellos vectores \mathbf{u} cuya dirección no se modifica al transformarlos mediante la matriz, i.e.,

$$C\mathbf{u} = \lambda\mathbf{u}$$

donde λ es un escalar que se denomina **valor propio** asociado al vector \mathbf{u} .

Nota: Si \mathbf{u} vector propio $\Rightarrow a\mathbf{u}$ también $(\forall a \in \mathbb{R})$.

Consideraremos $\|\mathbf{u}\| = 1$. Sin embargo, el signo queda indeterminado.

Ejemplo

$$C = \begin{pmatrix} 9 & \sqrt{6} \\ \sqrt{6} & 4 \end{pmatrix}$$

```
C <- matrix(c(9,sqrt(6), sqrt(6), 4), nrow=2)
```

```
eigen(C)
```

```
$values
```

```
[1] 10  3
```

```
$vectors
```

```
      [,1]      [,2]
```

```
[1,] -0.9258201  0.3779645
```

```
[2,] -0.3779645 -0.9258201
```

$$\left(\frac{\sqrt{6}}{\sqrt{7}} = 0.9258201, \quad \frac{1}{\sqrt{7}} = 0.3779645 \right)$$

Propiedades de la matrices definidas positivas

Dado C ($p \times p$),

- ▶ Los vectores propios son ortogonales entre ellos,

$$\mathbf{u}_i' \mathbf{u}_j = 0, \quad i \neq j.$$

Si además C es simétrica y semi-definida positiva:

- ▶ Existen p vectores propios de C .
Cuando se repite $m > 1$ veces un valor propio λ , éste tiene asociados m vectores propios ortogonales entre ellos.
- ▶ Todos los valores propios de C son ≥ 0 .

Diagonalización de matrices simétricas

Sea C una matriz simétrica ($p \times p$).

C se puede convertir en una matriz diagonal!

$$C = UDU'$$

donde

- ▶ $U = [\mathbf{u}_1, \dots, \mathbf{u}_p]$
- ▶ $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ son las matrices de los vectores propios y valores propios de C .

Propiedades de la diagonalización

C y D son esencialmente iguales.

- ▶ $\det(C) = \det(D)$
- ▶ $rg(C) = rg(D)$
- ▶ $tr(C) = tr(D)$

Consecuencias

- ▶ $\det(C) = \lambda_1 \cdot \dots \cdot \lambda_p$
- ▶ $rg(C) = \text{número de valores propios } \neq 0$
- ▶ $tr(C) = \lambda_1 + \dots + \lambda_p$

Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

- Análisis de coordenadas principales

- MDS métrico

- MDS no métrico

Introducción

Consideremos la matriz de datos de n objetos sobre los que se han observado p variables cuantitativas.

Matriz de datos X ($n \times p$).

Objetivo principal:

- ▶ Estudiar la posibilidad de representar adecuadamente esta información con menos variables, construyendo para eso nuevas variables.
- ▶ Estas nuevas variables se construirán como combinación lineal de las originales.

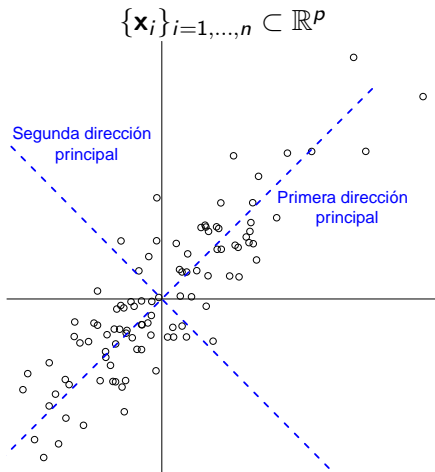
¿En qué nos puede ayudar un ACP?

Si la reducción de la dimensión es posible nos permitirá:

- ▶ Simplificar posteriores análisis a partir de $q < p$ variables.
- ▶ Una representación gráfica de los objetos.
- ▶ Examinar e interpretar las relaciones entre las variables observadas.

Planteamiento del problema

Podemos representar los datos X como una nube de n puntos en \mathbb{R}^p . (Supondremos que previamente hemos restado a cada variable su media, i.e., $\bar{\mathbf{x}} = 0$.)



Planteamiento del problema

- ▶ Buscamos la dirección \mathbf{u} ($\|\mathbf{u}\| = 1$) que minimice la distancia (al cuadrado) entre los puntos originales \mathbf{x}_i y sus proyecciones en dicha dirección \mathbf{y}_i .
 - ▶ $\mathbf{y}_i = (\mathbf{x}_i' \mathbf{u}) \mathbf{u}$.
 - ▶ $d_i = \|\mathbf{x}_i - \mathbf{y}_i\|$, Tma. Pitágoras, $d_i^2 = \|\mathbf{x}_i\|^2 - (\mathbf{x}_i' \mathbf{u})^2$
 - ▶ Sumando para todos los individuos minimizar $\sum_{i=1}^n d_i^2$,

$$\text{minimizar } \sum_{i=1}^n \|\mathbf{x}_i\|^2 - \sum_{i=1}^n (\mathbf{x}_i' \mathbf{u})^2 \Leftrightarrow \text{maximizar } \sum_{i=1}^n (\mathbf{x}_i' \mathbf{u})^2$$

Planteamiento del problema

$$\text{maximizar } \sum_{i=1}^n (\mathbf{x}'_i \mathbf{u})^2$$

- Buscamos la dirección \mathbf{u} ($\|\mathbf{u}\| = 1$) que maximice la varianza de los datos proyectados.

Esta dirección será la que determinará la primera componente principal.

Definición de las componentes principales

Hemos comentado que la primera componente principal C_1 será,

$$c_{i1} = u_{11}x_{i1} + \dots + u_{p1}x_{ip}, \quad i = 1, \dots, n$$

de manera que maximice la varianza $s^2(C_1) = \sum c_{i1}^2$ y $\|\mathbf{u}_1\| = 1$ ($u_{11}^2 + \dots + u_{p1}^2 = 1$).

Análogamente, construiremos la 2 componente principal C_2 tal que:

- ▶ $c_{i2} = u_{12}x_{i1} + \dots + u_{p2}x_{ip}$, $i = 1, \dots, n$
- ▶ $s^2(C_2)$ máxima
- ▶ $\|\mathbf{u}_2\| = 1$
- ▶ C_1 y C_2 incorrelacionadas, $\mathbf{u}_1' \mathbf{u}_2 = 0$

Definición de las componentes principales

De manera similar podemos ir fijando sucesivas direcciones.

Construiremos la k -ésima componente principal C_k tal que:

- ▶ $c_{ik} = u_{1k}x_{i1} + \dots + u_{pk}x_{ip}, i = 1, \dots, n$
- ▶ $s^2(C_k)$ máxima
- ▶ $\|\mathbf{u}_k\| = 1$
- ▶ C_k está incorrelacionada con todas las anteriores componentes, $\mathbf{u}_k' \mathbf{u}_j = 0, j < k$

Podemos construir tantas componentes principales como dimensión de los datos p .

Primera componente principal

La primera componente viene dada por la dirección \mathbf{u}_1 tal que:

- ▶ $c_{i1} = u_{11}x_{i1} + \dots + u_{p1}x_{ip}, \quad i = 1, \dots, n$
- ▶ $s^2(C_1) = \mathbf{u}_1' S \mathbf{u}_1$ máxima
- ▶ $\|\mathbf{u}_1\| = 1$

Solución:

- ▶ \mathbf{u}_1 es el vector propio asociado al mayor de los valores propios λ_1 de la matriz de varianzas-covarianzas S .

Segunda componente principal

La segunda componente viene dada por la dirección \mathbf{u}_2 tal que:

- ▶ $c_{i2} = u_{12}x_{i1} + \dots + u_{p2}x_{ip}, \quad i = 1, \dots, n$
- ▶ $s^2(C_2) = \mathbf{u}_2' S \mathbf{u}_2$ máxima
- ▶ $\|\mathbf{u}_2\| = 1$
- ▶ $\mathbf{u}_1' \mathbf{u}_2 = 0$

Solución:

- ▶ \mathbf{u}_2 es el vector propio asociado al segundo mayor de los valores propios λ_2 de la matriz de varianzas-covarianzas S .

Generalización

Puede verse que:

- ▶ En general hay p componentes principales ($rg(X) = rg(S) = p$).
- ▶ Sean $C = [\mathbf{c}_1, \dots, \mathbf{c}_p]$ ($n \times p$) y $U = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ ($p \times p$).
Entonces,

$$C = XU$$

donde $U'U = I$ (U ortogonal).

Por tanto, calcular las componentes principales equivale a aplicar una rotación/reflexión a los datos originales.

- ▶ El espacio de dimensión $q < p$ que mejor representa los datos es el definido por los vectores propios asociados a los q mayores valores propios de S . (Primeras q componentes principales).

Propiedades de las Componentes Principales

- ▶ Tienen media 0, $\bar{c}_i = 0$, $i = 1, \dots, p$.
- ▶ Varianza, $s^2(C_i) = \lambda_i$, $i = 1, \dots, p$.
- ▶ Variabilidad de los datos = Variabilidad de las componentes.

$$tr(S) = s^2(X_1) + \dots + s^2(X_p) = \lambda_1 + \dots + \lambda_p$$

- ▶ Porcentaje de variabilidad que aporta cada componente i -ésima:

$$\frac{\lambda_i}{\lambda_1 + \dots + \lambda_p}$$

- ▶ Porcentaje de variabilidad que aporta las primeras $q < p$ componentes:

$$\frac{\lambda_1 + \dots + \lambda_q}{\lambda_1 + \dots + \lambda_q + \dots + \lambda_p}$$

- ▶ Correlación entre variables originales y componentes:

$$Cor(C_i, X_j) = u_{ji} \frac{\sqrt{\lambda_i}}{s(X_j)}$$

Componentes Principales sobre correlaciones

Cuando las unidades en las que se han medido las variables X_1, \dots, X_p son muy distintas, las variables con valores más grandes y mayores varianzas tendrán más peso en el análisis.

Solución:

Estandarizar las variables ($X_i \rightarrow Z_i$) y aplicar este mismo análisis a las variables Z_1, \dots, Z_p .

Este análisis se obtiene calculando los vectores y valores propios de la **matriz de correlaciones** de las variables originales.

¿Cuántas componentes retener?

A fin de reducir la dimensión hay que decidir el número de componentes $q < p$ que se van a retener.

1. Si se retienen las p posibles no se pierde información pero no se simplifica el problema.
2. Si unas pocas componentes son capaces de explicar una buena parte de la varianza total, podrán sustituir a las variables primitivas sin pérdida grave de información.

$$q = ?$$

¿Cuántas componentes retener?

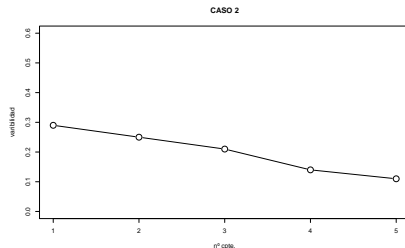
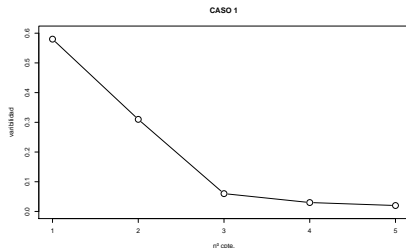
Reglas orientativas

Para ilustrar estas reglas nos basaremos en el siguiente ejemplo:

Cpte.	Variabilidad explicada $(\lambda_i)/(\lambda_1 + \dots + \lambda_p)$	
	CASO 1	CASO 2
1	58%	29%
2	31%	25%
3	6%	21%
4	3%	14%
5	2%	11%

Reglas orientativas para elegir q

Scree graph Se trata de representar de mayor a menor los valores propios o porcentaje de variabilidad explicada.



Reglas orientativas para elegir q

Criterio del cociente Se trata de representar de dividir sucesivamente la varibilidad explicada por cada componente con la siguiente.

Cocientes altos son indicativos de que la componente es más relevante que la siguiente.

	Componente			
	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5
CASO 1:	1.9	5.2	2.0	1.5
CASO 2:	1.2	1.2	1.5	1.3

Reglas orientativas para elegir q

Criterio de Kaiser Se trata de retener aquellas componentes que expliquen más variabilidad de la que harían por azar.

- ▶ Al trabajar con la matriz de correlaciones.

Retener las componentes con valor propio > 1 .

- ▶ Al trabajar con la matriz de covarianzas.

Retener las componentes con valor propio $> (\lambda_1 + \dots + \lambda_p)/p$.

Ejemplo

Los siguientes datos muestran la descomposicin de distintos alimentos en calorías, hidratos de carbono, grasas y proteínas

ALIMENTO	CALORÍAS	HIDR. CARB.	GRASAS	PROTEINAS
chocolate	564	508	376	89
queso	383	30	280	300
pollo	130	0	65	179
ternera	92	1	10	200
manzana	45	110	3	2
espinaca	29	33	2	36
trucha	86	0	30	147
pasta	363	699	6	238
patata	85	180	10	21
leche	63	48	34	31
fresa	27	53	4	9

Ejemplo

¿Están las variables correlacionadas?

Cor	CALORÍAS	HIDR. CARB.	GRASAS	PROTEINAS
CAL.	1.00	0.70	0.85	0.522
HIDR.	0.70	1.00	0.28	0.16
GRASA	0.85	0.28	1.00	0.34
PROT	0.52	0.16	0.34	1.00

Ejemplo - Resultados R

Análisis sobre la matriz de correlaciones.

```
> b <- prcomp(x[2:5],scale.=T)
> b$sdev
[1] 1.580991124 0.925575205 0.802349540 0.003580598
> b$rotation
```

	PC1	PC2	PC3	PC4
cal	-0.6293644	-0.07744606	-0.08626063	-0.7684150
hidr	-0.4305462	-0.68946513	0.44837404	0.3717910
grasa	-0.5221825	0.14604895	-0.68283528	0.4896234
prot	-0.3819108	0.70520185	0.57031202	0.1777040

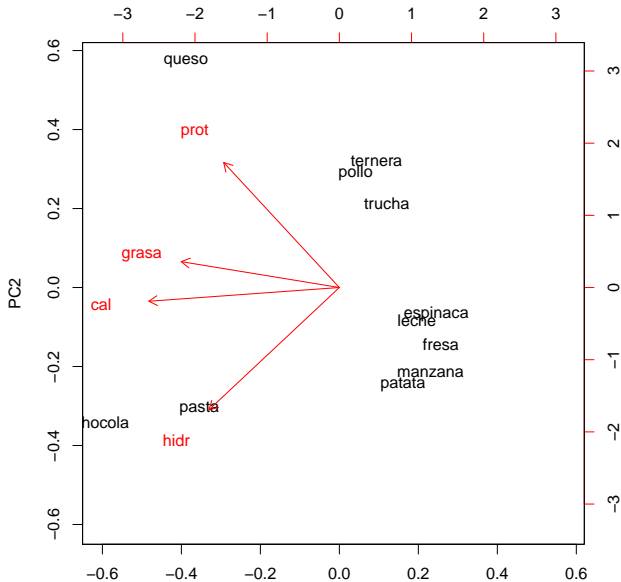
```
> b$x
```

	PC1	PC2	PC3	PC4
1	-3.1624431	-1.0462747	-1.24130859	-0.0015066479
2	-2.0364925	1.7595646	-0.40594941	0.0018760614
3	0.2161128	0.8928015	0.13692553	0.0034470469
4	0.4929391	0.9858002	0.56414313	-0.0072371946
5	1.2102409	-0.6638169	-0.25141561	0.0012023817
6	1.2870379	-0.2009561	-0.19991178	0.0009776766
7	0.6287237	0.6552209	0.16864562	0.0024303995
8	-1.8577572	-0.9348710	2.00217625	0.0006007588
9	0.8441258	-0.7506802	-0.06947114	0.0016278645
10	1.0294587	-0.2572144	-0.38475602	-0.0062366630
11	1.3480538	-0.4395738	-0.31907798	0.0028183161

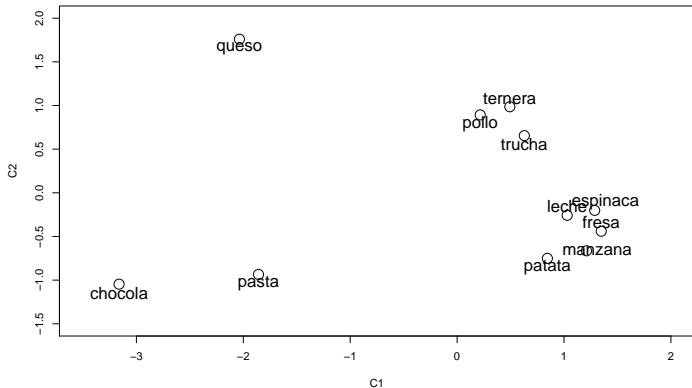
Ejemplo - Resultados R

1. ¿Cuántas variables retener? ¿Qué porcentaje de variabilidad explican?
2. Interpretación de las componentes.

Ejemplo - Biplot



Ejemplo



Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

Análisis de coordenadas principales

MDS métrico

MDS no métrico

Multidimensional Scaling - MDS

Introducción

Consideremos un conjunto Ω de n objetos y una matriz de disimilaridades $\Delta = (\delta_{ij})_{ij}$ ($n \times n$) entre los objetos.

Objetivo principal:

- Obtener una representación de los objetos en un espacio de dimensión pequeña.
Las inter-posiciones d_{ij} de los objetos en este espacio deben reproducir las disimilaridades originales δ_{ij} .

Introducción

Dentro MDS se agrupan muchas y diferentes técnicas:

- ▶ **Análisis de coordenadas principales - Classical scaling**

Ofrece una solución algebraica.

- ▶ **MDS métrico**

Estas técnicas tratan de obtener la configuración de los objetos a partir de los valores numéricos δ_{ij} .

- ▶ SMACOF: Ratio-MDS, Interval-MDS

- ▶ **MDS no-métrico**

En algunas situaciones el valor numérico δ_{ij} es poco fiable. En estas situaciones es preferible utilizar técnicas que se basan en las propiedades ordinales de δ_{ij} .

Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

- Análisis de coordenadas principales

 - MDS métrico

 - MDS no métrico

Análisis de coordenadas principales

Supongamos que tenemos n objetos y Δ la matriz de disimilaridades sobre ellos.

En el análisis de coordenadas principales supondremos que δ_{ij} son **distancias euclídeas**. Es decir, existen coordenadas $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ de manera que:

$$\delta_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j).$$

¿Podemos encontrar estas coordenadas?

Llamaremos **coordenadas principales** de Δ a las p coordenadas $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ de cada objeto i .

Obtención de las coordenadas principales

- ▶ Calcular la matriz $A = \left(-\frac{1}{2}\delta_{ij}^2\right)$
- ▶ Calcular la matrix $B = HAH$ donde H es la matriz de centrado $H = I_n - \mathbf{1}\mathbf{1}'/n$.
Supongamos que B tiene rango p .
- ▶ Hallar la descomposición espectral de B . Es decir,
 1. Los valores propios $\lambda_1 > \dots > \lambda_p \geq 0$ (*).
 2. Los vectores propios asociados $\mathbf{u}_1, \dots, \mathbf{u}_p$.

$$B\mathbf{u}_i = \lambda_i\mathbf{u}_i, \quad i = 1, \dots, p$$

- ▶ Construir las matrices $U = (\mathbf{u}_1, \dots, \mathbf{u}_p)$ y $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}$ y calcular:

$$X = U\Lambda^{1/2} \quad (n \times p)$$

Cada fila i de X contiene las coordenadas principales del objeto i .

Propiedades

- ▶ Las filas $\mathbf{x}_1, \dots, \mathbf{x}_n$ de X reproducen la matriz de distancias Δ .

$$(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) = \delta_{ij}^2$$

- ▶ Las columnas de X entendidas como variables, tienen media 0.
- ▶ Cada columna j de X tiene varianza igual a $\lambda_j/(n-1)$ ($j = 1, \dots, p$).
- ▶ Las columnas de X están incorrelacionadas (ortogonales).

Ejemplo "bebidas"

La siguiente matriz de disimilaridades entre 6 bebidas refrescantes se obtuvo a partir de la valoración dada por 38 estudiantes siguiendo una escala (1=no similar, 9=muy similar). Las similitudes acumuladas se transformaron en disimilaridades.

$$\Delta = \begin{pmatrix} 0 & . & . & . & . & . \\ 127 & 0 & . & . & . & . \\ 167 & 143 & 0 & . & . & . \\ 207 & 235 & 243 & 0 & . & . \\ 320 & 322 & 327 & 288 & 0 & . \\ 321 & 318 & 318 & 317 & 136 & 0 \end{pmatrix} \begin{array}{l} \text{PepsiCola} \\ \text{CocaCola} \\ \text{ClassicCC} \\ \text{DietPepsi} \\ \text{Diet7 - Up} \\ \text{SevenUp} \end{array}$$

Calcularemos sus coordenadas principales.

Ejemplo "bebidas". Computación con R

```
> espectral <- eigen(B)
> espectral
$values
[1] 1.113171e+05 3.459311e+04 1.332382e+04 7.844151e+03 6.558663e+03
[6] -3.620511e-12

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] -0.3367463 0.02728232 0.584673812 0.279020722 0.5472613 -0.4082483
[2,] -0.3385205 0.28563491 0.287841066 -0.371553697 -0.6451707 -0.4082483
[3,] -0.3254777 0.35097834 -0.731207906 -0.009182834 0.2635573 -0.4082483
[4,] -0.1370492 -0.84697640 -0.193646961 0.130693634 -0.2064022 -0.4082483
[5,] 0.5692291 -0.08217526 0.055920366 -0.633564950 0.3130929 -0.4082483
[6,] 0.5685647 0.26525608 -0.003580376 0.604587124 -0.2723386 -0.4082483
```

Ejemplo "bebidas". Computación con R

```
> U <- espectral$vector[,1:(n-1)]
> L <- diag(espectral$values[1:(n-1)])
> L <- sqrt(L)
>
> X <- U%*%L
> X
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-112.35277	5.07430	67.4882296	24.712088	44.32027
[2,]	-112.94472	53.12587	33.2251651	-32.907476	-52.24952
[3,]	-108.59309	65.27924	-84.4024926	-0.813298	21.34434
[4,]	-45.72539	-157.53102	-22.3524473	11.575171	-16.71560
[5,]	189.91882	-15.28396	6.4548239	-56.113083	25.35601
[6,]	189.69715	49.33557	-0.4132788	53.546598	-22.05550

```
> dist(X)
```

	1	2	3	4	5
2	127				
3	167	143			
4	207	235	243		
5	320	322	327	288	
6	321	318	318	317	136

Reducción de dimensión

Para reducir la dimensión a $q < p$ seleccionaremos las q primeras coordenadas principales. Es decir, representaremos el objeto i mediante:

$$\mathbf{x}_i^* = (x_{i1}, \dots, x_{iq})'$$

Llamaremos d_{ij}^* a las distancias calculadas a partir de estas coordenadas truncadas.

$$d_{ij}^{*2} = (\mathbf{x}_i^* - \mathbf{x}_j^*)(\mathbf{x}_i^* - \mathbf{x}_j^*)'$$

Reducción de dimensión - justificación

- Sea $\tilde{\mathbf{x}}_i$ una proyección de \mathbf{x}_i en un subespacio de dimensión q y \tilde{d}_{ij} las distancias derivadas de éstas.

Consideremos

$$S = \sum_i \sum_j (\delta_{ij}^2 - \tilde{d}_{ij}^2),$$

El mínimo de S se alcanza para $\tilde{d}_{ij} = d_{ij}^*$. Es decir, considerando las q primeras coordenadas de principales.

¿Cuántas dimensiones retener?

Para seleccionar número de coordenadas principales q que vamos a retener nos basaremos en la siguiente propiedad:

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \delta_{ij}^2 = n(\lambda_1 + \dots + \lambda_p)$$

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^{*2} = n(\lambda_1 + \dots + \lambda_q)$$

- Porcentaje de variabilidad explicada por las q primeras dimensiones:

$$\frac{\lambda_1 + \dots + \lambda_q}{\lambda_1 + \dots + \lambda_q + \dots + \lambda_p}$$

Ejemplo "bebidas". Computación con R

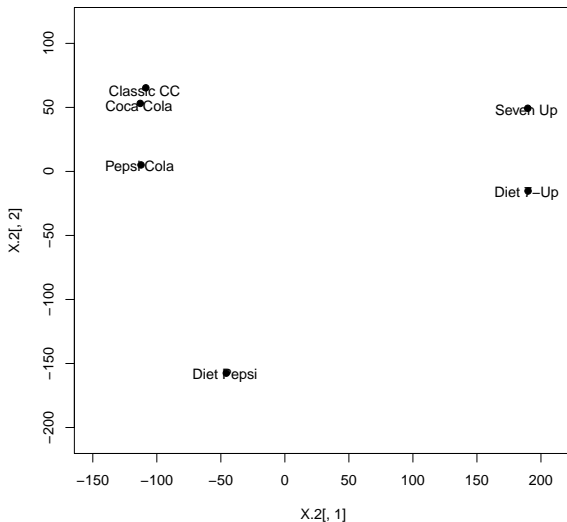
```
# Retenemos 2 dimensiones
> X.2 <- X[,1:2]

> dist(X.2)
      1          2          3          4          5
2  48.05522
3  60.32222  12.90895
4 175.72620 221.12160 231.50974
5 302.95639 310.49353 309.19215 275.24974
6 305.27564 302.66561 298.71603 313.39681 64.61991

> dist(X)
      1    2    3    4    5
2 127
3 167 143
4 207 235 243
5 320 322 327 288
6 321 318 318 317 136

> sum(espectral$values[1:2])/sum(espectral$values)
[1] 0.8403182
```

Ejemplo "bebidas" Computación con R



Coordenadas principales y Componentes principales

Sea X una matriz de datos ($n \times p$) centrada ($\bar{X}_j = 0, j = 1, \dots, p$). Calculamos,

1. sus componentes principales a partir de la matriz de covarianzas $\frac{1}{n-1}X'X$.
2. la matriz de distancias euclídeas,

$$\delta_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j),$$

y las coordenadas principales correspondientes.

Resultado:

Las componentes principales coinciden con las coordenadas principales.

Ejemplo

Retomamos el ejemplo de algunos alimentos y su descomposición en calorías, hidratos de carbono, grasas y proteínas

ALIMENTO	CALORÍAS	HIDR. CARB.	GRASAS	PROTEINAS
chocolate	564	508	376	89
queso	383	30	280	300
pollo	130	0	65	179
ternera	92	1	10	200
manzana	45	110	3	2
espinaca	29	33	2	36
trucha	86	0	30	147
pasta	363	699	6	238
patata	85	180	10	21
leche	63	48	34	31
fresa	27	53	4	9

Ejemplo

Ya habíamos visto el resultado del ACP aplicado a la matriz de correlaciones. Comprobemos que podemos obtener la misma información a través de las coordenadas principales.

Componentes principales

```
x.acp <- prcomp(x, scale=T)
x.acp$x[,1:2]
      PC1    PC2
chocola -3.16 -1.05
queso    -2.04  1.76
pollo     0.22  0.89
ternera   0.49  0.99
manzana   1.21 -0.66
espinaca  1.29 -0.20
trucha    0.63  0.66
pasta    -1.86 -0.93
patata    0.84 -0.75
leche     1.03 -0.26
fresa     1.35 -0.44
```

Coordenadas principales

```
x.s <- scale(x)
d <- dist(x.s)
x.mds <- cmdscale(d, eig=T, k=4)
x.mds$points[,1:2]
      [,1] [,2]
chocola -3.16  1.05
queso    -2.04 -1.76
pollo     0.22 -0.89
ternera   0.49 -0.99
manzana   1.21  0.66
espinaca  1.29  0.20
trucha    0.63 -0.66
pasta    -1.86  0.93
patata    0.84  0.75
leche     1.03  0.26
fresa     1.35  0.44
```

Cuando Δ no es euclídea

Dada la matriz de distancias Δ puede ser que no sea euclídea.
Esta situación es fácil de detectar ya que:

$$\Delta \text{ es euclídea} \Leftrightarrow \text{todos los v.p. de } B \lambda_1, \dots, \lambda_q \geq 0$$

Cuando Δ no es euclídea

Sea Δ no euclídea y $\lambda_1 > \dots > \lambda_k > 0 > \lambda'_1 > \dots > \lambda'_{k'}$ los valores propios de B .

Normalmente esta situación se afronta como:

- Ignorar las dimensiones correspondientes a valores propios negativos y medir la precisión de la aproximación q -dimensional como:

$$\frac{\lambda_1 + \dots + \lambda_q}{\sum_h |\lambda_h|}$$

- Transformar la matriz Δ en $\tilde{\Delta}$ de manera que $\tilde{\Delta}$ sí sea euclídea.

Transformaciones para construir $\tilde{\Delta}$

Sea Δ no eucídea.

Transformaciones para construir $\tilde{\Delta}$:

1. Transformación q -aditiva

$$\tilde{\delta}_{ij} = (\delta_{ij}^2 + c)^{1/2} \quad \text{con} \quad c > -2\lambda_{q'}$$

2. Transformación aditiva

$$\tilde{\delta}_{ij} = \delta_{ij} + c \quad \text{con} \quad c \text{ cte. 'apropiada'}$$

De esta manera $\tilde{\Delta}$ es euclídea y conserva la preordenación de Δ .

Ejemplo - "Municipios"

Hemos obtenido la distancia (en Km) entre 7 municipios de la CAV (Guía Michelin).

La matriz de distancias es la siguiente:

>d

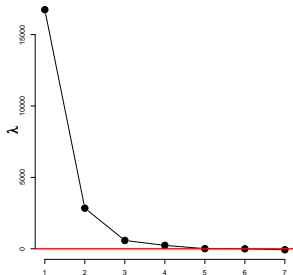
	Donostia	Bilbo	Gasteiz	Azpeitia	Irun	Oñati	Hondarribia
Don.	0	99	113	41	22	76	23
Bil.	99	0	65	70	118	73	118
Gas.	113	65	0	77	129	48	129
Azp.	41	70	77	0	61	46	61
Irun	22	118	129	61	0	95	4
Oa.	76	73	48	46	95	0	95
Hon.	23	118	129	61	4	95	0

Ejemplo - "Municipios"

¿Es euclídea?

Calculemos los valores propios asociados de B :

```
espectral <- eigen(B)
espectral$values
16744.2  2844.9   583.7   239.4    10.1
-4.121148e-13 -67.8
```



Diferentes posibilidades:

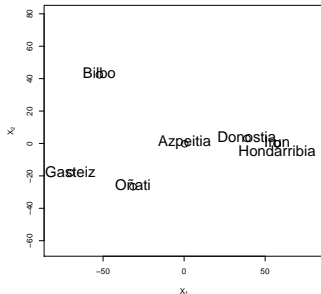
- S1: Ignorar los valores propios negativos.
- S2: Transformación aditiva de la matriz de distancias.
- S3: Transformación q -aditiva de la matriz de distancias.

Ejemplo - "Municipios"

S1: Ignorar los valores propios negativos

Representaremos la solución en 2 dimensiones.

```
> d.mds$points[,1:2]
      [,1] [,2]
Donostia 38.73  3.11
Bilbo    -52.28 42.34
Gasteiz  -70.15 -18.19
Azpeitia   0.35 -0.18
Irun      57.46 -0.19
Onati     -31.51 -26.68
Hondarribia 57.41 -0.22
```



¿Porcentaje?

Ejemplo - "Municipios"

S2: Transformación aditiva de la matriz de distancias

Representaremos la solución en 2 dimensiones.

```
> d.mds <- cmdscale(d,k=6, eig=T, add=T)
```

```
> d.mds$eig
```

```
1.764346e+04 3.174171e+03
```

```
1.844816e+03 3.818309e+02
```

```
3.275763e+01 6.576901e-01
```

```
> d.mds$points[,1:2]
```

```
      [,1]      [,2]
```

```
Donostia    43.7   -0.87
```

```
Bilbo       -50.8   42.06
```

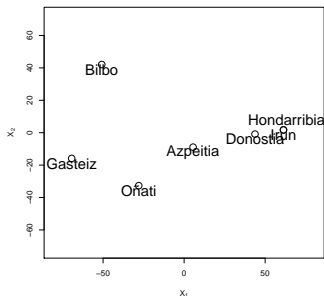
```
Gasteiz     -69.4  -15.85
```

```
Azpeitia      5.5   -8.91
```

```
Irun         61.3    1.72
```

```
Onati        -28.0  -32.67
```

```
Hondarribia  61.3    1.73
```



¿Porcentaje?

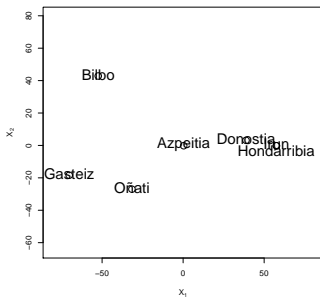
Ejemplo - "Municipios"

S3: Transformación q -aditiva de la matriz de distancias

Representaremos la solución en 2 dimensiones.

```
> dtilde <- d*d+1  
> diag(dtilde) <- 0  
> dtilde <- sqrt(dtilde)  
  
> eigen(B)$values  
1.68e+04 2.92e+03 6.59e+02  
3.14e+02 8.51e+01 7.16e+00  
3.73e-13
```

¿Porcentaje?



Características básicas del MDS

En general, los métodos de MDS se engloban dentro del modelo:

$$d_{ij} = f(\delta_{ij}) + \epsilon_{ij} \quad i, j = 1, \dots, n$$

donde:

- ▶ δ_{ij} disimilaridades originales/observadas.
- ▶ d_{ij} distancias euclídeas, i.e., una configuración X euclídea de los objetos (en \mathbb{R}^q). Es decir, $(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) = d_{ij}^2$.
Generalmente se buscan las coordenadas X en un espacio de "pequeña" dimensión.
- ▶ f función monótona no decreciente.

Características básicas del MDS

Supongamos que d_{ij} una configuración euclídea en dimensión q .

Para que ésta sea una *buena* configuración pediremos que:

- ▶ $d_{ij} \approx \delta_{ij}$, o bien,
- ▶ $d_{ij} \approx f(\delta_{ij})$, flexibilizando la condición anterior, mediante una función *f adecuada*.

¿Qué significa

- ▶ *f adecuada?*
- ▶ \approx ?

Características básicas del MDS

¿f adecuada?

- ▶ f monótona no decreciente.

f preserva la preordenación de las distancias originales δ_{ij}

$$\delta_{i_1j_1} < \delta_{i_2j_2} < \dots < \delta_{i_Nj_N} \Rightarrow f(\delta_{i_1j_1}) \leq f(\delta_{i_2j_2}) \leq \dots \leq f(\delta_{i_Nj_N})$$

- ▶ $f(\delta_{ij}) = \hat{d}_{ij}$ disparidad

Características básicas del MDS

$$d_{ij} \approx f(\delta_{ij}) ?$$

Minimizar:

- S_r , raw stress

$$S_r^2 = \sum_{i < j} (d_{ij} - f(\delta_{ij}))^2$$

- S , stress

$$S^2 = \frac{\sum_{i < j} (d_{ij} - f(\delta_{ij}))^2}{\sum_{i < j} d_{ij}^2}$$

- S_w , stress ponderado

$$S_w^2 = \frac{\sum_{i < j} w_{ij} (d_{ij} - f(\delta_{ij}))^2}{\sum_{i < j} d_{ij}^2}$$

Características básicas del MDS

Proceso obtención de la configuración euclídea

Pasos generales:

Paso 1 Se parte de una configuración inicial X en \mathbb{R}^q

$$d_{ij} = [(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)]^{1/2}.$$

Paso 2 Se calculan las disparidades $f(\delta_{ij}) = \hat{d}_{ij}$

Paso 3 Se calcula una nueva configuración euclídea $\mathbf{x} = (\mathbf{x}'_i)_i$ que minimice alguna versión del stress

$$S(\mathbf{x}) = \left(\frac{\sum_{i < j} (d_{ij}(\mathbf{x}_i, \mathbf{x}_j) - \hat{d}_{ij})^2}{\sum_{i < j} \hat{d}_{ij}^2} \right)^{1/2}$$

Paso 4 Se vuelve al paso 2 hasta convergencia.

Particularidades en los pasos 2 y 3 llevan a diferentes métodos de MDS.

SMACOF: Algoritmo iterativo de mayorización

Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

Análisis de coordenadas principales

MDS métrico

MDS no métrico

MDS métrico - Ratio and Interval MDS

Establecen funciones lineales para buscar las disparidades

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

- ▶ Ratio-MDS: $\hat{d}_{ij} = b\delta_{ij}$ con $b > 0$
- ▶ Interval-MDS: $\hat{d}_{ij} = a + b\delta_{ij}$ con $b > 0$

Paso 2 Se calculan las disparidades $\hat{d}_{ij} = a + b\delta_{ij}$ pero... ¿ a , b ?
Se utilizan los coeficientes que minimicen el stress.

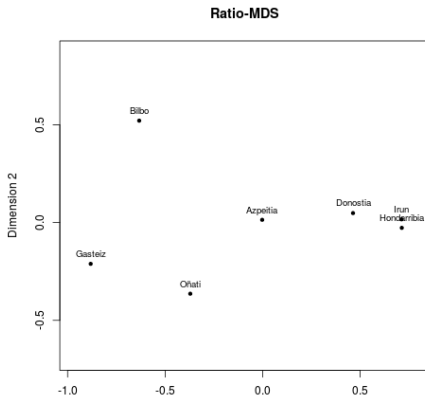
Paso 3 Dadas las disparidades se busca una nueva configuración que minimice S

Ejemplo - "Municipios"

Ratio-MDS

Representaremos la solución para $q = 2$ dimensiones.

```
> library(smacof)
> ratio.mds <- mds(d, type="ratio")
> ratio.mds$stress
[1] 0.02637932
```



Métodos de Reducción Dimensión - Índice

Preliminares - Cálculo matricial

Análisis de Componentes Principales

Multidimensional Scaling

Análisis de coordenadas principales

MDS métrico

MDS no métrico

MDS no métrico - Ordinal MDS

En algunas situaciones el valor numérico exacto de las disimilitudes δ_{ij} puede ser poco fiable. En estos casos es preferible utilizar métodos que utilicen sólo sus propiedades **ordinales**.

Paso 2 Se calculan las disparidades $f(\delta_{ij})$ mediante una función que conserve la preordenación inicial

$$\hat{d}_{ij} \leq \hat{d}_{kl} \quad \text{siempre que} \quad \delta_{ij} < \delta_{kl}$$

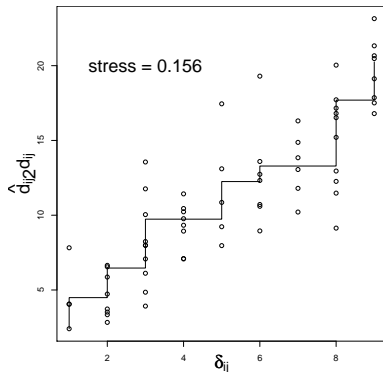
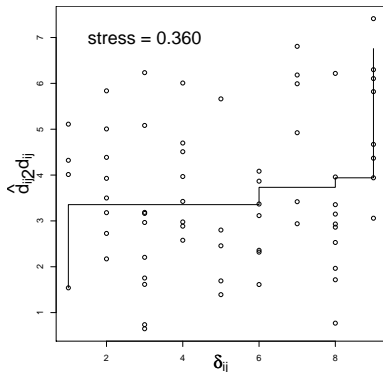
Éstas se ajustan a través de la **regresión monótona o isotónica**.

Paso 3 Se busca una nueva configuración que minimice S

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

Bondad del ajuste

- ▶ La práctica sugiere la siguiente interpretación del stress como bondad del ajuste.
 - ▶ $S \sim 0.2$, ajuste pobre.
 - ▶ $S \sim 0.05$, ajuste bueno.
 - ▶ $S \sim 0.025$, ajuste excelente.
- ▶ Shepard's diagram (δ_{ij} vs. \hat{d}_{ij} y d_{ij})



Ejemplo "bebidas"

Ya habíamos visto este ejemplo donde la matriz de disimilaridades entre 6 bebidas refrescantes se obtuvo a partir de la valoración dada por 38 estudiantes siguiendo una escala (1=no similar, 9=muy similar). Las similaridades acumuladas se transformaron en disimilaridades.

$$\Lambda = \begin{pmatrix} 0 & . & . & . & . & . \\ 127 & 0 & . & . & . & . \\ 167 & 143 & 0 & . & . & . \\ 207 & 235 & 243 & 0 & . & . \\ 320 & 322 & 327 & 288 & 0 & . \\ 321 & 318 & 318 & 317 & 136 & 0 \end{pmatrix} \begin{array}{l} \text{PepsiCola} \\ \text{CocaCola} \\ \text{ClassicCC} \\ \text{DietPepsi} \\ \text{Diet7 - Up} \\ \text{SevenUp} \end{array}$$

Encontraremos una representación en \mathbb{R}^2 con MDS no métrico.

Ejemplo "bebidas" – Resultados R

```
> library(smacof)}  
> ratio.mds <- mds(d, type="ordinal")  
> ratio.mds$stress  
[1] 0.001301569
```

