

10. ESCALADO MULTIDIMENSIONAL - EMD

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

- El **escalado multidimensional** es una técnica diseñada para construir “mapas” que muestran las relaciones entre un número de objetos/observaciones a partir de una tabla de distancias.
- Mapas 1D (si los objetos pueden representarse por una recta)
- Mapas 2D (si los objetos pueden representarse en un plano)
- Mapas 3D (si los objetos pueden representarse en un espacio 3D)

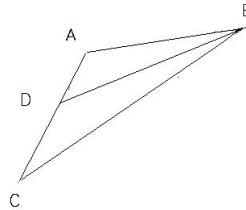
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

Ejemplo 1

- Sean 4 objetos A, B, C, D y su matriz de distancias:

	A	B	C	D
A	0,0	6,0	6,0	2,5
B	6,0	0,0	9,5	7,8
C	6,0	9,5	0,0	3,5
D	2,5	7,8	3,5	0,0



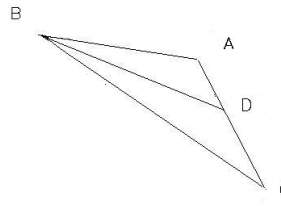
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

Ejemplo 1

- Sin embargo, la imagen espejo también tiene por base la misma matriz de distancias entre los objetos. Por consiguiente, la reconstrucción del mapa original puede estar sujeta a una inversión, como en este caso.

	A	B	C	D
A	0,0	6,0	6,0	2,5
B	6,0	0,0	9,5	7,8
C	6,0	9,5	0,0	3,5
D	2,5	7,8	3,5	0,0



AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

Ejemplo 2

- Si se considera más de 3 objetos/observaciones, probablemente no puedan representarse por un plano. En ese caso, la matriz de distancias implícitamente contendrá dicha información.

	A	B	C	D
A	0	1	$\sqrt{2}$	$\sqrt{2}$
B	1	0	1	1
C	$\sqrt{2}$	1	0	$\sqrt{2}$
D	$\sqrt{2}$	1	$\sqrt{2}$	0

- Por ejemplo, en este caso se requieren 3D para representar los 4 objetos/observaciones.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

- Desafortunadamente, con datos reales usualmente no se conoce el número de dimensiones que se requiere para poder representarlos gráficamente.
- La utilidad del escalado multidimensional viene del hecho que, a menudo surgen situaciones en que la relación entre los objetos no es conocida, pero puede estimarse una matriz de distancias.
- Las principales aplicaciones del escalado multidimensional han tenido lugar en la psicología y la sociología.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

“MAPAS” a partir de matrices de distancia

Existe una variedad de técnicas de análisis de datos bajo la denominación “escalado multidimensional”.

Consideraremos solamente los métodos clásicos más simples propuestos por:

- Torgerson (1952)
- y
- Kruskal (1964).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Procedimiento para el escalado multidimensional

- El escalado multidimensional clásico parte de una matriz de distancias entre n objetos, en que δ_{ij} representa la distancia entre el objeto i y el objeto j , en la i -ésima fila y j -ésima columna.

- El número de dimensiones, t , para el mapeo de los objetos se fija para cada solución particular.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Procedimiento para el escalado multidimensional

1. Se fija una configuración inicial para los n objetos en t dimensiones. Es decir, se asumen las coordenadas (x_1, x_2, \dots, x_t) para cada objeto en el espacio tD .
2. Se calculan las distancias Euclídeas entre los objetos/observaciones, para la configuración inicial. d_{ij} representa la distancia entre los objetos i y j en el espacio tD .
3. Se establece una regresión de d_{ij} en función de δ_{ij} . La regresión puede ser lineal, polinómica o monotónica. En el caso de la regresión líneal

$$d_{ij} = \alpha + \beta_{ij} + \varepsilon_{ij}$$

donde ε representa el "error", α y β son constantes.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Procedimiento para el escalado multidimensional

3. Las distancias que se obtienen de la ecuación de regresión:

$$\hat{d}_{ij} = \alpha + \beta_{ij}$$

se denominan "disparidades". Es decir, las disparidades \hat{d}_{ij} son las distancias de los datos (δ_{ij}) escaladas para coincidir con la distancia de la configuración (d_{ij}) tan estrechamente como sea posible.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Procedimiento para el escalado multidimensional

4. La bondad del ajuste entre las distancias de la configuración y las disparidades se mide por estadígrafos adecuados. Una posibilidad es la denominada "fórmula de estrés 1" de Kruskal:

$$STRESS1 = \sqrt{\frac{\sum (d_{ij} - \hat{d}_{ij})^2}{\sum \hat{d}_{ij}^2}}$$

Este estadígrafo es una medida de hasta que punto hay que "forzar" la configuración espacial de puntos para obtener las distancias de los datos (δ_{ij}).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

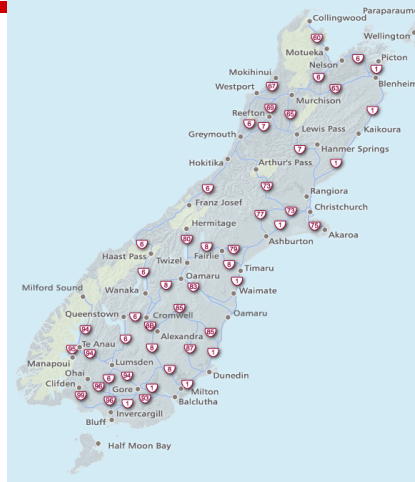
Procedimiento para el escalado multidimensional

5. Las coordenadas de cada objeto (x_1, x_2, \dots, x_t) se modifican suavemente de modo de reducir el "stress".
6. Los pasos 2 a 5 se repiten hasta que el "stress" no pueda reducirse más. **El resultado final del análisis son las coordenadas de los n objetos en t dimensiones.** Estas coordenadas pueden utilizarse para dibujar un mapa que muestre la forma en que los individuos están relacionados. Por lo tanto, es deseable encontrar una buena solución en 3D o 2D. Obviamente esto no siempre es posible.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej. 10.1 Distancias carreteras entre ciudades de Nueva Zelanda

Objetivo: Construir un "mapa" de la isla sur de Nueva Zelanda, a partir de una tabla de distancias carreteras entre 13 ciudades.



AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej. 10.1 Distancias carreteras entre ciudades de Nueva Zelanda

- Si las distancias carreteras fueran proporcionales a las distancias geográficas sería posible reconstruir el verdadero mapa exactamente. Sin embargo, debido a la ausencia de conexiones carreteras directas entre algunas ciudades, las distancias carreteras son en esos casos mayores que las distancias geográficas. Por consiguiente, sólo puede obtenerse una reconstrucción aproximada.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej. 10.1 Distancias carreteras entre ciudades de Nueva Zelanda

□ Matriz de distancias

	Alexandra	Balclutha	Blenheim	Christchurch	Dunedin	Franz Josef	Greymouth	Invercargill	Milford	Nelson	Queenstown	Te Anau	Timaru
Alexandra	0												
Balclutha	100	0											
Blenheim	485	478	0										
Christchurch	284	276	201	0									
Dunedin	126	50	427	226	0								
Franz Josef	233	493	327	247	354	0							
Greymouth	347	402	214	158	352	114	0						
Invercargill	138	89	567	365	139	380	493	0					
Milford	248	213	691	489	263	416	555	174	0				
Nelson	563	537	73	267	493	300	187	632	756	0			
Queenstown	56	156	494	305	192	228	341	118	178	572	0		
Te Anau	173	138	615	414	188	366	480	99	75	681	117	0	
Timaru	197	177	300	99	127	313	225	266	377	366	230	315	0

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej. 10.1 Distancias carreteras entre ciudades de Nueva Zelanda

XLSTAT 7.0 - Multidimensional Scaling (MDS) - 03/07/2006 at 11:41:42 a.m.
Dissimilarity matrix: workbook = New Zealand.xls / sheet = Sheet1 / range = \$B\$2:\$N\$14 / 13 rows and 13 columns
Uniform weighting (default)
No missing values
Metric Multidimensional Scaling
Multidimensional Scaling model: absolute
Stress used for the results: Kruskal's stress-1
Dimension of the representation space: 2
Repetitions: 10
Seed of the pseudo-random numbers generator: 3803407336
Iterations: 50
Convergence: 0,0001

Space with 2 Dimensions:

Model: $D_{ij} = P_{ij}$

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.