Techniques and Applications of Multivariate Analysis

7. Multidimensional Scaling (MDS)

Contents

7.1 Comprehension of MDS
7.2 Metric MDS
7.3 Non-Metric MDS
7.4 Three-way matrix MDS
7.5 Nonlinear MDS
7.6 R for MDS : Practice Time

7.1 Comprehension of MDS

- **[Definition] Choi (2014, Sec 1.1)**
  - Visualization Techniques for finding a low dimensional configuration space to match nearly the original similarities (or distances) between every pair of n objects.

- Applications
  - **Market research, Political sciences, Psychology, Physics, Text mining, etc.**

- **General relationship :**

$$d_{ij} = f(\delta_{ij}) + \varepsilon_{ij}$$

Distances in High Dimensional Space

Distances in Configuration Space

Relationship of d_{ij} and δ_{ij}

Scales

1) Absolute : $d_{ij} = \delta_{ij} + \varepsilon_{ij}$,

2) Interval : $d_{ij} = a + b\delta_{ij} + \varepsilon_{ij}$, $b > 0$

3) Ordinal : $d_{ij} = f(\delta_{ij}) + \varepsilon_{ij}$.

where ε_{ij} is a combination of errors of measurement and $f(\cdot)$ is non-decreasing function.

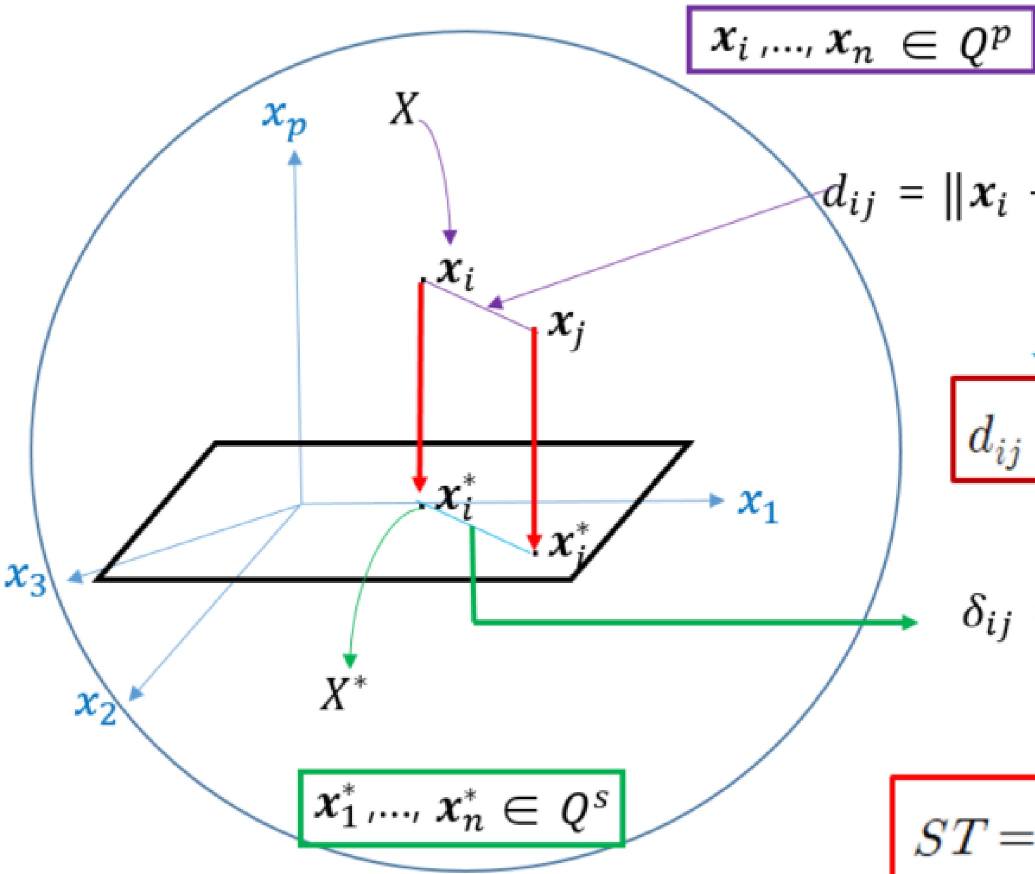
Torgerson Algorithm :

Metric MDS: classical scaling → cmdscale()

Non-metric MDS: ordinal scaling → isoMDS()

Kruskal-Shepard Algorithm :

Geometric Concept of MDS



$$d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \in Q^p$$

$$\Rightarrow n \times n \; D = (d_{ij})$$

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

$$\delta_{ij} = \|\boldsymbol{x}_i^* - \boldsymbol{x}_j^*\| \in Q^s$$

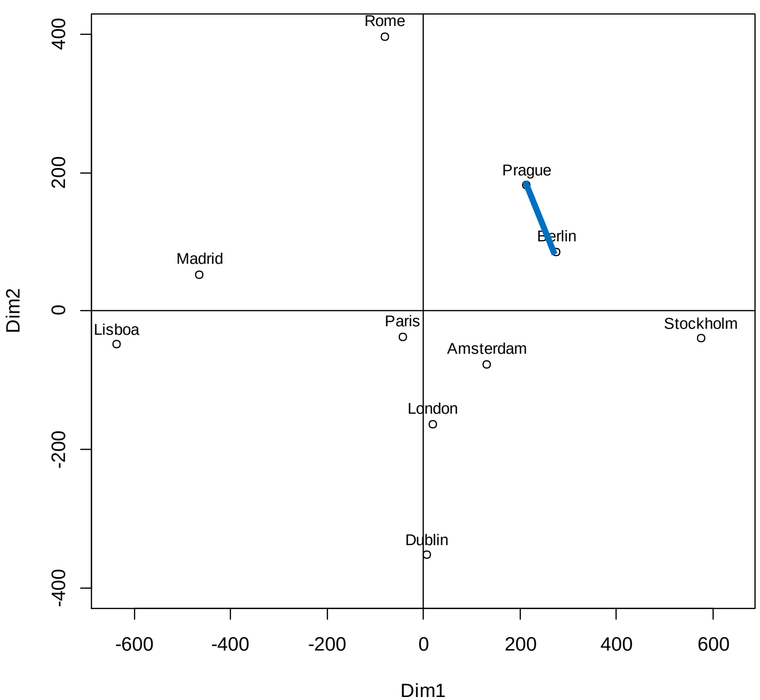
$$\boxed{ST = \left[\sum_{i<j} (d_{ij} - f(\delta_{ij}))^2 / \sum_{i<j} (d_{ij})^2 \right]^{1/2}}$$

$$\Rightarrow \quad Min_{f(\,.\,)} ST \Leftrightarrow Min_{X^*} ST$$

7.1 Comprehension of MDS

- **[Example 7.2.2] Metric MDS of distance between 10 European cities**
  - **[Data 7.2.1] Distance matrix of 10 European cities (eurodist.txt)**

London	0									
Stockholm	569	0								
Lisboa	667	1212	0							
Madrid	530	1043	201	0						
Paris	141	617	596	431	0					
Amsterdam	140	446	768	608	177	0				
Berlin	357	325	923	740	340	218	0			
Prague	396	423	882	690	337	272	114	0		
Rome	569	787	714	516	436	519	472	364	0	
Dublin	190	648	714	622	320	302	514	573	755	0



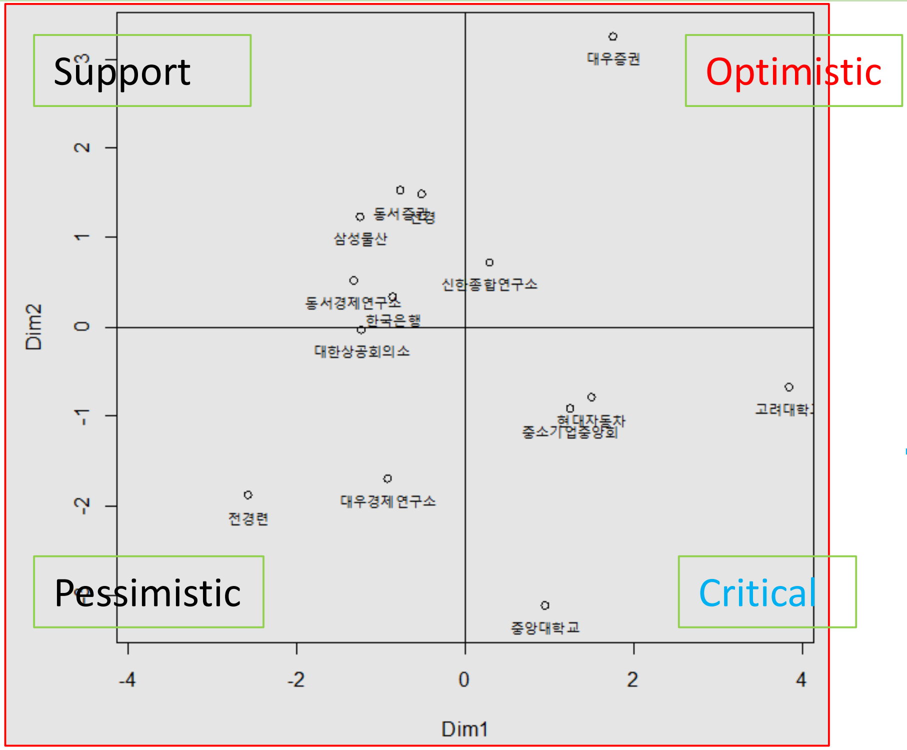
GOF = 99.91%

The relative positions of cities on the Map

7.1 Comprehension of MDS

- **[Example 7.2.3] Metric MDS on the economic view about the 1990 [R-code 7.2.3]**

- **[Table 7.2.2] Cluster and characteristics of MDS plot on economic view (data 1.3.5: economicview.txt)**



GOF = 55.44%

D : standardized Euclidean distance

Growth /Export/Import/ Black-ink balance/ Foreign debt /Exchange rate/ Unemployed rate /Consumer price rate /Wage increase rate

기관	성장률	GNP	수출	수입	국제 흑자	연말 외채	연말 환율	실업률	소비 물가	임금 상승
한국은행	8.2	1950	698	650	98	280	630	3.0	5.7	12.0
대우증권	9.5	2100	710	620	110	270	640	2.7	4.5	12.0
동서증권	9.0	2000	690	630	100	290	630	2.6	6.0	12.0
전경련	7.8	1850	668	660	88	280	620	3.0	6.4	13.8
대한상공회의소	8.5	1928	710	670	90	290	620	3.0	6.0	10.0
중소기업중앙회	9.0	1958	710	615	95	280	603	4.0	6.0	10.0
현대자동차	8.5	1900	700	610	100	250	620	3.2	5.5	14.0
삼성물산	8.0	1900	700	640	100	280	640	2.7	5.5	10.0
선경	8.5	1950	700	620	120	300	630	2.7	7.0	12.0
대우경제연구소	7.9	1900	697	645	95	280	610	2.9	6.8	15.0
신한종합연구소	9.0	2030	700	620	100	275	630	3.0	7.0	13.0
동서경제연구소	8.5	1950	690	630	90	290	630	2.9	6.0	12.0
고려대학교	9.9	1870	729	649	123	260	616	3.4	6.1	15.8
중앙대학교	8.5	1700	700	630	90	260	620	4.0	6.0	14.0

Low High

사분면	기 관	특 성
제1	대우증권, 신한종합연구소	낙관론
제2	동서증권, 선경, 삼성물산, 동서경제연구소, 한국은행	정책옹호
제3	대한상공회의소, 전경련, 대우경제연구소	비관론
제4	현대자동차, 중소기업중앙회, 고려대학교, 중앙대학교	정책비판

7.1 Comprehension of MDS

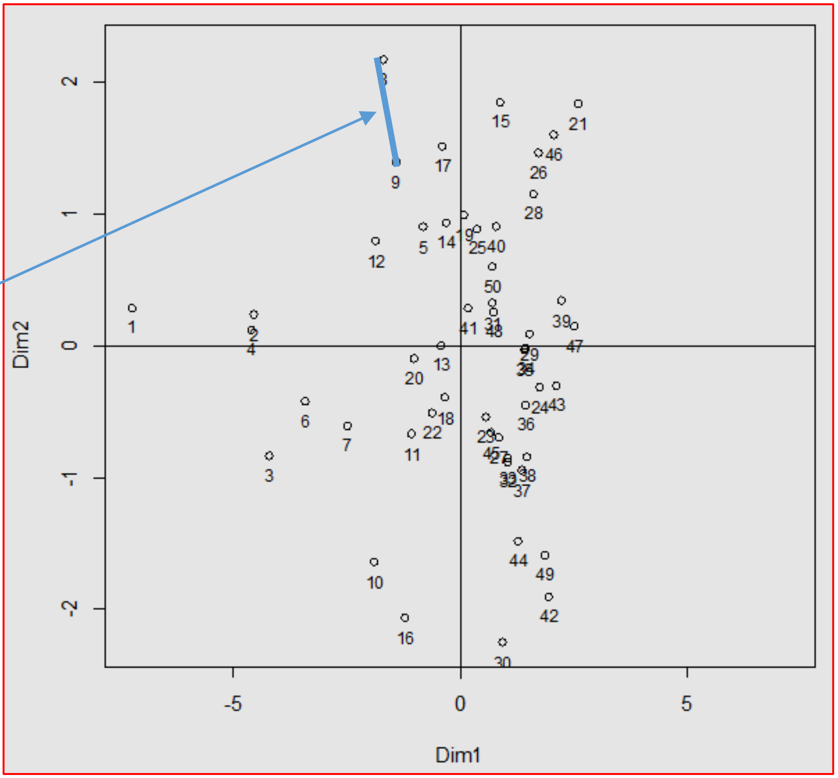
- **[Example 7.2.4] Metric MDS of KLPGA Data** D : standardized Euclidean distance
- **[Data 1.3.2] KLPGA player's grade (klpga.txt)**

GOF = 90.48%

Putting average /Green in regulation % / Par save % / Par break % / Scoring average Prize rate

선수	평균퍼팅수	그린적중률	파세이브율	파브레이크율	평균타수	상금률
1	30.36	82.72	90.12	23.77	69.58	100.0
2	30.85	76.94	85.29	23.69	70.85	63.7
3	31.33	79.63	88.52	19.26	70.93	59.4
4	30.64	79.32	87.65	21.30	70.47	50.2
5	30.97	68.86	79.97	17.17	72.79	44.0
6	31.09	78.59	86.03	20.48	71.23	38.0
7	31.25	77.47	86.88	16.82	71.61	30.5
8	29.81	69.53	85.94	15.63	71.84	21.8
9	30.36	70.37	83.95	16.82	71.97	21.1
10	31.91	78.28	87.71	15.32	71.73	19.2
⋮						
48	31.20	70.19	80.19	14.44	73.53	6.2
49	32.50	71.35	78.13	12.67	74.06	5.9
50	31.07	68.72	80.36	14.88	73.39	5.1

축	선 수	특 성
제1축 왼편	1, 2, 3, 4, 6, 7, 8, 9, 10	상위권
제1축 가운데	5, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	중위권
제1축 오른편	30위-50위	하위권



Top Middle Low

7.2 Metric MDS

[Table 7.1.1] Process Steps for MDS : Metric MDS or Non-metric MDS

[Step 1] Obtain the $n \times n$ distance matrix $D = (d_{ij})$, $i, j = 1, \ldots, n$ based on the dissimilarities among objects.

[Step 2] Select a MDS algorithm.

[Step 3] Calculate a goodness-of-fit of MDS.

[Step 4] Interpret the MDS plot and attempt cluster of objects.

How well dose MDS plot represent the data?

- **[Example 7.2.2] Metric MDS of distance between 10 European cities**
  - **[Data 7.2.1] Distance matrix of 10 European cities (euroddist.txt) :**
    D = Euclidian distance on a map of Europe

    GOF = 99.91%

- **[Example 7.2.3] Metric MDS on the economic view**
  - **[Data 7.2.1] Distance matrix of 10 Institutes (economicview.txt)**
    D = Standardized Euclidian distance from economicview.txt

    GOF = 55.44%

- **[Example 7.2.4] Metric MDS of KLPGA Data**
  - **[Data 1.3.2] KLPGA player's grade (klpga.txt)**
    D = Standardized Euclidian distance from klpga.txt

    GOF = 90.48%

7.2 Metric MDS

Torgerson, W.S.(1952) Lincoln Laboratories, Massachusetts Institute of Technology, USA

[Table 7.2.1] Torgerson Algorithm for Metric MDS

[Step 1] Calculate an $n \times n$ matrix $\boxed{A = (a_{ij})}$ from $D = (d_{ij})$, $i, j = 1, \ldots, n$

where, $\boxed{a_{ij} = -\frac{1}{2} d_{ij}^2}$, $i, j = 1, \ldots, n$.

[Step 2] Calculate $\boxed{B = (b_{ij})} = HAH$ from $A = (a_{ij})$

where, $\boxed{b_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}}$

$$a_{ij} = -\frac{1}{2} d_{ij}^2, \quad \overline{a}_{i.} = \frac{1}{n} \sum_{j=1}^{n} a_{ij}, \quad \overline{a}_{.j} = \frac{1}{n} \sum_{i=1}^{n} a_{ij}, \quad \overline{a}_{..} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}.$$

$H = I_n - J_n / n$: centering matrix

[Step 3] Calculate spectral decomposition: $B = V D_l V^t$

where, $D_l = diag(l_1, \ldots, l_n)$: Diagonal matrix with $l_1 \geq \cdots \geq l_n > 0$,

$V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$: Orthogonal matrix

[Step 4] Obtain an $n \times s$ coordinate matrix:

$$C_{(s)} = V_{(s)} D_{l(s)}^{1/2} = (\boldsymbol{v}_1 l_1^{1/2}, \cdots, \boldsymbol{v}_s l_s^{1/2})$$

[Step 5] Goodness- of-fit: $fit_1 = \dfrac{\sum_{i=1}^{s} l_i^2}{\sum_{i=1}^{n} l_i^2} \times 100\%$ for $B \geq 0$

Deriving B from $X=(x_{ik})$ and $A=(a_{ij})$:

Consider the centred data matrix $X=(x_{ik})$, $i=1,\dots,n$; $k=1,\dots,p$ with $\Sigma_{i=1}^{n} x_{ij}=0$.

Let $\boldsymbol{x}_i=(x_{i1},\dots,x_{ip})^T$ be ith observation which is row of X and coordinate of data points in p-dimensional space.

Then the squared Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is given by

$$\boxed{d_{ij}^2 = (\boldsymbol{x}_i-\boldsymbol{x}_j)^T(\boldsymbol{x}_i-\boldsymbol{x}_j)}$$

$$= \boldsymbol{x}_i^T\boldsymbol{x}_i + \boldsymbol{x}_j^T\boldsymbol{x}_j - 2\boldsymbol{x}_i^T\boldsymbol{x}_j. \tag{7.2.2}$$

Let the inner product matrix be $B=(b_{ij})$ where $\boxed{b_{ij} = \boldsymbol{x}_i^T\boldsymbol{x}_j.}$

Hence from (7.2.2)

$$\boxed{d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}} \tag{7.2.3}$$

Deriving B from $X=(x_{ik})$ and $A=(a_{ij})$:

From (7.2.3), we have

$$\frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 = b_{ii} + \frac{1}{n}\sum_{j=1}^{n} b_{jj},$$

$$\frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} b_{ii} + b_{jj},$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} b_{ii} + \frac{1}{n}\sum_{j=1}^{n} b_{jj}$$

From (7.2.3),

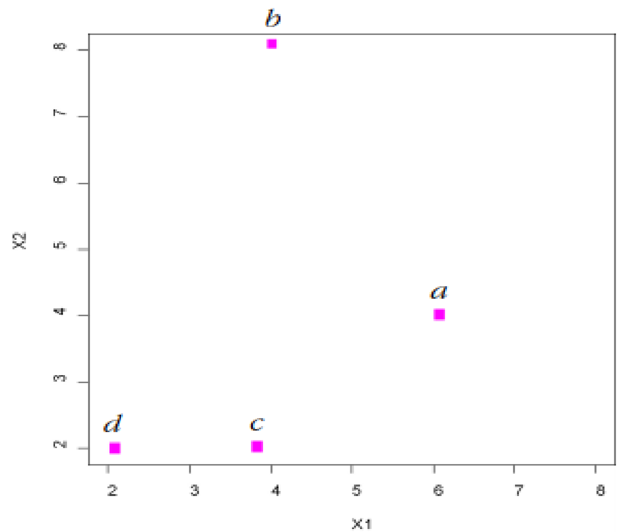
$$\begin{aligned} b_{ij} &= -\frac{1}{2}\left(d_{ij}^2 - b_{ii} - b_{jj}\right) \\ &= -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\right) \\ &= \boxed{a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}} \end{aligned}$$

where $a_{ij} = -\frac{1}{2}d_{ij}^2$, $\bar{a}_{i.} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$, $\bar{a}_{.j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$, $\bar{a}_{..} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$.

7.2 Metric MDS

- **[Example 7.2.1] Scatter plot, 3-dim spatial representation of distance, and MDS for 4x2 X's**

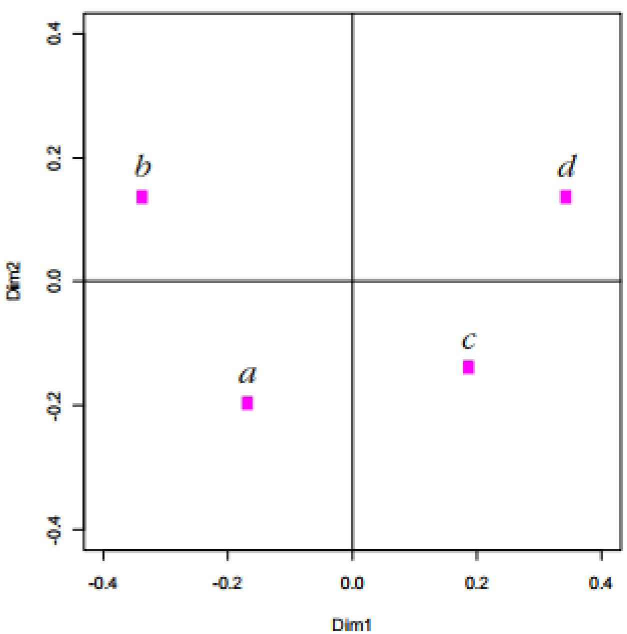
$$X = \begin{bmatrix} 6 & 4 \\ 4 & 8 \\ 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$



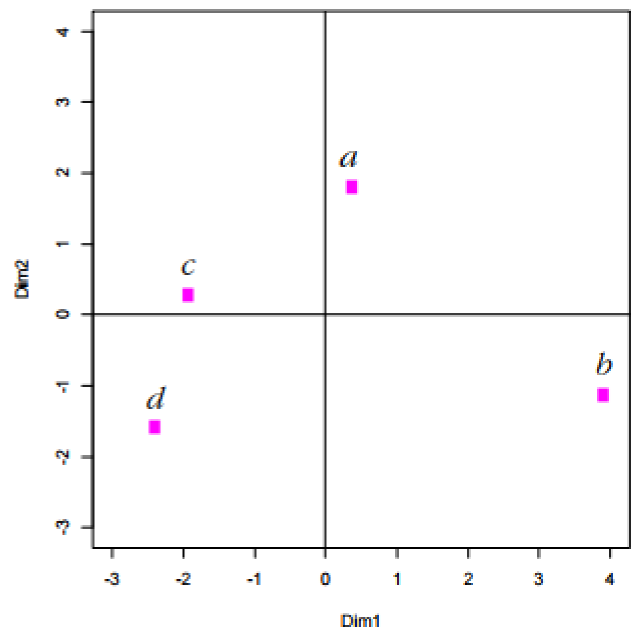
X's 2-dim scatter plot

distance $d_{ij},\ i,j=1,\ldots,4$	Clark Distance $d_{ij} = \sqrt{\sum_{k=1}^{2}\left(\frac{x_{ik}-x_{jk}}{x_{ik}+x_{jk}}\right)^2}$	Euclidean Distance $d_{ij} = \sqrt{\sum_{k=1}^{2}(x_{ik}-x_{jk})^2}$
dissimilarity matrix $D=(d_{ij})$	$\begin{bmatrix} 0 & 0.389 & 0.389 & 0.601 \\ 0.389 & 0 & 0.600 & 0.686 \\ 0.389 & 0.600 & 0 & 0.333 \\ 0.601 & 0.686 & 0.333 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 4.472 & 2.828 & 4.472 \\ 4.472 & 0 & 6.000 & 6.325 \\ 2.828 & 0.600 & 0 & 2.000 \\ 4.472 & 6.325 & 2.000 & 0 \end{bmatrix}$

7.2 Metric MDS - [Figure 7.2.1]



Metric MDS plot of **Clark dist**



Metric MDS plot of **Euclid dist**

GOF = 96.39%

GOF = 100%

Torgerson Algorithm for Metric MDS

7.2 Metric MDS

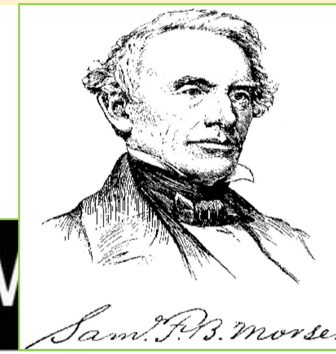
- **[Example 7.2.6] Metric MDS of Morse Code**

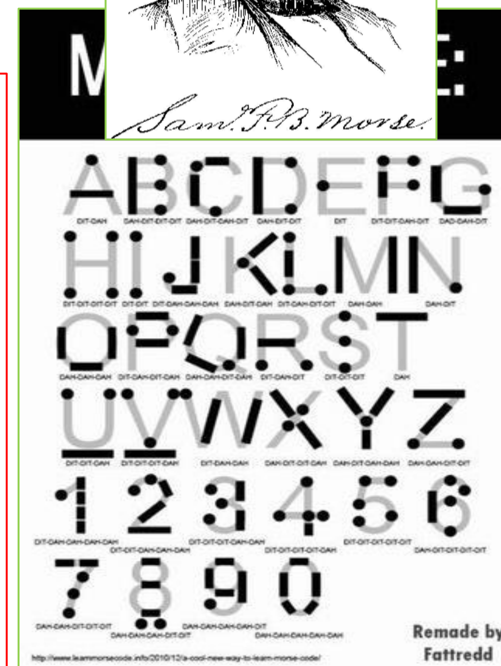
- [Data 7.2.3] Similarity Matrix $C = (c_{ij})$ of Morse code(morse.txt)

- 598 subjects determined whether the digits of each pair were the same or different. ➡ % of times in correct

	1	2	3	4	5	6	7	8	9	0
1.----	84									
2..---	62	89								
3...--	16	59	86							
4....-	6	23	38	89						
5.....	12	8	27	56	90					
6-....	12	14	33	34	30	86				
7--...	20	25	17	24	18	65	85			
8---..	37	25	16	13	10	22	65	88		
9----.	57	28	9	7	5	8	31	58	91	
0-----	52	18	9	7	5	18	15	39	79	94

- Each pair of signals was presented twice(1 then 2, and 2 then 1).
- dots(.) & dashes(-) are coded for digits 0 ~ 9.





7.2 Metric MDS

- **[Example 7.2.6] Metric MDS of Morse Code**

GOF = 56.22%

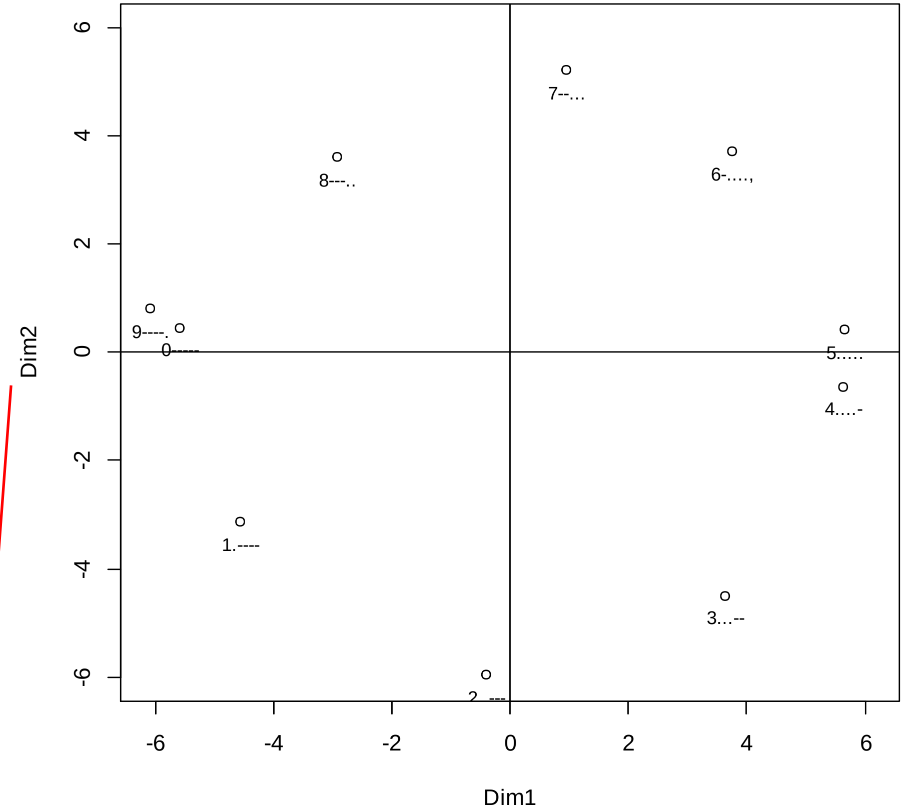
$$d_{ij} = \sqrt{c_{ii} - 2c_{ij} + c_{jj}}$$

$$\Rightarrow \quad D = (d_{ij}),\ i,\ j = 1,\ \ldots,\ 10$$

```
$points
          [,1]       [,2]
 [1,] -4.5683272 -3.1377040
 [2,] -0.4079107 -5.9626237
 [3,]  3.6389106 -4.4990989
 [4,]  5.6363286 -0.6275020
 [5,]  5.6464513  0.4205213
 [6,]  3.7426401  3.7170866
 [7,]  0.9389537  5.2143848
 [8,] -2.9284754  3.6160868
 [9,] -6.0970041  0.8048214
[10,] -5.6015668  0.4540277

$GOF
[1] 0.5582581 0.5622199
```

- **[Figure 7.2.6] Metric MDS and MDS Map**



Heterogeneity of the signal

#s of . are increasing

7.3 Non-Metric MDS

Kruskal, J.B. (January 29, 1928 – September 19, 2010) : American mathematician, statistician, computer scientist and psychometrician.

Shepard, R.N. (January 30, 1929 – May 30, 2022), USA,.Cognitive Scientist

[Table 7.3.1] Kruskal -Shepard Algorithm for Non-metric MDS

[STEP 1] Order $m = n(n-1)/2$ non-diagonal elements from $D = (d_{ij})$ in ascending as

$$d_{i_1j_1} \le d_{i_2j_2} \le \cdots \le d_{i_mj_m}$$

where the subscript indicates the pair of objects that are at least similar

[STEP 2] If $D^* = (\delta^*_{ij})$ is a dissimilarity matrix related with a coordinate matrix X^* of s-dimensional configuration space, $\delta^*_{i_1j_1}, \delta^*_{i_2j_2}, \ldots, \delta^*_{i_mj_m}$ must be corresponded to $d_{i_1j_1}, d_{i_2j_2}, \ldots, d_{i_mj_m}$, respectively.

But these can not guarantee original ordering : $d_{i_1j_1} \le d_{i_2j_2} \le \cdots \le d_{i_mj_m}$

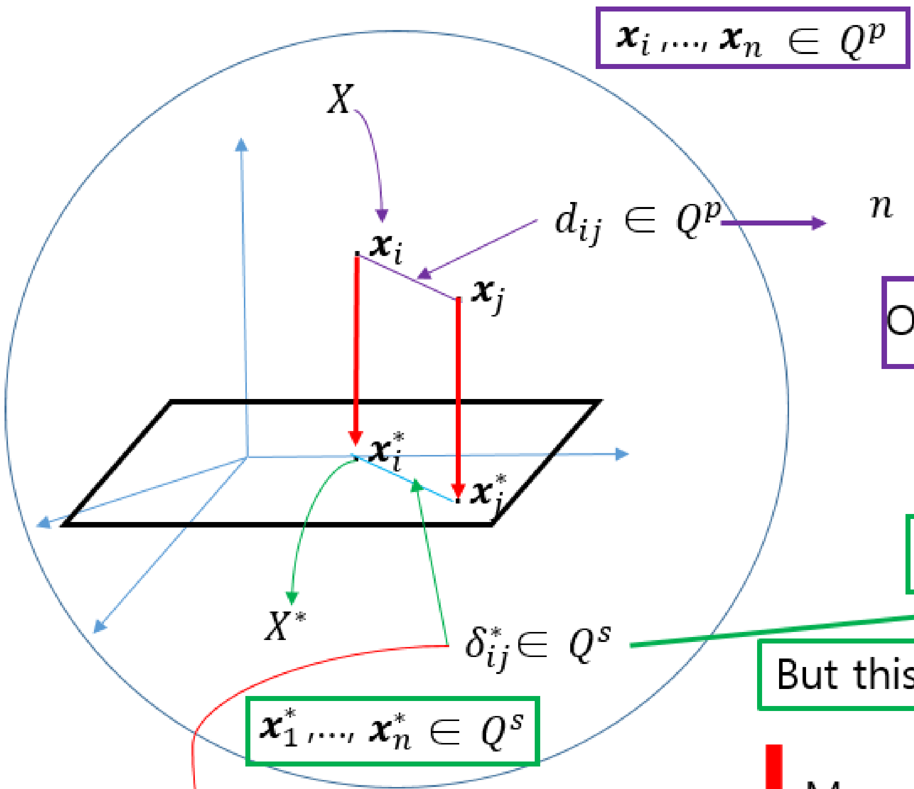
[STEP 3] Calculate the rank image $\hat d_{i_1j_1} \le \hat d_{i_2j_2} \le \cdots \le \hat d_{i_mj_m}$ which is monotonic related with $\delta^*_{i_1j_1} \le \delta^*_{i_2j_2} \le \cdots \le \delta^*_{i_mj_m}$ and non-decreasing with $d_{i_1j_1} \le d_{i_2j_2} \le \cdots \le d_{i_mj_m}$

[STEP 4] Define the Kruskal's stress :

$$\left[\sum_{i<}\sum_{j} (\delta^*_{ij} - \hat d_{ij})^2 / \sum_{i<}\sum_{j} (\delta^*_{ij})^2 \right]^{1/2} \qquad (7.3.1)$$

and find a coordinate matrix X^* of s-dimensional configuration space minimizing (7.3.1).

Geometric Concept of Non-metric MDS



$$n \times n \;\; D = (d_{ij}) : m = \frac{n(n-1)}{2}$$

Ordering : $d_{ij(1)} \leq \cdots \leq d_{ij(m)}$

Hope : $\delta^*_{ij(1)} \leq \cdots \leq \delta^*_{ij(m)}$

But this 1-1 ^ onto ordering can not be guaranteed.

Monotonic non-decreasing with $d_{ij(1)} \leq \cdots \leq d_{ij(m)}$

from $\delta^*_{ij(1)} \leq \cdots \leq \delta^*_{ij(m)}$

Rank Image : $\hat{d}_{ij(1)} \leq \cdots \leq \hat{d}_{ij(m)}$

$$ST = \left[\sum_{i<} \sum_{j} \left(\delta^*_{ij} - \hat{d}_{ij}\right)^2 \Big/ \sum_{i<} \sum_{j} \left(\delta^*_{ij}\right)^2 \right]^{1/2}$$

7.3 Non-Metric MDS

❖ **Stress (Kruskal and Wish, 1978, pp. 24-25)**

$$ST = (RSS / SC)^{1/2}$$

$$= \left[\sum_{i<} \sum_{j} \left(\delta^*_{ij} - \hat{d}_{ij}\right)^2 \Big/ \sum_{i<} \sum_{j} \left(\delta^*_{ij}\right)^2 \right]^{1/2}$$

→ $Min_{X^*} ST$ for s-dimensional Non-metric MDS Map.

❖ **Kruskal(1964)'s Criterion for the number of dimensions**

Stress(%)	0.0	2.5	5.0	10.0	20.0
Goodness of fit	Perfect	Excellent	Good	Fair	Poor

How well dose Non-metric MDS plot represent the data?

7.3 Non-Metric MDS

- **[Example 7.3.1] Non-Metric MDS for Dissimilarity Matrix Data among 9 Jobs**

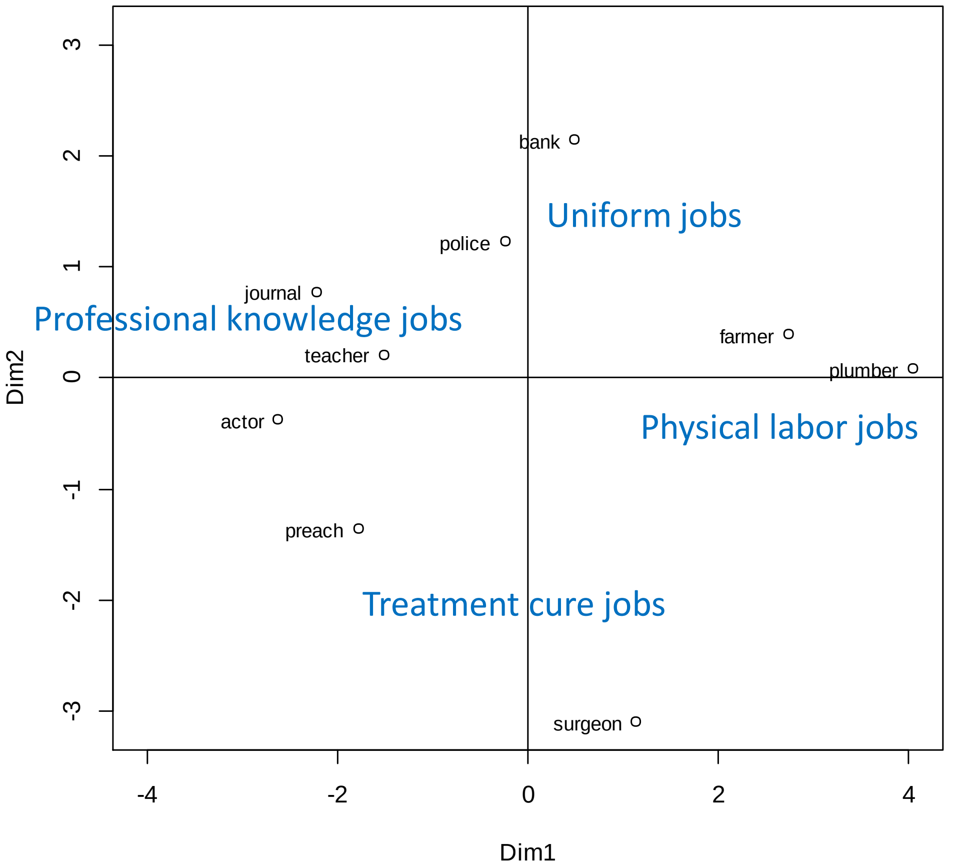
[Data 7.3.1] Dissimilarity Matrix (jobs.txt)

Preacher	0.00								
Surgeon	3.06	0.00							
Teacher	2.14	3.95	0.00						
Journalist	3.21	3.95	3.03	0.00					
Policeman	3.51	4.17	2.82	3.33	0.00				
Plumber	4.40	3.77	3.86	4.14	3.60	0.00			
Farmer	3.64	3.69	3.47	3.90	3.56	2.53	0.00		
Actor	3.12	4.13	2.72	2.58	3.68	4.17	4.10	0.00	
Bank clerk	3.73	4.05	3.31	3.46	3.29	3.92	3.59	3.72	0.00

- 0(very similar) ~ 5(very different) : 36 cases of pair of jobs
- Average the rating of 62 people

[Figure 7.3.1] Non-metric MDS Map

GOF = 5.60% **Good**



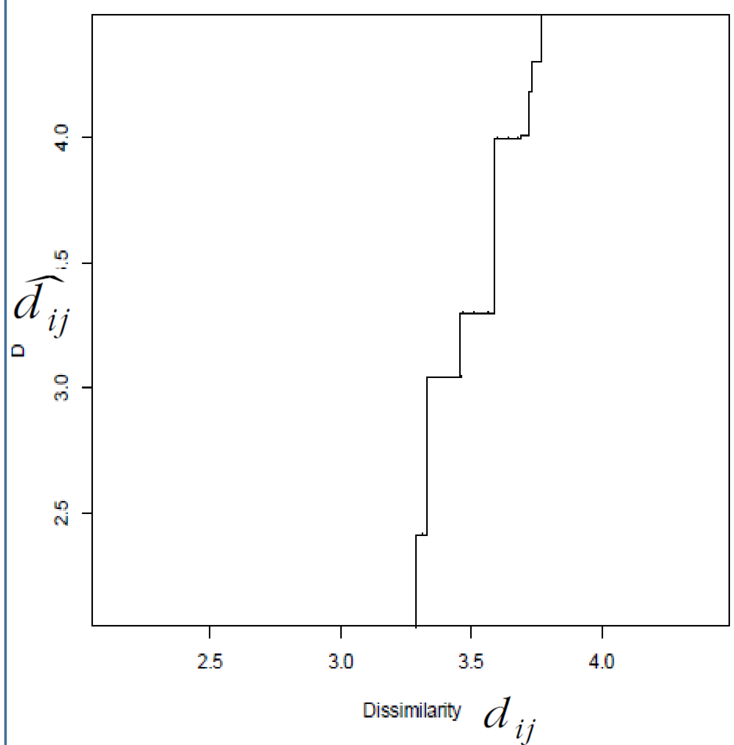
7.3 Non-Metric MDS

- **[Example 7.3.1] Non-Metric MDS for Dissimilarity Matrix Data among 9 Jobs**

GOF = 5.60% : Good

Non-metric MDS plot represents the data good !

	d_{ij} 원거리	δ^*_{ij} FITDIST	$\hat{d}_{ij}$ FITDATA
[1,]	2.14	1.5972510	1.326969
[2,]	2.53	1.3432663	1.326969
[3,]	2.58	1.2215031	1.326969
[4,]	2.72	1.2567262	1.326969
[5,]	2.82	1.6318242	1.326969
[6,]	3.03	0.9112457	1.326969
[7,]	3.06	3.3988974	2.012820
[8,]	3.12	1.2946962	2.012820
[9,]	3.21	2.1861394	2.012820
[10,]	3.29	1.1715476	2.012820
--- 생략 ---			
[35,]	4.17	6.6752779	6.335556
[36,]	4.40	5.9958335	6.335556



Shepard Diagram

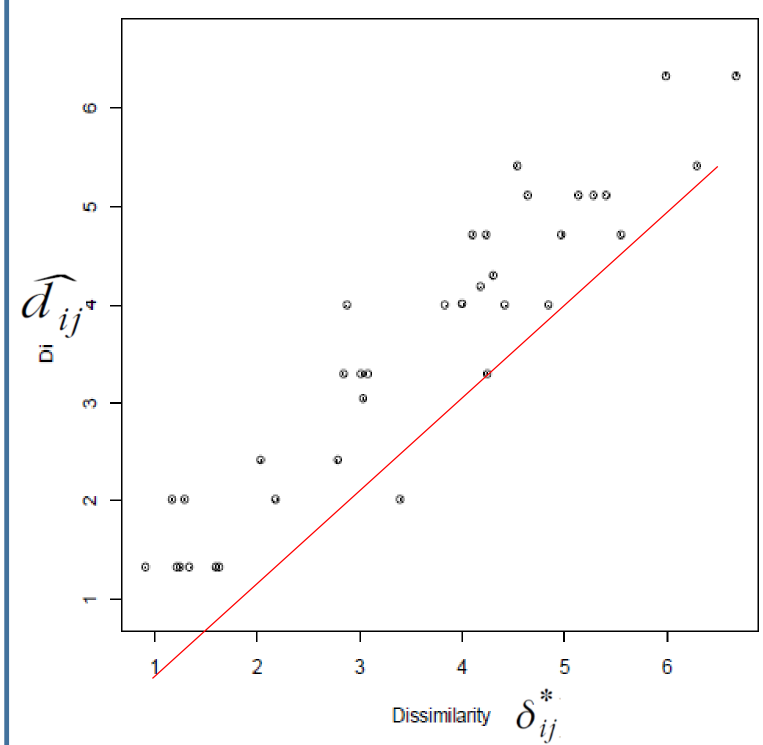


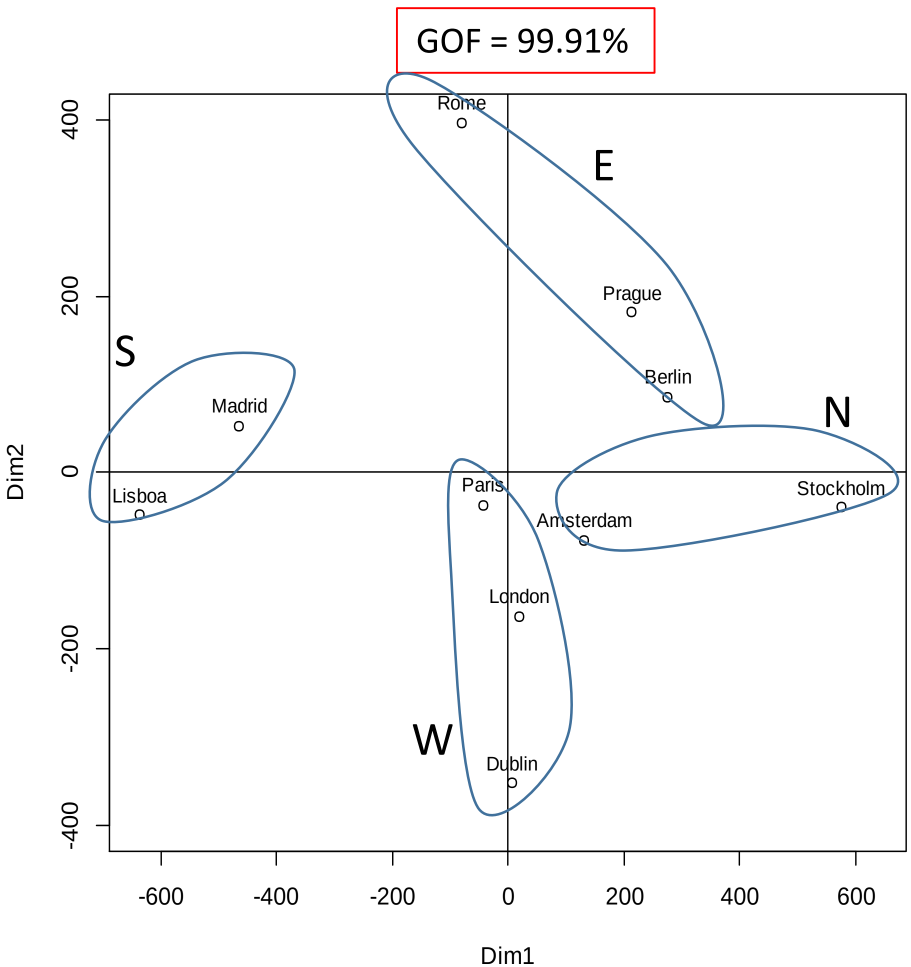
Image Diagram

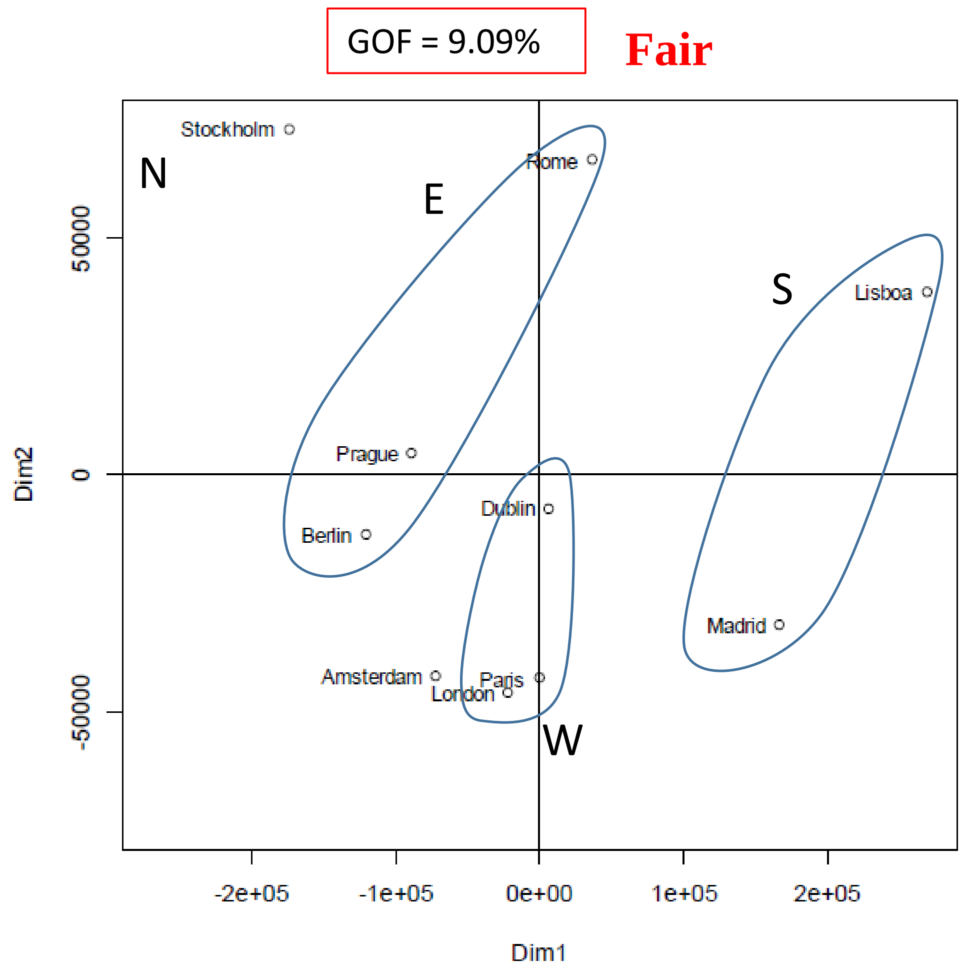
[Figure 7.3.1] Guttman(1968)' Shepard Diagram and Image Diagram

7.3 Non-Metric MDS

- **[Example 7.3.2] Non-Metric MDS of distance between 10 European cities**

[Figure 7.2.2] Metric MDS Map	[Figure 7.3.2] Non-metric MDS Map





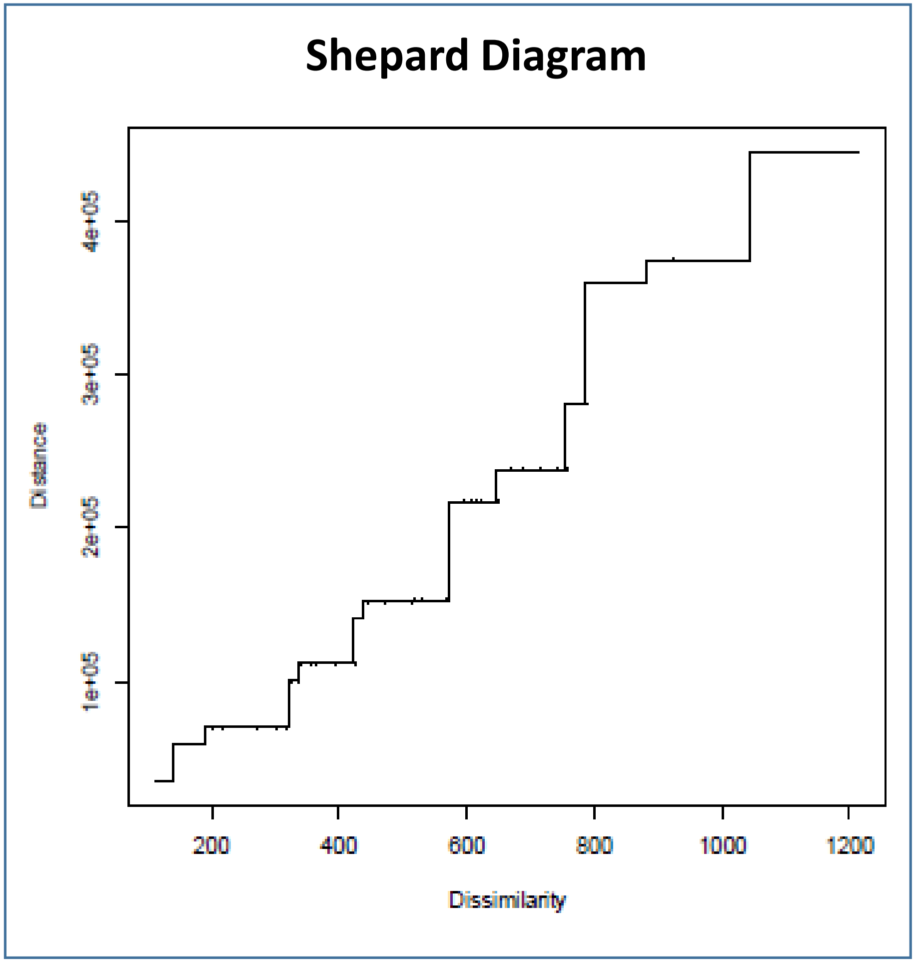
Similar geographical positions but Metric MDS's performance is better than No-metric MDS's

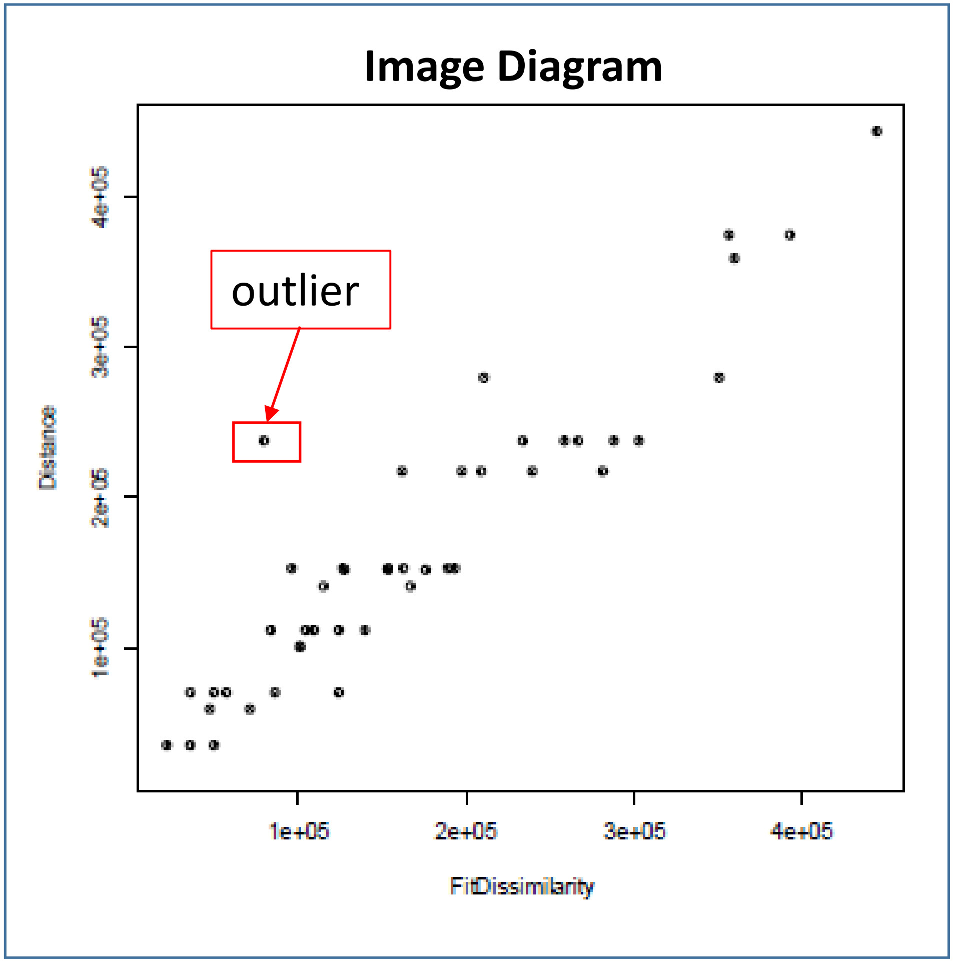
7.3 Non-Metric MDS

- **[Example 7.3.2] Non-Metric MDS of distance between 10 European cities**

GOF = 9.09% : Fair

Non-metric MDS plot represents the data fairly !



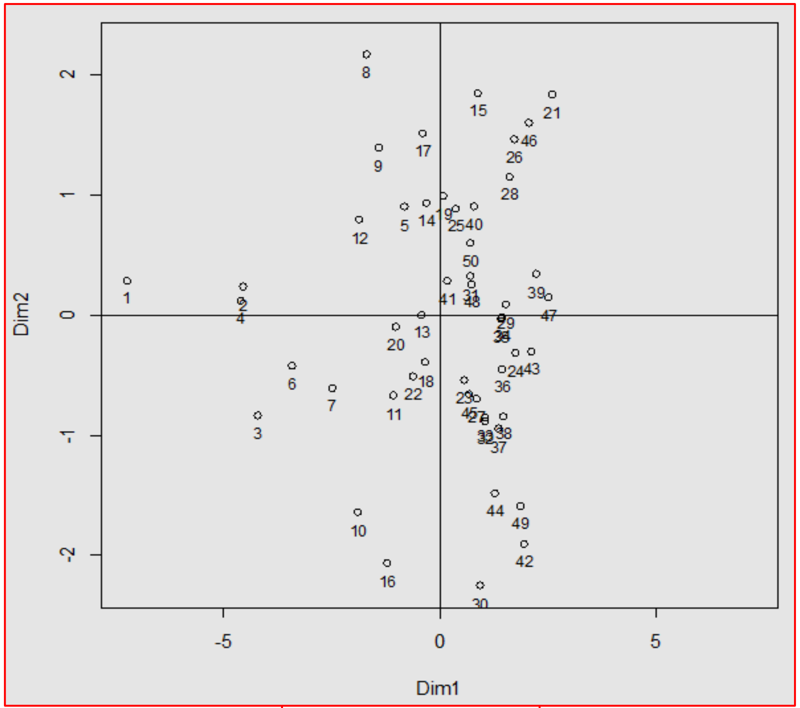


[Figure 7.3.2] Shepard Diagram and Image Diagram

7.3 Non-Metric MDS

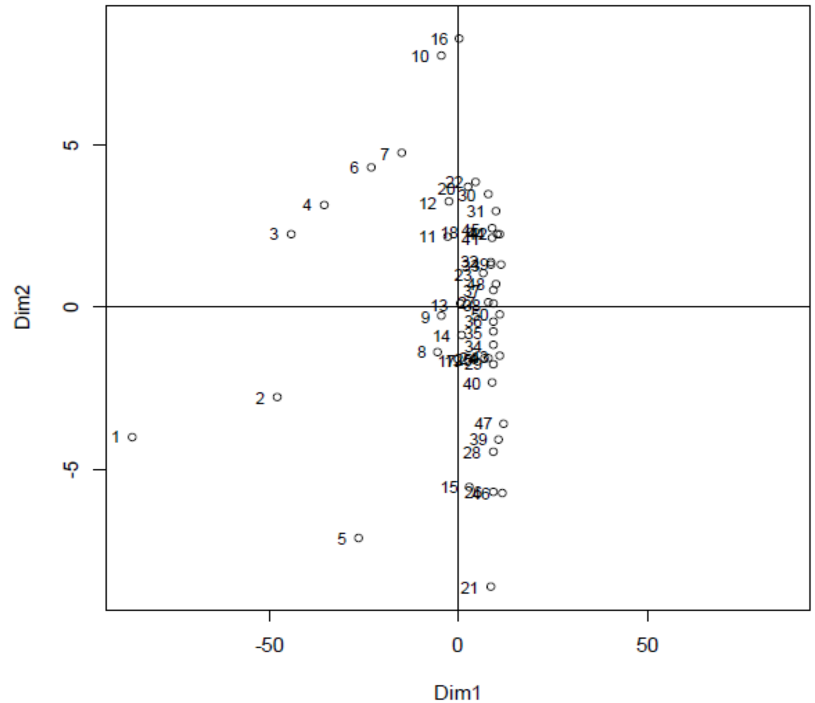
- **[Example 7.3.3] Non-Metric MDS of KLPGA Data**

[Figure 7.2.4] Metric MDS Map



GOF = 90.48%

[Figure 7.3.3] Non-metric MDS Map



GOF = 1.86% **Excellent**

축	선 수	특 성
제1축 왼편	1, 2, 3, 4, 6, 7, 8, 9, 10	상위권
제1축 가운데	5, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	중위권
제1축 오른편	30위-50위	하위권

Non-metric MDS plot represents the data excellently!

7.6 R for MDS : Practice Time

R-Code:

Metric MDS	**Non-metric MDS**
cmdscale()	library(MASS) isoMDS()
Torgerson Algorithm : Spectral Decomposition	Shepard() Kruskal-Shepard Algorithm

7.6 R for MDS : Practice Time

R-code list of Chapter 7 Cluster Analysis

example721-metricMDS.R	[R-코드 7.2.1]	4×2 자료행렬 X의 계량형 MDS
eurodist-metricMDS.R	[R-코드 7.2.2]	유럽 10대 도시간 거리의 계량형 MDS
economicview-metricMDS.R	[R-코드 7.2.3]	경제전망 자료의 표준화 유클리드 거리에 대한 계량형 MDS
klpga-metricMDS.R	[R-코드 7.2.4]	KLPGA 선수의 성적에 대한 계량형 MDS
bankbinary-metricMDS.R	[R-코드 7.2.5]	은행 경영평가 자료의 계량형 MDS
morse-metricMDS.R	[R-코드 7.2.6]	모스부호 자료의 계량형 MDS
jobs-nonmetricMDS.R	[R-코드 7.3.1]	9가지 직업의 비계량형 MDS
eurodist-nonmetricMDS.R	[R-코드 7.3.2]	유럽 10대 도시간 거리의 비계량형 MDS
klpga-nonmetricMDS.R	[R-코드 7.3.3]	KLPGA 선수의 성적에 대한 비계량형 MDS

7.6 R for MDS : Practice Time

[R-code 7.2.1] Metric MDS for 4 x 2 Data Matrix based on the Torgerson Algorithm

```
#[보기] 7.2.1] Metric MDS based ont the Torgerson's Algorithm
#Clark distance from data matrix
X <- matrix(c(6,4,4,8,4,2,2,2),byrow=T, nrow=4)
        n <- nrow(X)
rownames(X) <- c("a","b","c","d")
# Scatter Plot
par(pty="s")
lim <- range(pretty(X))
plot(X[,1],X[,2],xlab="X1",ylab="X2",xlim=lim,ylim=lim,pch=15,col=2)
abline(v=0,h=0)
text(X[,1],X[,2],rownames(X),cex=1,col=2,pos=3)
```

```
# Clark's distance
D <- matrix(0,n,n)
for (i in 1:n) { for (j in 1:n) {
tmp <- as.matrix((X[i,]-X[j,])/(X[i,]+X[j,]))
d <- sqrt(t(tmp)%*%tmp)
D[i,j] <- d}}
round(D, 3)
```

```
#Metric MDS: Torgerson's Algorithm
        A <- -(D^2)/2
        J <- matrix(1,n,n)
        H <- diag(1,n)-J/n
        B <- H%*%A%*%H
        eg.B <- eigen(B)
        tmp <- sum(eg.B$values>=0)
        eig <- eg.B$values[1:tmp]
        per <- eig/sum(eig)*100
        gof <- sum(per[1:2])
        V <- eg.B$vectors[,1:2]
        C <- V%*%diag(sqrt(eig[1:2]))
list(eig, gof)
# MDS Map
par(pty="s")
        lim <- range(pretty(C))
        plot(C[,1],C[,2],xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=15,col=2)
        abline(v=0,h=0)
        text(C[,1],C[,2],rownames(X),cex=1,col=2,pos=3)
```

7.6 R for MDS : Practice Time

[R-code 7.2.2] Metric MDS for **distance between 10 European cities**

```
# 10 European Cities's Distances
Data7.2.1<-read.table("eurodist.txt", header=T)
D<-Data7.2.1
도시=colnames(D)
n<-nrow(D)

# Metric MDS
win.graph()
con<-cmdscale(D, k=2, eig=T)
con
x<-con$points[, 1]
y<-con$points[, 2]
lim1<-c(-max(abs(x)), max(abs(x)))
lim2<-c(-max(abs(y)), max(abs(y)))

plot(x, y, xlab="Dim1", ylab="Dim2", xlim=lim1, ylim=lim2)
text(x, y, 도시, cex=0.8, pos=3)
abline(v=0, h=0)
```

7.6 R for MDS : Practice Time

[R-code 7.3.1] Non-metric MDS for **9 Jobs**

```
# Nine Jobs's Dissimilarity
Data7.3.1<-read.table("jobs.txt", header=T)
D<-Data7.3.1
직업=colnames(D)

# Nonmetric MDS
win.graph()
library(MASS)
con<-isoMDS(D, k=2)
con
x<-con$points[,1]
y<-con$points[,2]
lim1<-c(-max(abs(x)), max(abs(x)))
lim2<-c(-max(abs(y)), max(abs(y)))
plot(x,y, xlab="Dim1", ylab="Dim2", xlim=lim1, ylim=lim2)
text(x,y,직업, cex=0.8, pos=2)
abline(v=0, h=0)

# Shepard Diagram
win.graph()
```

```
dist_sh <- Shepard(D[lower.tri(D)], con$points)
dist_sh
cdist_sh=cbind(dist_sh$x, dist_sh$y, dist_sh$yf)
```

cdist_sh d_{ij} $\widehat{d}_{ij}$

```
plot(cdist_sh[,1], cdist_sh[,3], pch = ".", xlab = "Dissimilarity", ylab = "Distance",
xlim = range(cdist_sh[,1]), ylim = range(cdist_sh[,1]))
lines(cdist_sh[,1], cdist_sh[,3], type = "S")

# Image Diagram
```

win.graph() δ^*_{ij} $\widehat{d}_{ij}$

```
plot(cdist_sh[,2], cdist_sh[,3], pch = ".", xlab = "FitDissimilarity", ylab = "Distance",
xlim = range(cdist_sh[,2]), ylim = range(cdist_sh[,2]))
lines(cdist_sh[,2], cdist_sh[,3], type = "p")
```