Multidimensional Scaling

Leonardo Araujo

April 16, 2010

Outline I

Multidimensional scaling

- Why multidimensional scaling?
- Multidimensional scaling in psychophysics.

Multidimensional scaling

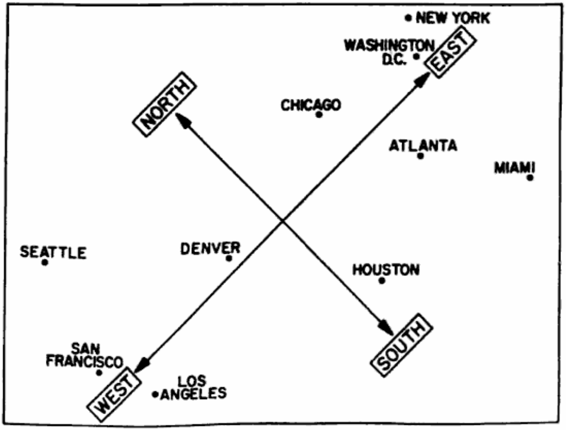
- Why multidimensional scaling?
- Multidimensional scaling in psychophysics.

Example - distance between cities

CITIES	ATLA.	CHIC.	DENV.	HOUS.	L.A.	MIAMI	N.Y.	S.F.	SEAT.	WASH D.C.
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO	587		920	940	1745	1188	713	1858	1737	597
DENVER	1212	920		879	831	1726	1631	949	1021	1494
HOUSTON	701	940	879		1374	968	1420	1645	1891	1220
LOS ANGELES	1936	1745	831	1374		2339	2451	347	959	2300
MIAMI	604	1188	1726	968	2339		1092	2594	2734	923
NEW YORK	748	713	1631	1420	2451	1092		2571	2408	205
SAN FRANCISCO	2139	1858	949	1645	347	2594	2571		678	2442
SEATTLE	2182	1737	1021	1891	959	2734	2408	678		2329
WASHINGTON D.C.	543	597	1494	1220	2300	923	205	2442	2329	

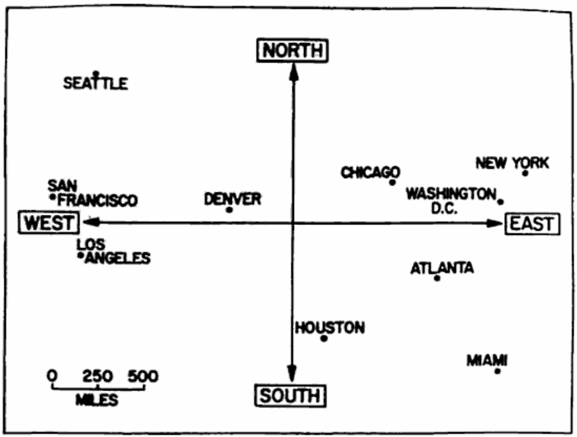
(B) AIRLINE DISTANCES BETWEEN TEN U.S. CITIES

Example - distance between cities



(C) CONFIGURATION OBTAINED BY APPLYING MULTIDIMENSIONAL SCALING TO THE AIRLINE DISTANCES

Example - distance between cities



(A) GEOGRAPHIC LOCATIONS OF TEN U.S. CITIES

Outline I

Some Methods for Obtaining Proximities Data

'Psychological distance' (or closeness) of the stimulus objects.

- Subjective sorting or clustering.
- Stimulus confusability - same or different.
- Triadic Comparison.

Some Methods for Obtaining Proximities Data

'Psychological distance' (or closeness) of the stimulus objects.

- Subjective sorting or clustering.
- Stimulus confusability - same or different.
- Triadic Comparison.

Some Methods for Obtaining Proximities Data

'Psychological distance' (or closeness) of the stimulus objects.

- Subjective sorting or clustering.
- Stimulus confusability - same or different.
- Triadic Comparison.

Outline I

An Application of MDS to Morse Code Confusions Data

1. Stimulus : 36 auditory Morse code signals.
2. Sujects : 150 people who did not know Morse code.
3. Task : subjects listened to a pair of signals (produced at a fixed rapid rate by machine and separated by a quiet period of 1.4 seconds), and were required to state whether the two signals they heard were the same or different.

Data collected by Rothkopf (1957).

An Application of MDS to Morse Code Confusions Data

1. Stimulus : 36 auditory Morse code signals.
2. Sujects : 150 people who did not know Morse code.
3. Task : subjects listened to a pair of signals (produced at a fixed rapid rate by machine and separated by a quiet period of 1.4 seconds), and were required to state whether the two signals they heard were the same or different.

Data collected by Rothkopf (1957).

An Application of MDS to Morse Code Confusions Data

1. Stimulus : 36 auditory Morse code signals.
2. Sujects : 150 people who did not know Morse code.
3. Task : subjects listened to a pair of signals (produced at a fixed rapid rate by machine and separated by a quiet period of 1.4 seconds), and were required to state whether the two signals they heard were the same or different.

Data collected by Rothkopf (1957).

An Application of MDS to Morse Code Confusions Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	
A	92	4	6	13	3	14	10	13	46	5	22	3	25	34	6	6	9	35	23	6	37	13	17	12	7	3	2	7	5	5	8	6	5	6	2	3	A
B	5	84	37	31	5	28	17	21	5	19	34	40	6	10	12	22	25	16	18	2	18	34	8	84	30	42	12	17	14	40	32	74	43	17	4	4	B
C	4	38	87	17	4	29	13	7	11	19	24	35	14	3	9	51	34	24	14	6	6	11	14	32	82	38	13	15	31	14	10	30	28	24	18	12	C
D	8	62	17	88	7	23	40	36	9	13	81	56	8	7	9	27	9	45	29	6	17	20	27	40	15	33	3	9	6	11	9	19	8	10	5	6	D
E	6	13	14	6	97	2	4	4	17	1	5	6	4	4	5	1	5	10	7	67	3	3	2	5	6	5	4	3	5	3	5	2	4	2	3	3	E
F	4	51	33	19	2	90	10	29	5	33	16	50	7	6	10	42	12	35	14	2	21	27	25	19	27	13	8	16	47	25	26	24	21	5	5	5	F
G	9	18	27	38	1	14	90	6	5	22	33	16	14	13	62	52	23	21	5	3	15	14	32	21	23	39	15	14	5	10	4	10	17	23	20	11	G
H	3	45	23	25	9	32	8	87	10	10	9	29	5	8	8	14	8	17	37	4	36	59	9	33	14	11	3	9	15	43	70	35	17	4	3	3	H
I	64	7	7	13	10	8	6	12	93	3	5	16	13	30	7	3	5	19	35	16	10	5	8	2	5	7	2	5	8	9	6	8	5	2	4	5	I
J	7	9	38	9	2	24	18	5	4	85	22	31	8	3	21	63	47	11	2	7	9	9	9	22	32	28	67	66	33	15	7	11	28	29	26	23	J
K	5	24	38	73	1	17	25	11	5	27	91	33	10	12	31	14	31	22	2	2	23	17	33	63	16	18	5	9	17	8	8	18	14	13	5	6	K
L	2	69	43	45	10	24	12	26	9	30	27	86	6	2	9	37	36	28	12	5	16	19	20	31	25	59	12	13	17	15	26	29	36	16	7	3	L
M	24	12	5	14	7	17	29	8	8	11	23	8	96	62	11	10	15	20	7	9	13	4	21	9	18	8	5	7	6	6	5	7	11	7	10	4	M
N	31	4	13	30	8	12	10	16	13	3	16	8	59	93	5	9	5	28	12	10	16	4	12	4	16	11	5	2	3	4	4	6	2	2	10	2	N
O	7	7	20	6	5	9	76	7	2	39	26	10	4	8	86	37	35	10	3	4	11	14	25	35	27	27	19	17	7	7	6	18	14	11	20	12	O
P	5	22	33	12	5	36	22	12	3	78	14	46	5	6	21	83	43	23	9	4	12	19	19	19	41	30	34	44	24	11	15	17	24	23	25	13	P
Q	8	20	38	11	4	15	10	5	2	27	23	26	7	6	22	51	91	11	2	3	6	14	12	37	50	63	34	32	17	12	9	27	40	58	37	24	Q
R	13	14	16	23	5	34	26	15	7	12	21	33	14	12	12	29	8	87	16	2	23	23	62	14	12	13	7	10	13	4	7	12	7	9	1	2	R
S	17	24	5	30	11	26	5	59	16	3	13	10	5	17	6	6	3	18	96	9	56	24	12	10	6	7	8	2	2	15	28	9	5	5	5	2	S
T	13	10	1	5	46	3	6	6	14	6	14	7	6	5	6	11	4	4	7	96	8	5	4	2	2	6	5	5	3	3	3	8	7	6	14	6	T
U	14	29	12	32	4	32	11	34	21	7	44	32	11	13	6	20	12	40	51	6	93	57	34	17	9	11	6	6	16	34	10	9	9	7	4	3	U
V	5	17	24	16	9	29	6	39	5	11	26	43	4	1	9	17	10	17	11	6	32	92	17	57	35	10	10	14	28	79	44	36	25	10	1	5	V
W	9	21	30	22	9	36	25	15	4	25	29	18	15	6	26	20	25	61	12	4	19	20	86	22	25	22	10	22	19	16	5	9	11	6	3	7	W
X	7	64	45	19	3	28	11	6	1	35	50	42	10	8	24	32	61	10	12	3	12	17	21	91	48	26	12	20	24	27	16	57	29	16	17	6	X
Y	9	23	62	15	4	26	22	9	1	30	12	14	5	6	14	30	52	5	7	4	6	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	Y
Z	3	46	45	18	2	22	17	10	7	23	21	51	11	2	15	59	72	14	4	3	9	11	12	36	42	87	16	21	27	9	10	25	66	47	15	15	Z
1	2	5	10	3	3	5	13	4	2	29	5	14	9	7	14	30	28	9	4	2	3	12	14	17	19	22	84	63	13	8	10	8	19	32	57	55	1
2	7	14	22	5	4	20	13	3	25	26	9	14	2	3	17	37	28	6	5	3	6	10	11	17	30	13	62	89	54	20	5	14	20	21	16	11	2
3	3	8	21	5	4	32	6	12	2	23	6	13	5	2	5	37	19	9	7	6	4	16	6	22	25	12	18	64	86	31	23	41	16	17	8	10	3
4	6	19	19	12	8	25	14	16	7	21	13	19	3	3	2	17	29	11	9	3	17	55	8	37	24	3	5	26	44	89	42	44	32	10	3	3	4
5	8	45	15	14	2	45	4	67	7	14	4	41	2	0	4	13	7	9	27	2	14	45	7	45	10	10	14	10	30	69	90	42	24	10	6	5	5
6	7	80	30	17	4	23	4	14	2	11	11	27	6	2	7	16	30	11	14	3	12	30	9	58	38	39	15	14	26	24	17	88	69	14	5	14	6
7	6	33	22	14	5	25	6	4	6	24	13	32	7	6	7	36	39	12	6	2	3	13	9	30	30	50	22	29	18	15	12	61	85	70	20	13	7
8	3	23	40	6	3	15	15	6	2	33	10	14	3	6	14	12	45	2	6	4	6	7	5	24	35	50	42	29	16	16	9	30	60	89	61	26	8
9	3	14	23	3	1	6	14	5	2	30	6	7	16	11	10	31	32	5	6	7	6	3	8	11	21	24	57	39	9	12	4	11	42	56	91	78	9
0	9	3	11	2	5	7	14	4	5	30	8	3	2	3	25	21	29	2	3	4	5	3	2	12	15	20	50	26	9	11	5	22	17	52	81	94	0

Figure: Rothkopf's Data on Similarities Among Morse Code Symbols.

An Application of MDS to Morse Code Confusions Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	
A	92	4	6	13	3	14	10	13	46	5	22	3	25	34	6	6	9	35	23	6	37	13	17	12	7	3	2	7	5	5	8	6	5	6	2	3	A
B	5	84	37	31	5	28	17	21	5	19	34	40	6	10	12	22	25	16	18	2	18	34	8	84	30	42	12	17	14	40	32	74	43	17	4	4	B
C	4	38	87	17	4	29	13	7	11	19	24	35	14	3	9	51	34	24	14	6	6	11	14	32	82	38	13	15	31	14	10	30	28	24	18	12	C
D	8	62	17	88	7	23	40	36	9	13	81	56	8	7	9	27	9	45	29	6	17	20	27	40	15	33	3	9	6	11	9	19	8	10	5	6	D
E	6	13	14	6	97	2	4	4	17	1	5	6	4	4	5	1	5	10	7	67	3	3	2	5	6	5	4	3	5	3	5	2	4	2	3	3	E
F	4	51	33	19	2	90	10	29	5	33	16	50	7	6	10	42	12	35	14	2	21	27	25	19	27	13	8	16	47	25	26	24	21	5	5	5	F
G	9	18	27	38	1	14	90	6	5	22	33	16	14	13	62	52	23	21	5	3	15	14	32	21	23	39	15	14	5	10	4	10	17	23	20	11	G
H	3	45	23	25	9	32	8	87	10	10	9	29	5	8	8	14	8	17	37	4	36	59	9	33	14	11	3	9	15	43	70	35	17	4	3	3	H
I	64	7	7	13	10	8	6	12	93	3	5	16	13	30	7	3	5	19	35	16	10	5	8	2	5	7	2	5	8	9	6	8	5	2	4	5	I
J	7	9	38	9	2	24	18	5	4	85	22	31	8	3	21	63	47	11	2	7	9	9	9	22	32	28	67	66	33	15	7	11	28	29	26	23	J
K	5	24	38	73	1	17	25	11	5	27	91	33	10	12	31	14	31	22	2	2	23	17	33	63	16	18	5	9	17	8	8	18	14	13	5	6	K
L	2	69	43	45	10	24	12	26	9	30	27	86	6	2	9	37	36	28	12	5	16	19	20	31	25	59	12	13	17	15	26	29	36	16	7	3	L
M	24	12	5	14	7	17	29	8	8	11	23	8	96	62	11	10	15	20	7	9	13	4	21	9	18	8	5	7	6	6	5	7	11	7	10	4	M
N	31	4	13	30	8	12	10	16	13	3	16	8	59	93	5	9	5	28	12	10	16	4	12	4	16	11	5	2	3	4	4	6	2	2	10	2	N
O	7	7	20	6	5	9	76	7	2	39	26	10	4	8	86	37	35	10	3	4	11	14	25	35	27	27	19	17	7	7	6	18	14	11	20	12	O
P	5	22	33	12	5	36	22	12	3	78	14	46	5	6	21	83	43	23	9	4	12	19	19	19	41	30	34	44	24	11	15	17	24	23	25	13	P
Q	8	20	38	11	4	15	10	5	2	27	23	26	7	6	22	51	91	11	2	3	6	14	12	37	50	63	34	32	17	12	9	27	40	58	37	24	Q
R	13	14	16	23	5	34	26	15	7	12	21	33	14	12	12	29	8	87	16	2	23	23	62	14	12	13	7	10	13	4	7	12	7	9	1	2	R
S	17	24	5	30	11	26	5	59	16	3	13	10	5	17	6	6	3	18	96	9	56	24	12	10	6	7	8	2	2	15	28	9	5	5	5	2	S
T	13	10	1	5	46	3	6	6	14	6	14	7	6	5	6	11	4	4	7	96	8	5	4	2	2	6	5	5	3	3	3	8	7	6	14	6	T
U	14	29	12	32	4	32	11	34	21	7	44	32	11	13	6	20	12	40	51	6	93	57	34	17	9	11	6	6	16	34	10	9	9	7	4	3	U
V	5	17	24	16	9	29	6	39	5	11	26	43	4	1	9	17	10	17	11	6	32	92	17	57	35	10	10	14	28	79	44	36	25	10	1	5	V
W	9	21	30	22	9	36	25	15	4	25	29	18	15	6	26	20	25	61	12	4	19	20	86	22	25	22	10	22	19	16	5	9	11	6	3	7	W
X	7	64	45	19	3	28	11	6	1	35	50	42	10	8	24	32	61	10	12	3	12	17	21	91	48	26	12	20	24	27	16	57	29	16	17	6	X
Y	9	23	62	15	4	26	22	9	1	30	12	14	5	6	14	30	52	5	7	4	6	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	Y
Z	3	46	45	18	2	22	17	10	7	23	21	51	11	2	15	59	72	14	4	3	9	11	12	36	42	87	16	21	27	9	10	25	66	47	15	15	Z
1	2	5	10	3	3	5	13	4	2	29	5	14	9	7	14	30	28	9	4	2	3	12	14	17	19	22	84	63	13	8	10	8	19	32	57	55	1
2	7	14	22	5	4	20	13	3	25	26	9	14	2	3	17	37	28	6	5	3	6	10	11	17	30	13	62	89	54	20	5	14	20	21	16	11	2

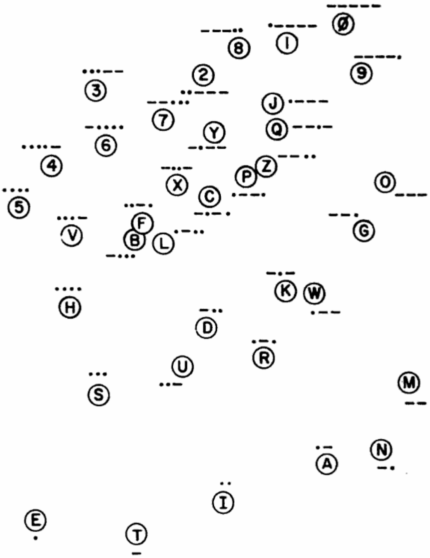
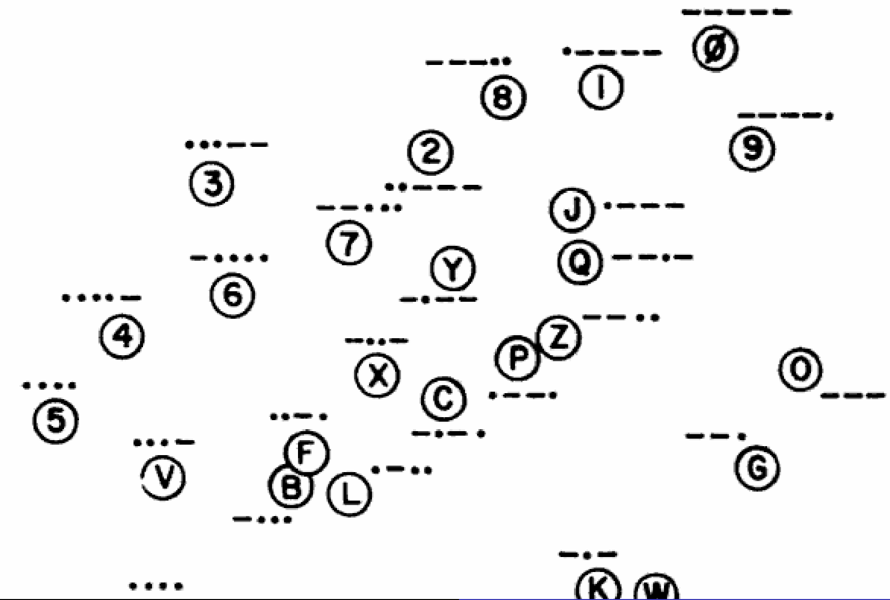
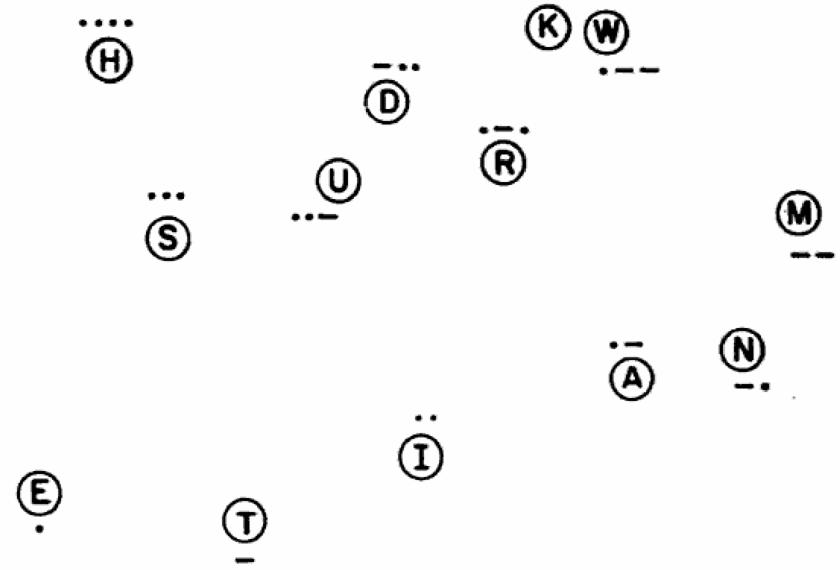


Figure: Configuration Resulting from Morse Code Similarities


Ø
8
1
9
3
2
J
7
Q
6
Y
4
Z
P
X
O
5
C
G
V
F
B
L
K
W


H
D
K
W
R
U
S
M
N
A
I
E
T

Interpretation of Results

Shepard's interpretation

- length of the signal
- number of components
- number of dots and dashes

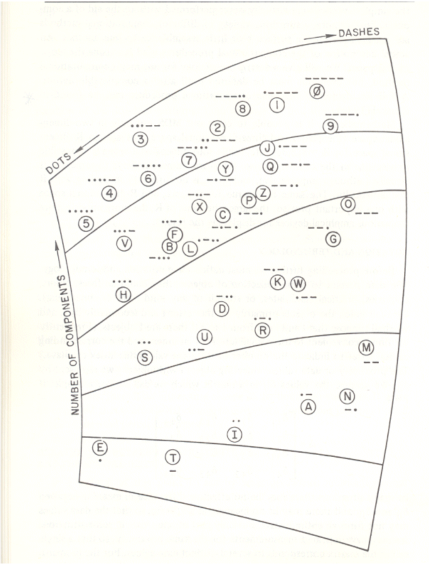
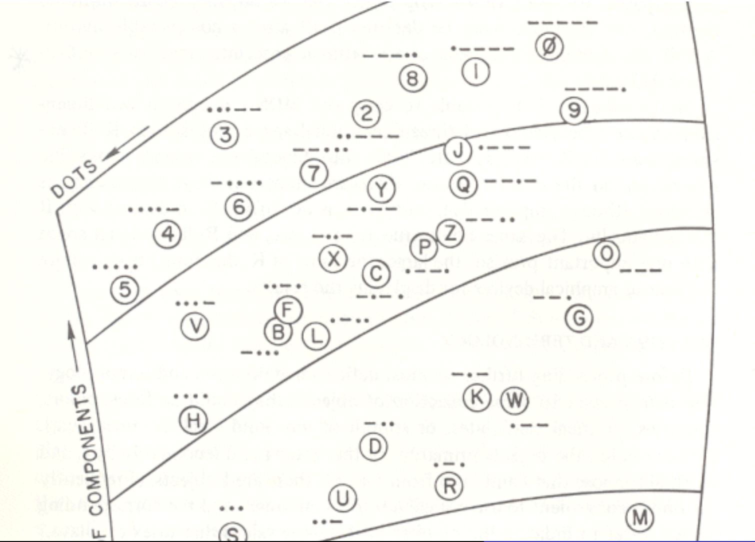
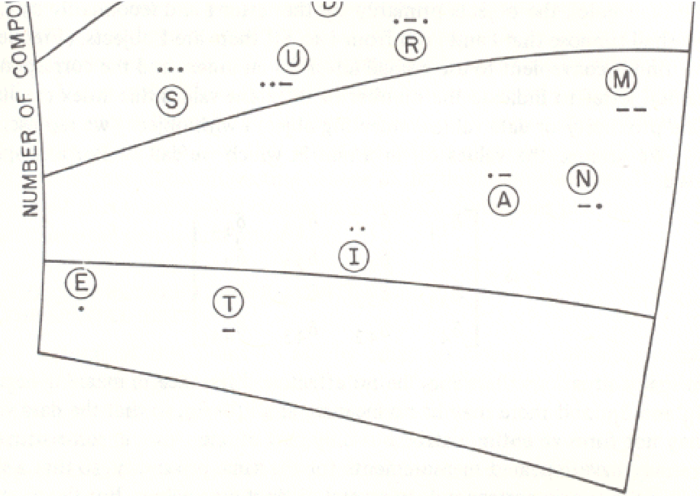


Figure:

Shepard's interpretation



Shepard's interpretation



Outline I

Multidimensional Scaling

MDS's basics

- A Multidimensional Scaling may be carried out in R-dimensional space for R=1,2,3,4,etc.
- Metric and non-metric MDS.
- Dissimilarities measures between objects or stimulus.

Multidimensional Scaling

MDS's basics

- A Multidimensional Scaling may be carried out in R-dimensional space for R=1,2,3,4,etc.
- Metric and non-metric MDS.
- Dissimilarities measures between objects or stimulus.

Multidimensional Scaling

MDS's basics

- A Multidimensional Scaling may be carried out in R-dimensional space for R=1,2,3,4,etc.
- Metric and non-metric MDS.
- Dissimilarities measures between objects or stimulus.

Dissimilarity Matrix

The input for a MDS method is a dissimilarity (or similarity) matrix Δ, as below:

$$\boldsymbol{\Delta} = \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,N} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N,1} & \delta_{N,2} & \cdots & \delta_{N,N} \end{pmatrix} \tag{1}$$

The dissimilarity (distance) or data value connecting object i with object j we represent by $\delta_{i,j}$.[1]

[1] In many situations there may be no effective difference in meaning between $\delta_{i,j}$ and $\delta_{j,i}$, and there may be no meaning at all for $\delta_{i,i}$, so that the data values may not form an entire matrix, but only part of one.

Objects Vectors

The output of a MDS method is a set of N R-dimensional vectors representing the objects (or stimulus) subjected to the current study:

$$\begin{aligned}
\mathbf{x_1} &= \left(x_{1,1}, \ldots, x_{1,R}\right)^T \\
\mathbf{x_2} &= \left(x_{2,1}, \ldots, x_{2,R}\right)^T \\
&\vdots \\
\mathbf{x_N} &= \left(x_{N,1}, \ldots, x_{N,R}\right)^T
\end{aligned} \tag{2}$$

Outline I

Distance Between Objects

The distance between two objects $\mathbf{x_i}$ and $\mathbf{x_j}$ $(d_{i,j})$ may be given according to different metrics.

- Euclidean distance:

$$\begin{aligned} d_{i,j} &= \|\mathbf{x_i} - \mathbf{x_j}\| \\ &= \sqrt{\langle \mathbf{x_i} - \mathbf{x_j}, \mathbf{x_i} - \mathbf{x_j} \rangle} \\ &= \sqrt{\sum_{r=1}^{R} (x_{i,r} - x_{j,r})^2} \end{aligned} \tag{3}$$

Distance Between Objects

- Minkowski distance:

$$d_{i,j} = \left(\sum_{r=1}^{R} (x_{i,r} - x_{j,r})^k \right)^{\frac{1}{k}} \tag{4}$$

which is a generalization of the Euclidean distance.

Distance Between Objects

- Weighted Euclidean distance:

$$d_{i,j} = \sqrt{\sum_{r=1}^{R} w_r \left(x_{i,r} - x_{j,r}\right)^2} \tag{5}$$

where the weights w_i are previously defined according to some choice.

Distance Between Objects

- Mahalanobis distance:

$$\begin{aligned} d_{i,j} &= \sqrt{(\mathbf{x_i} - \mathbf{x_j})^T \mathbf{S}^{-1} (\mathbf{x_i} - \mathbf{x_j})} \\ &= \sqrt{\sum_{r=1}^{R} \frac{(x_{i,r} - x_{j,r})^2}{\sigma_r^2}} \end{aligned} \tag{6}$$

where $\mathbf{S}$ is the covariance matrix for $\mathbf{x_i}$ and $\mathbf{x_j}$.

The Choice of a Metric

The choice of different metrics can leads the problem to completely disparate solutions. Lets see it through the example given by Jäkel et al. (2007).

Classification Task

- Prototype based classification.
- Use of an inner product generalization to derive a distance metric.

The Choice of a Metric

The choice of different metrics can leads the problem to completely disparate solutions. Lets see it through the example given by Jäkel et al. (2007).

Classification Task

- Prototype based classification.
- Use of an inner product generalization to derive a distance metric.

Example: Classification Task - Prototype

Suppose a classification problem where we have 2 classes A and B Hence, to separate A from B the arithmetic means of all examples in A and B are calculated:

$$\bar{a} = \frac{1}{|A|} \sum_{a \in A} a \quad \text{and} \quad \bar{b} = \frac{1}{|B|} \sum_{b \in B} b \tag{7}$$

A new pattern x is classified as belonging to class A if it is closer to $\bar{a}$ (the mean of A) than to $\bar{b}$ (the mean of B). We

Example: Classification Task - Prototype Classifier

A Euclidean prototype classifier decides that a new stimulus x belongs to class A over class B whenever

$$
\begin{aligned}
\langle x-\bar{b}, x-\bar{b}\rangle &> \langle x-\bar{a}, x-\bar{a}\rangle \\
\langle x, x\rangle - 2\langle \bar{b}, x\rangle + \langle \bar{b}, \bar{b}\rangle &> \langle x, x\rangle - 2\langle \bar{a}, x\rangle + \langle \bar{a}, \bar{a}\rangle \\
2\langle \bar{a}, x\rangle - 2\langle \bar{b}, x\rangle &> \langle \bar{a}, \bar{a}\rangle - \langle \bar{b}, \bar{b}\rangle \\
\langle \bar{a}-\bar{b}, x\rangle &> \frac{1}{2}\left(\langle \bar{a}, \bar{a}\rangle - \langle \bar{b}, \bar{b}\rangle\right) \\
\langle w, x\rangle &> \theta \qquad (8)
\end{aligned}
$$

It can be seen that the prototype classifier defines a hyperplane in the input space just like a perceptron does.[2]

[2] The weight vector w is the difference between the means, $\bar{a}-\bar{b}$ and the threshold θ is given in the right-hand side $\frac{1}{2}\left(\langle \bar{a}, \bar{a}\rangle - \langle \bar{b}, \bar{b}\rangle\right)$.

Example: Prototype Classifier using Standard Inner Product

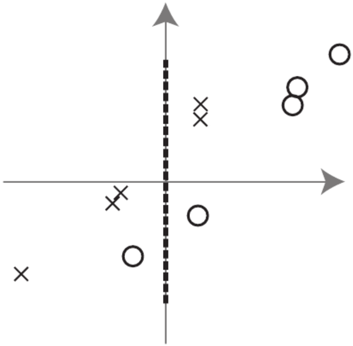


Figure: The thick dashed line depicts the decision bound for a prototype classifier when the standard inner product is used.

Inner Product Generalization

A very important generalization of inner product is given by

$$\begin{aligned} \langle w, v \rangle_K &= \sum_{i=1}^{n} \sum_{i=1}^{n} w_i v_j k_{i,j} \\ &= w^T K v \end{aligned} \quad (9)$$

when K is the identity matrix, the standard inner product is recovered.

Inner Product Generalization

In order for this definition (9) to result in an inner product the three axioms have to be fulfilled:

1. symmetry : $\langle w, v\rangle = \langle v, w\rangle$,
2. linearity : $\langle a(u+w), v\rangle = a\langle u, v\rangle + a\langle w, v\rangle$,
3. positive definiteness : $\langle w, w\rangle \geq 0$ for all w (positive) and $\langle w, w\rangle = 0$ if and only if $w = 0$ (definiteness).

The matrix K must be positive define so that (9) defines a inner product.

Inner Product Generalization

In order for this definition (9) to result in an inner product the three axioms have to be fulfilled:

1. symmetry : $\langle w, v\rangle = \langle v, w\rangle$,
2. linearity : $\langle a(u + w), v\rangle = a\langle u, v\rangle + a\langle w, v\rangle$,
3. positive definiteness : $\langle w, w\rangle \geq 0$ for all w (positive) and $\langle w, w\rangle = 0$ if and only if $w = 0$ (definiteness).

The matrix K must be positive define so that (9) defines a inner product.

Inner Product Generalization

In order for this definition (9) to result in an inner product the three axioms have to be fulfilled:

1. symmetry : $\langle w, v \rangle = \langle v, w \rangle$,
2. linearity : $\langle a(u + w), v \rangle = a\langle u, v \rangle + a\langle w, v \rangle$,
3. positive definiteness : $\langle w, w \rangle \geq 0$ for all w (positive) and $\langle w, w \rangle = 0$ if and only if $w = 0$ (definiteness).

The matrix K must be positive define so that (9) defines a inner product.

Inner Product Generalization

In order for this definition (9) to result in an inner product the three axioms have to be fulfilled:

1. symmetry : $\langle w, v\rangle = \langle v, w\rangle$,
2. linearity : $\langle a(u + w), v\rangle = a\langle u, v\rangle + a\langle w, v\rangle$,
3. positive definiteness : $\langle w, w\rangle \geq 0$ for all w (positive) and $\langle w, w\rangle = 0$ if and only if $w = 0$ (definiteness).

The matrix K must be positive define so that (9) defines a inner product.

Inner Product Generalization

Since K is a square $(n \times n)$ symmetric matrix it has n linearly independent eigenvectors. As it is also positive semidefinite, the eigendecomposition applies. It may be then factorized as

$$K = U \Lambda U^T \tag{10}$$

where U is the square matrix $(n \times n)$ which columns are eigenvectors of K and Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues.

Inner Product Generalization

Using (10) and (9) we have

$$\begin{aligned}
\langle w, v \rangle_K &= w^T K v \\
&= w^T (U \Lambda U^T) v \\
&= w^T (\sqrt{\Lambda} U)^T (\sqrt{\Lambda} U) v \\
&= (\sqrt{\Lambda} U w)^T (\sqrt{\Lambda} U v) \\
&= \langle \sqrt{\Lambda} U w, \sqrt{\Lambda} U v \rangle
\end{aligned} \tag{11}$$

U peforms a rotation on w and v; and $\sqrt{\Lambda}$ makes a rescaling. It is a transformation of w and v into a new coordinate system, in which the inner products $\langle w, v \rangle_K$ amounts to the standard inner product.

Prototype Classifier using a Different Inner Product

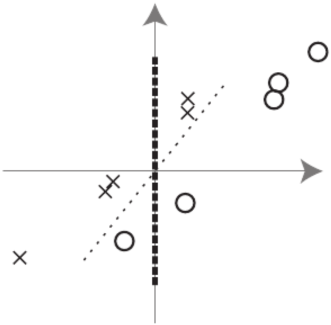


Figure: Comparison of Prototype classifier with Euclidean inner product and classifier with a different inner product.

Prototype Classifier using a Different Inner Product

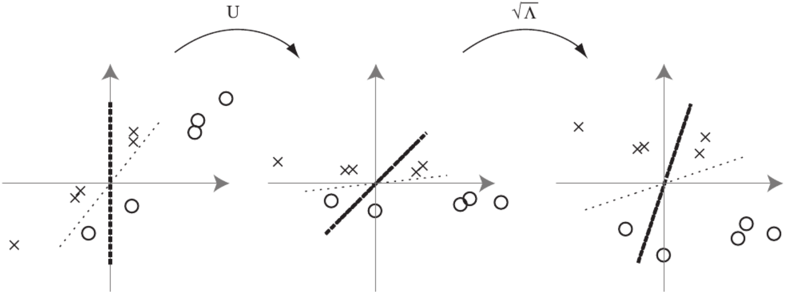


Figure: Whether a prototype classifier can separate two classes depends also on the inner product that is chosen. The left panel shows two classes (circles and crosses) with highly correlated dimensions–in this case the standard inner product is not appropriate. The thick dashed line depicts the decision bound for a prototype classifier when the standard inner product is used. The dotted line depicts a decision bound with a different inner product. This inner product corresponds to the standard inner product in the space depicted on the right that can be obtained by rotating and rescaling the original space.

Distance and Dissimilarity

The central motivating concept of multidimensional scaling is that the distances $d_{i,j}$ between the points should correspond to the proximities $\delta_{i,j}$.

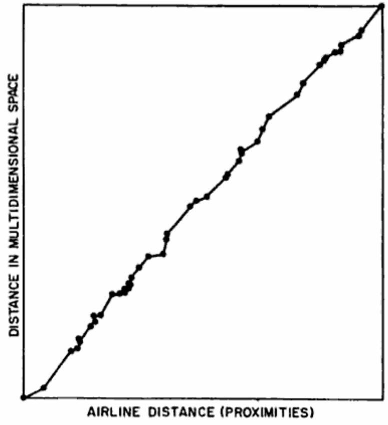


Figure: Scatter diagram (or Shepard diagram)

Scatter Diagram of Morse Code

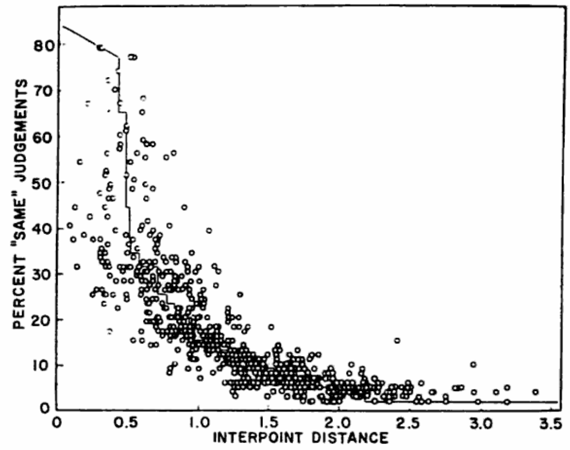


Figure: Scatter Diagram Associated with Morse Code Configuration Problem

Distance - Similarity Relationship

The relationship between δ and d is usually described through a function

$$d = f(\delta) \tag{12}$$

We can adjust the numerical calculation procedure to aim for any type of relationship we wish, and each type of relationship corresponds to a different type of MDS.

Metric and Nonmetric MDS

- *metric* MDS : if we aim a relationship specified by a function $f(\delta)$ (like $d = a + b\delta$).
- *nonmetric* MDS : If we want a relationship which is not described by any formula, but is only a rising pattern (any increasing curve is as good as any other), i.e. f is arbitrary, but must obey the monotonicity constraint

$$\delta_{i,j} \leq \delta_{i',j'} \Rightarrow f(\delta_{i,j}) \leq f(\delta_{i',j'}) \ , \tag{13}$$

Metric and Nonmetric MDS

- *metric* MDS : if we aim a relationship specified by a function $f(\delta)$ (like $d = a + b\delta$).
- *nonmetric* MDS : If we want a relationship which is not described by any formula, but is only a rising pattern (any increasing curve is as good as any other), i.e. f is arbitrary, but must obey the monotonicity constraint

$$\delta_{i,j} \leq \delta_{i',j'} \Rightarrow f(\delta_{i,j}) \leq f(\delta_{i',j'}) \ , \tag{13}$$

Goodness-of-Fit Function

We define a objective function (error function or goodness-of-fit function) as

$$\text{f-stress} = \sqrt{\frac{\sum_i \sum_j \left(f(\delta_{i,j}) - d_{i,j}\right)^2}{\text{scale factor}}} \tag{14}$$

where the scale factor most commonly used is

$$\text{scale factor} = \sum_i \sum_j d_{i,j}^2 \tag{15}$$

Outline I

Classical Scaling

Classical scaling originated in the 1930s when Young and Householder (1938) showed how starting with a matrix of distances between points in a Euclidean space, coordinates for the points can be found such that distances are preserved. Torgerson (1952) brought the subject to popularity using the technique for scaling.

How to Calculate

1. Take the dissimilarity matrix D (where, in this case, $D \approx \Delta$) and build the matrix A in the following way: $[A]_{i,j} = a_{i,j} = -\frac{1}{2} d_{i,j}^2$;
2. Calculate the matrix $B = HAH$, where A is given above and H is the following matrix $H = I - n^{-1}\mathbf{1}\mathbf{1}^T$;
3. B is usually a positive-semidefinite matrix, so the Singular value decomposition (SVD) applies in the following form: $B = V \Lambda V^T$;
4. B is a matrix of rank p, then $n - p$ eigenvalues of B are null and we have: $B = V_1 \Lambda_1 V_1^T$, where $\Lambda_1 = diag(\lambda_1, \ldots, \lambda_p)$ and $V_1 = [v_1, \ldots, v_p]$;
5. As $B = XX^T$, X is given by $X = V_1 \sqrt{\Lambda_1}$.

Distance

The distance between two objects

$$\begin{aligned} d_{i,j}^2 &= (x_i - x_j)^T (x_i - x_j) \\ &= x_i^T x_i + x_j^T x_j - 2 x_i^T x_j \end{aligned} \tag{16}$$

To overcome the indeterminacy of the solution due to arbitrary translation, the centroid of the configuration of points is placed at the origin.

$$\sum_{i=1}^{N} x_{i,r} = 0 \quad (r = 1, \ldots, R) \tag{17}$$

Step 1a

From (16) and (17)

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} d_{i,j}^2 &= \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i + \frac{1}{N}\sum_{i=1}^{N} x_j^T x_j - 2\frac{1}{N}\sum_{i=1}^{N} x_i^T x_j \\ &= \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i + x_j^T x_j \end{aligned} \tag{18}$$

since $\frac{1}{N}\sum_{i=1}^{N} x_j^T x_j = x_j^T x_j$ and $\frac{1}{N}\sum_{i=1}^{N} x_i^T x_j = 0$.

Step 1b

Analogously

$$\frac{1}{N}\sum_{j=1}^{N} d_{i,j}^2 = x_i^T x_i + \sum_{j=1}^{N} x_j^T x_j \tag{19}$$

Step 2

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}d_{i,j}^2 &= \frac{1}{N}\sum_{i=1}^{N}x_i^T x_i + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}x_j^T x_j \\
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}d_{i,j}^2 &= \frac{1}{N}\sum_{i=1}^{N}x_i^T x_i + \sum_{j=1}^{N}x_j^T x_j \qquad (20)
\end{aligned}$$

Step 3

From (16) and using the results from (18), (19) and (20) we have

$$
\begin{aligned}
x_i^T x_j &= \frac{1}{2}\left[x_i^T x_i + x_j^T x_j - d_{i,j}^2\right] \\
&= \frac{1}{2}\left[\left(\frac{1}{N}\sum_{j=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{j=1}^{N} x_j^T x_j\right) + \right. \\
&\qquad \left. \left(\frac{1}{N}\sum_{i=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i\right) - d_{i,j}^2\right] \qquad (21)
\end{aligned}
$$

Step 3

$$
\begin{aligned}
x_i^T x_j &= -\frac{1}{2}\Bigg[d_{i,j}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{j=1}^{N} d_{i,j}^2 + \\
&\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^T x_i\right) + \\
&\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{j=1}^{N} x_j^T x_j\right)\Bigg] \qquad (22) \\
&= -\frac{1}{2}\left[d_{i,j}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{i,j}^2 - \frac{1}{N}\sum_{j=1}^{N} d_{i,j}^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j}^2\right]
\end{aligned}
$$

Matrix B

The elements of Matrix A are defined as

$$[A]_{i,j} = a_{i,j} = -\frac{1}{2} d_{i,j}^2 \tag{23}$$

Matrix H is given by

$$H = I - N^{-1}\mathbf{1}\mathbf{1}^T \tag{24}$$

Matrix B

Calculating $C = AH$, the elements of C are given by

$$c_{ij} = \frac{1}{2N} \sum_{j' \neq j} d^2_{ij'} - \frac{N-1}{2N} d^2_{ij} \tag{25}$$

Calculating now $B = HC$, the elements of B are given by

$$b_{ij} = -\frac{1}{N} \sum_{i' \neq i} c_{i'j} + \frac{N-1}{N} c_{ij} \tag{26}$$

Matrix B

Using the result of (25) in (26) we have

$$b_{ij} = -\frac{1}{2N^2}\left[(N-1)^2 d_{ij}^2 - (N-1)\sum_{i' \neq i} d_{i'j}^2 - (N-1)\sum_{j' \neq j} d_{ij'}^2 + \sum_{i' \neq i}\sum_{j' \neq j} d_{i'j'}^2\right] \tag{27}$$

Comparing (27) and (23), we conclude that

$$B = x^T x \tag{28}$$

and we may apply the SVD procedure, as described before (slide 42).

Example

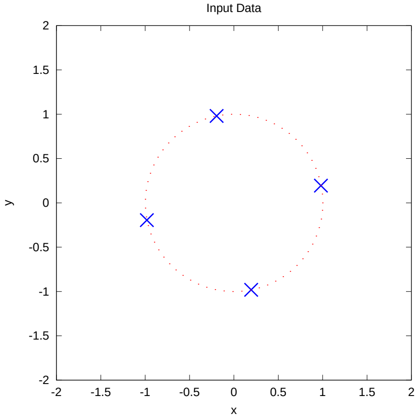


Figure: Input Data.

Example

The data matrix from the point shown in the previous plot is:

$$X = \begin{pmatrix} 0.98 & -0.20 & -0.98 & 0.20 \\ 0.20 & 0.98 & -0.20 & -0.98 \end{pmatrix} \tag{29}$$

Example

Based on the previous input data, the distance matrix D is calculated using euclidean distance, and its result is shown below:

$$D = \begin{pmatrix} 0.00 & 1.41 & 2.00 & 1.41 \\ 1.41 & 0.00 & 1.41 & 2.00 \\ 2.00 & 1.41 & 0.00 & 1.41 \\ 1.41 & 2.00 & 1.41 & 0.00 \end{pmatrix} \tag{30}$$

Example

From D we calculate the matrix A, where $a_{i,j} = -\frac{1}{2} d_{i,j}^2$, then

$$A = \begin{pmatrix} -0.00 & -1.00 & -2.00 & -1.00 \\ -1.00 & -0.00 & -1.00 & -2.00 \\ -2.00 & -1.00 & -0.00 & -1.00 \\ -1.00 & -2.00 & -1.00 & -0.00 \end{pmatrix} \tag{31}$$

Example

Matrix H is given by

$$H = \begin{pmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{pmatrix} \tag{32}$$

Example

Matrix $H = I - n^{-1}\mathbf{1}\mathbf{1}^T$ is given by

$$H = \begin{pmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{pmatrix} \tag{33}$$

Example

$$B = HAH$$

$$B = \begin{pmatrix} 1.00 & 0.00 & -1.00 & 0.00 \\ 0.00 & 1.00 & -0.00 & -1.00 \\ -1.00 & -0.00 & 1.00 & -0.00 \\ 0.00 & -1.00 & -0.00 & 1.00 \end{pmatrix} \tag{34}$$

Example

Performing the SVD, we have $B = USV^T$

$$U = \begin{pmatrix} -0.71 & 0.00 & -0.00 & 0.71 \\ -0.00 & -0.71 & 0.71 & 0.00 \\ 0.71 & -0.00 & 0.00 & 0.71 \\ -0.00 & 0.71 & 0.71 & 0.00 \end{pmatrix} \tag{35}$$

$$S = \begin{pmatrix} 2.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 2.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{pmatrix} \tag{36}$$

$$V = \begin{pmatrix} -0.71 & -0.00 & 0.00 & -0.71 \\ -0.00 & -0.71 & 0.71 & 0.00 \\ 0.71 & -0.00 & 0.00 & -0.71 \\ -0.00 & 0.71 & 0.71 & 0.00 \end{pmatrix} \tag{37}$$

Example

The rank of B is 2, that means it has 2 not null eigenvalues. We may then consider only the elements $(1,1)$, $(1,2)$, $(2,1)$ and $(2,2)$ in matrices S, V and U.

$$U_1 = \begin{pmatrix} -0.71 & 0.00 \\ -0.00 & -0.71 \\ 0.71 & -0.00 \\ -0.00 & 0.71 \end{pmatrix} \tag{38}$$

$$S_1 = \begin{pmatrix} 2.00 & 0.00 \\ 0.00 & 2.00 \end{pmatrix} \tag{39}$$

$$U_1 = \begin{pmatrix} -0.71 & -0.00 & 0.00 & -0.71 \\ -0.00 & -0.71 & 0.71 & 0.00 \end{pmatrix} \tag{40}$$

Example

$$X' = U_1 \sqrt{S_1}$$

$$X' = \begin{pmatrix} -1.00 & 0.00 \\ -0.00 & -1.00 \\ 1.00 & -0.00 \\ -0.00 & 1.00 \end{pmatrix} \tag{41}$$

Example

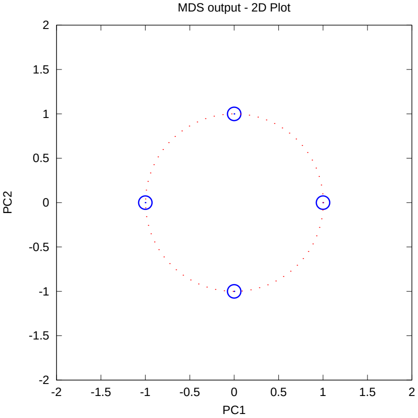


Figure: Output Result from MDS in 2 dimensions.

Example

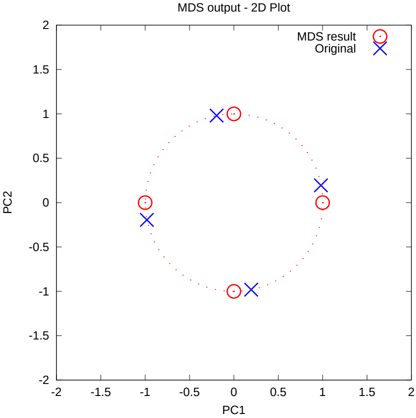


Figure: Output Result from MDS in 2 dimensions.

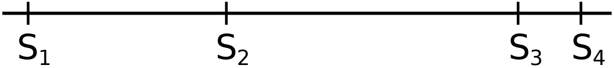
Outline I

Psychophysics

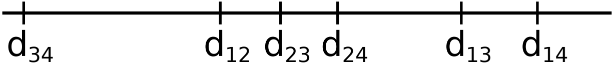
- dimensions
- number of relevant dimensions
- intuitively given dimensions

One-dimensional Scale

Given four stimulus-objects designated S_1, S_2, S_3 and S_4.

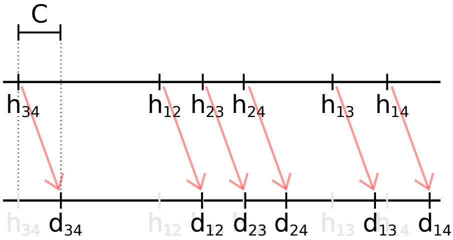


Given the same four stimulus-objects, the scale of comparative distances locates, with respect to one another on a distance continuum, the six inter-stimulus distances, d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}.



One-dimensional Scale

A comparative distance is not a “distance” in the usual sense of the term, but is a distance minus an unknown constant.



Thus, a comparative distance h_{jk} plus an unknown additive constant C gives the corresponding absolute distance d_{jk}.

$$h_{jk} + C = d_{jk} \tag{42}$$

Estimating the Additive Constant

It is assumed that the value of the additive constant is the value which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality.

Example: consider the following comparative interpoint distances

$$h_{12} = 2, \quad h_{14} = 2, \quad h_{23} = 2, \quad h_{25} = 1, \quad h_{35} = 0,$$
$$h_{13} = 3, \quad h_{15} = 0, \quad h_{24} = 5, \quad h_{34} = 2, \quad h_{45} = 1.$$

Estimating the Additive Constant

The value of C which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality is $C = 3$.

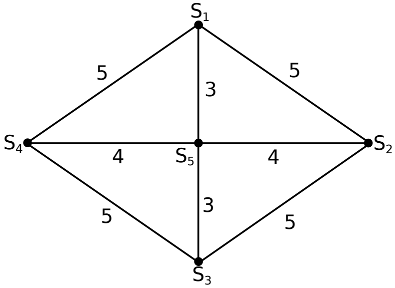
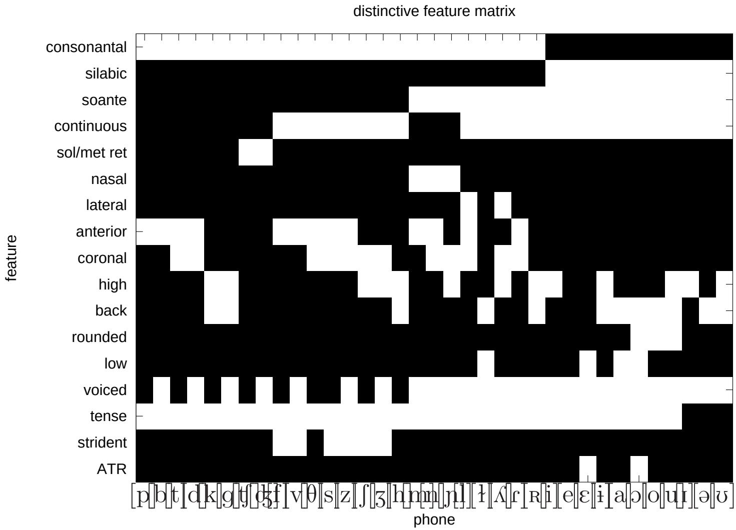
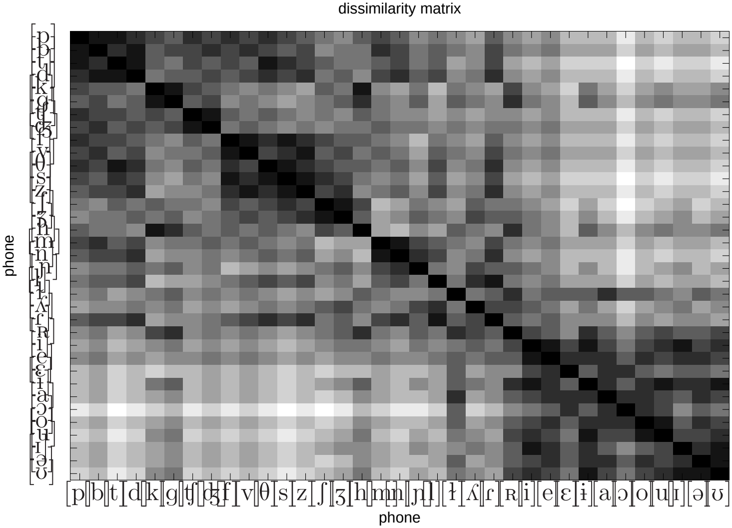


Figure: Representation of the Stimulus in the smallest possible dimensionality, with $C = 3$

$$\begin{aligned} d_{12} &= 5, & d_{13} &= 6, \\ d_{14} &= 5, & d_{15} &= 3, \\ d_{23} &= 5, & d_{24} &= 8, \\ d_{25} &= 4, & d_{34} &= 5, \\ d_{35} &= 3, & d_{45} &= 4. \end{aligned}$$

Outline I


distinctive feature matrix
feature
consonantal
silabic
soante
continuous
sol/met ret
nasal
lateral
anterior
coronal
high
back
rounded
low
voiced
tense
strident
ATR
[p][b][t][d][k][g][tʃ][dʒ][f][v][θ][s][z][ʃ][ʒ][h][m][n][ɲ][l][ɫ][ʎ][ɾ][ʀ][i][e][ɛ][ɨ][a][ɔ][o][u][ɪ][ə][ʊ]
phone


dissimilarity matrix
phone
phone
[p][b][t][d][k][g][tʃ][dʒ][f][v][θ][s][z][ʃ][ʒ][h][m][n][ɲ][l][ɫ][ʎ][ɾ][ʀ][i][e][ɛ][ɨ][a][ɔ][o][u][ɪ][ə][ʊ]

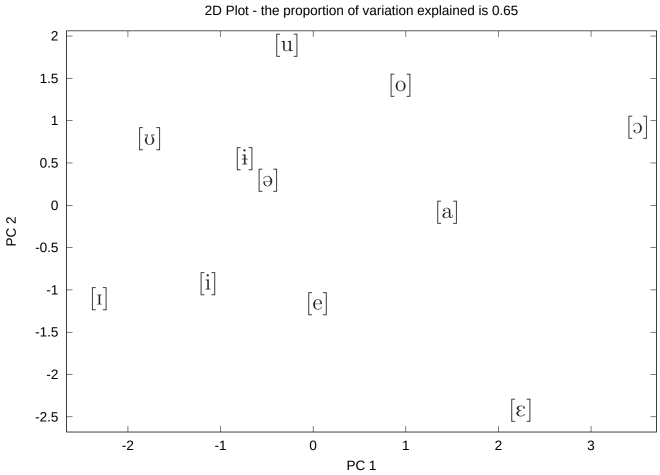


Figure: MDS - Vowels Portuguese

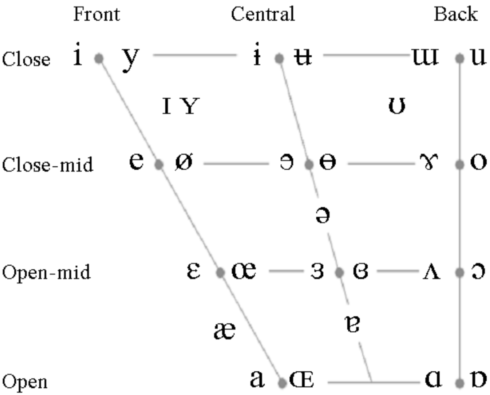


Figure: Vowel Trapezoid

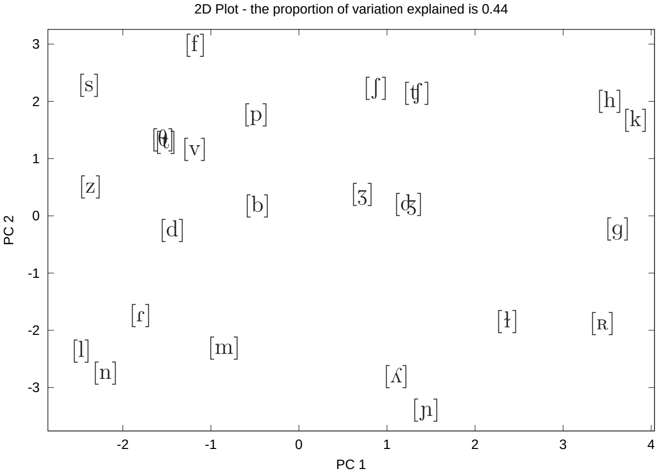


Figure: MDS - Consonants Portuguese

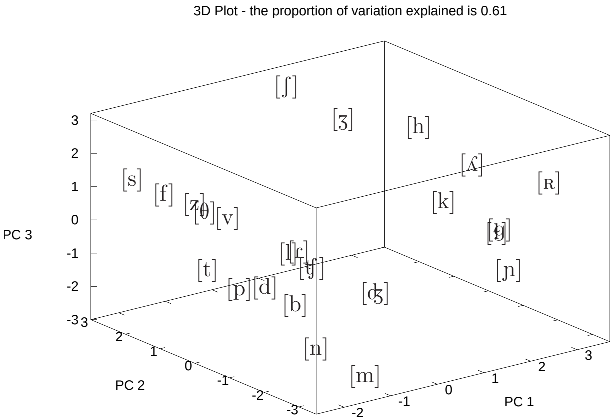


Figure: MDS - Consonants Portuguese

Outline I

Jäkel, F., Schölkopf, B., and Wichmann, F. A. (2007). A tutorial on kernel methods for categorization. *Journal of Mathematical Psychology*, 51:343–358.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Sage.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 53:94–101.

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 5:33–48.

Torgerson, W. (1952). Multidimensional scaling: 1. theory and method. *Psychometrika*, 17:401–419.

Young, G. and Householder, A. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22.