HUDM5124: Introduction to Multidimensional Scaling, Clustering, and Related Methods

Session 8: Analyzing (and Computing) Similarity (MMDS Ch. 6)

I. Models of Psychological Similarity

- Geometric model of similarity
- Similarity as feature-matching
- Similarity as structural-alignment
- Similarity as transformation

II. Computing a similarity measure from multivariate data

→ Getting issue I right helps us find the right model for analyzing similarity data

→ Getting issue II right helps us compute similarity from multivariate data so that it is maximally meaningful

Readings

(I.) Similarity as feature-matching:

**Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 325-340.

Similarity as structural- alignment:

*Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*, 517-535.

*Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431-467.

Similarity as transformation:

*Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87*, 1-32.

(II.) Gathering / computing proximities **Modern MDS Chapter 6

** = required

More on Similarity

I. On the nature of similarity (similarity as a psychological construct)

- Is it a useful theoretical construct?
- Does a given (dis)similarity measure obey the metric axioms?
- Are spatial models appropriate to model psychological similarity? If not, then what?

II. Computing similarity measures from multivariate data

- What measures are possible / appropriate?
- How do they differ?

Is similarity useful as an explanatory psychological construct?

Pro:

- Similarity explains generalization (Pavlov; Shepard)
- Explains mental associations
- Relevant in choice and decision making
- Relevant in problem solving
- Etc. etc.

Con:

- Context effects (Torgerson etc.)
- Nelson Goodman (philosopher) *“Similarity.. is a fraud”*
- Medin et al. “Respects for similarity”

The Nature of Psychological Similarity

(Given a set of psychological data interpretable as proximities:)

Question: Do the data satisfy the metric axioms?

If no, should we transform the data, or generalize our models?

If the latter, which axioms should we relax or drop?

- Symmetry: proximity can be asymmetric

  simple spatial models have trouble accounting for this

- Minimality: Is there some meaningful sense of "self-similarity"?

  Yes, particularly for some types of confusion, identification, or choice data (Shepard, 1958)

Similarity and spatial / geometric models

Some psychological dimensions can be attended to separately, some cannot: separable vs. integral (ref. Garner; Attneave)

A geometric distance model (Euclidean or other Minkowski / power) satisfies the metric axioms plus the property of *segmental additivity* (Beals, Krantz & Tversky,1968); these imply the more general *additive difference metric*, with properties:

- Intradimensional subtractivity
- Interdimensional additivity

→ such properties are testable as psychological hypotheses

But what if we have <u>discrete</u> attributes, rather than continuous? (Tversky, 1977) e.g. colas data → other metric spaces

Reconsideration of similarity

Similarity as feature-matching:

- Tversky's (1977) "contrast model" of similarity (a.k.a. the "feature-matching" model):

$$S(a,b) = \theta\, f(A \cap B) - \alpha\, g(A-B) - \beta\, g(B-A)$$

"common features" "distinctive features" (of A) (of B)

- "focusing hypothesis": some tasks can focus S more on stimulus A than on B (the natural salience of A vs. B can also be a factor) → more weight for distinctive features of A → model predicts asymmetries

"Canada is like the US" > "The US is like Canada"

Reconsideration of similarity

Similarity as feature-matching:

- Tversky's (1977) "contrast model" of similarity (a.k.a. the "feature-matching" model):

$$S(a,b) = \theta\, f(A \cap B) - \alpha\, g(A\text{-}B) - \beta\, g(B\text{-}A)$$

"common features" "distinctive features" (of A) (of B)

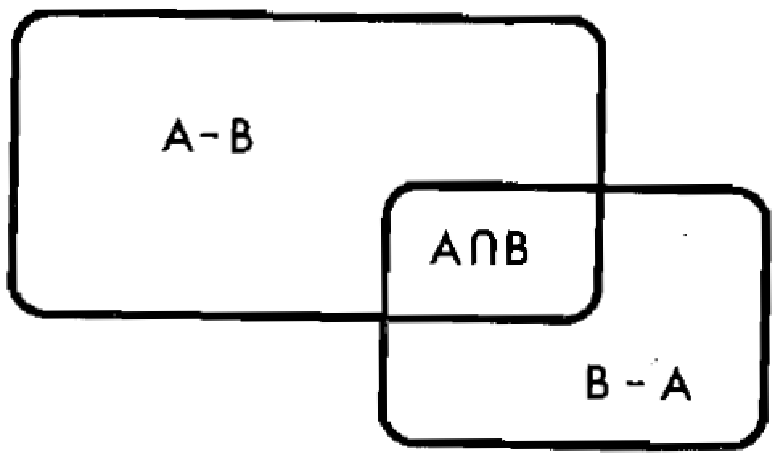


Figure 1. A graphical illustration of the relation between two feature sets.

Contrast model (cont.)

- This model can predict / model asymmetric similarity
- It can also explain why similarity and dissimilarity judgments can differ, and are not simply inverses of each other. Namely, these tasks differ in how they lead raters to emphasize common features vs. distinctive features.

Psychology of judgment:

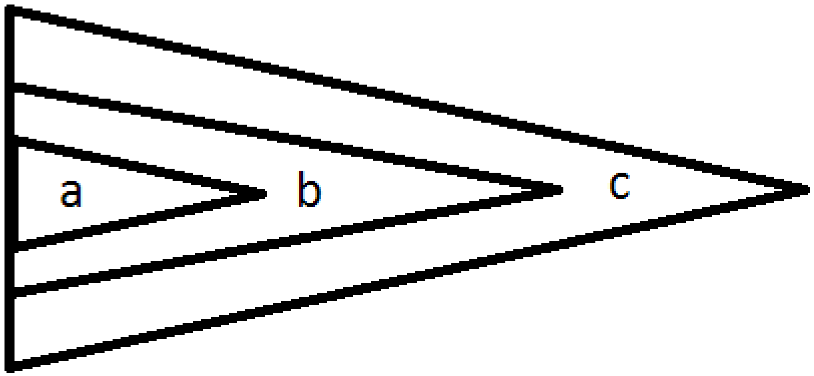
judgments of similarity → emphasize common features

judgments of dissimilarity → emphasize distinctive features

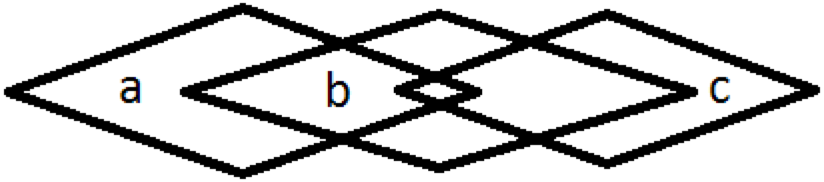
Contrast model (cont.)

Note that we can model continuous graded or “dimensional” differences among stimuli with discrete feature structures (Gati & Tversky, 1982):

Quantitative dimension: objects ~ nested feature sets



Qualitative dimension: objects ~ chains of feature sets



Contrast model (cont.)

An identifiability issue:

NOTE: If the functions f and g are simple additive set functions that sum the weights or “measure” of each feature set, then a common features model can always be represented as a distinctive features model and vice-versa (though the feature sets may need to be “extended”) (Sattath & Tversky, 1987)

Contrast model: A representation theorem

In addition to matching (1), monotonicity (2), and independence (3), we also assume solvability (4), and invariance (5). Solvability requires that the feature space under study be sufficiently rich that certain (similarity) equations can be solved. Invariance ensures that the equivalence of intervals is preserved across factors. A rigorous formulation of these assumptions is given in the Appendix, along with a proof of the following result.

Representation theorem. Suppose Assumptions 1, 2, 3, 4, and 5 hold. Then there exist a similarity scale S and a nonnegative scale f such that for all a,b,c,d in Δ,

(i). $S(a,b) \geq S(c,d)$ iff $s(a,b) \geq s(c,d)$;

(ii). $S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, for some $\theta, \alpha, \beta \geq 0$;

(iii). f and S are interval scales.

A “structural alignment” view of similarity (e.g. Markman & Gentner, 1993)

Objects or entities are understood as “<u>structured objects</u>” (a term from computer science), usually referred to as *scripts* for events, *frames* for objects and other concepts.

Example:

Computer:

Type: laptop ←”Type“ = *slot*, “laptop” = *slot value*

Maker: Dell

Model: Latitude D830

Operating System: Windows 7

etc.

Markman & Genter model (cont.)

How are structured objects compared?

Ss try to align the frames for two objects. Some slots will appear in both object frames, some will not. This leads to a distinction between "alignable differences" and "non-alignable differences"

Similarity is assumed to increase with the number (and salience) of alignable differences, decrease with number of non-alignable differences

→ Computational model = "structure-mapping engine" (Forbus et al.)

→ Another model for mapping of structured objects: LISA model (Hummel & Holyoak)

Similarity as Transformation

- In some applications, similarity as spatial distance or similarity as feature-matching do not seem completely appropriate. For some such cases, the similarity between two stimuli might be a function of how easily one stimulus can be transformed into another

  (E.g. Hahn, Chater & Richardson, 2003; Larkey & Markman, 2005)

  This approach seems natural to model / explain asymmetric similarities.

- Related: the idea of “graph-edit distance” in computer science (since an edit is a transformation).

Examples: How similar are the strings “ABCDEFG” and “ABCFG”, or “I eat fish on Fridays” vs. “On Fridays I eat fish”.

II. Computing Similarity Measures from Multivariate Data

If we wish to use a method for analyzing proximity data with rectangular multivariate data (subjects x variables), we may need to start by computing a proximity measure (similarity or association).

Example: (imagine profiles of 35 subjects on 12 scales, items, or other variables, e.g. a symptom checklist)

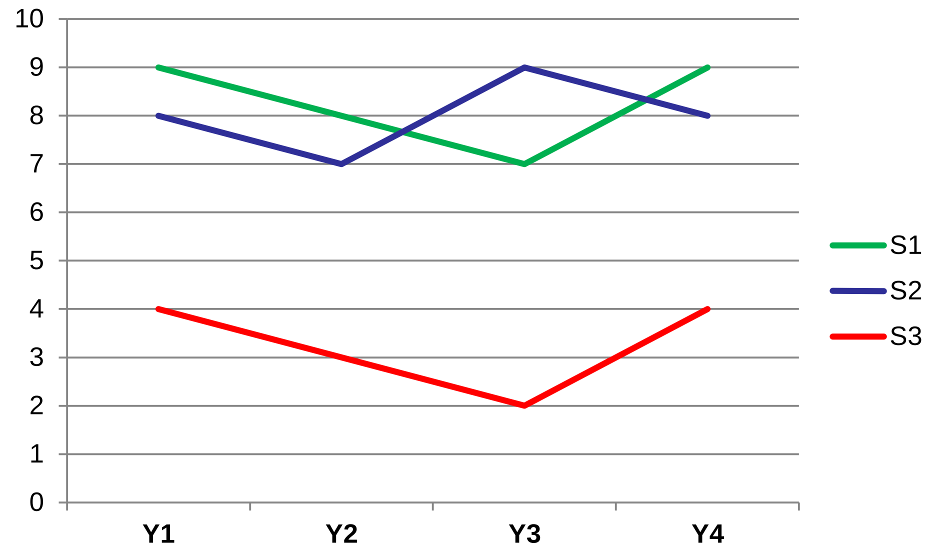
Two frequently used "similarity coefficients":

1) Compute correlations among variables
2) Compute Euclidean distances among variables

Another example: How about if we want to study proximities among *subjects*?

Example: A profile similarity problem

<u>Multivariate data</u>:
3 cases (S1-S3) measured on 4 numeric variables (Y1-Y4) (e.g., attitude subscales)



Data:	Y1	Y2	Y3	Y4			
S1	9	8	7	9			
S2	8	7	9	8			
S3	4	3	2	4			
						D(S1,S3)=	10.00

Types of explicit proximity measures

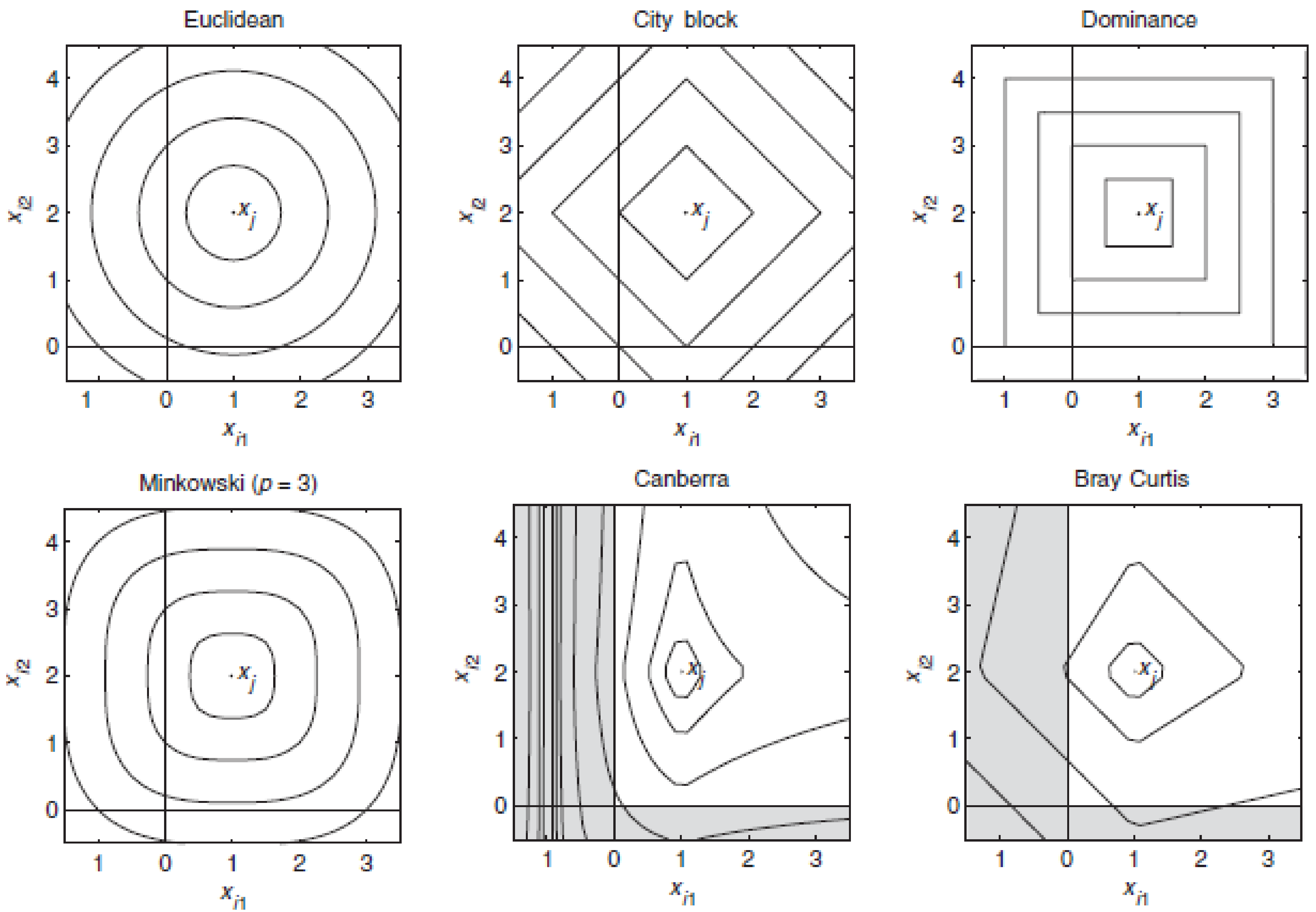
1. correlation measures (Pearson, phi, rank-order, etc.)
2. distance measures (Euclidean, city-block, other Minkowski metrics; Mahalonobis distance, etc.)
3. association measures (phi, lambda, etc)
4. probabilistic similarity measures (e.g., based on information, uncertainty)

Some proximity measures (Table 6.2, MMDS)

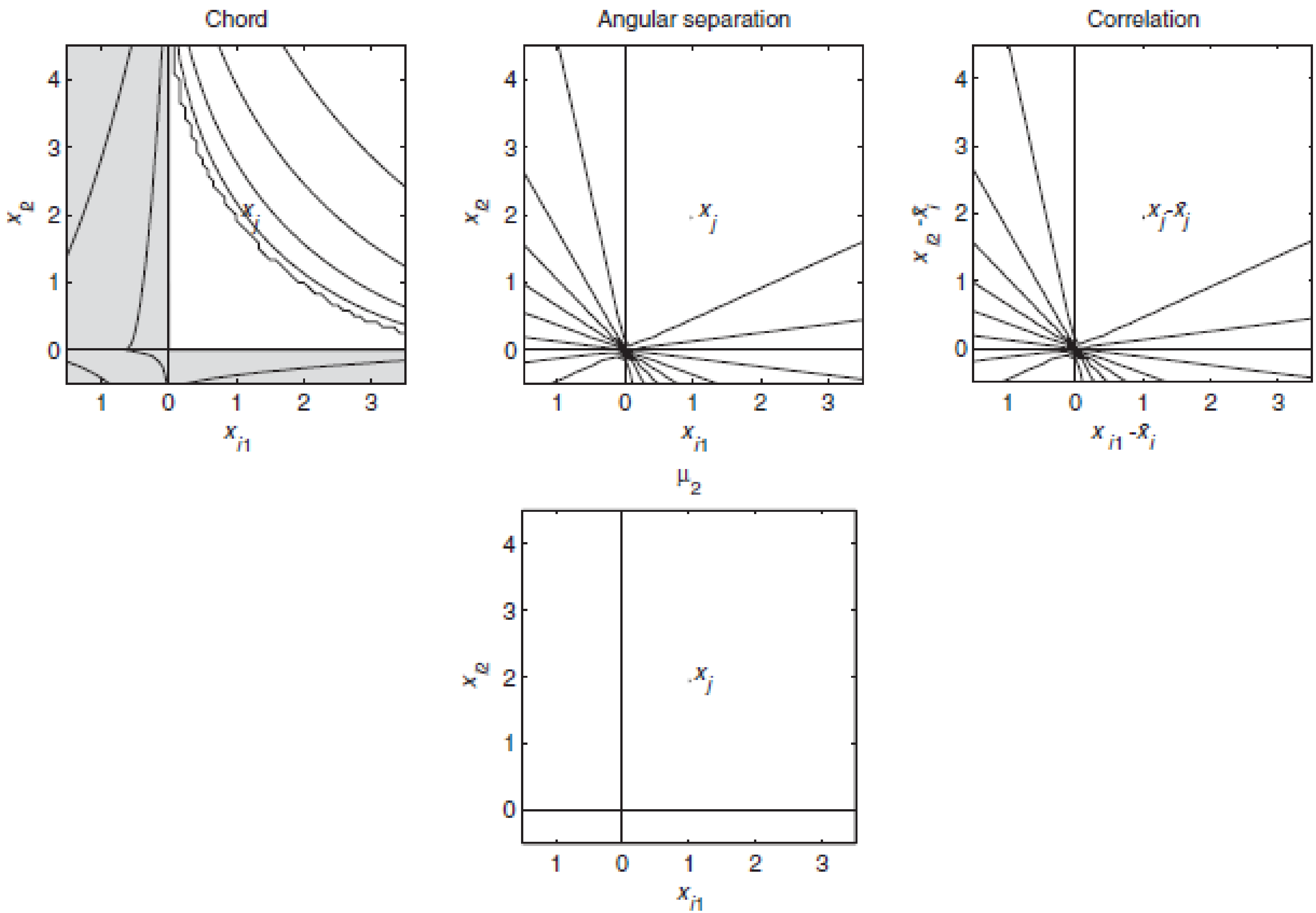
TABLE 6.2. Summary of measures of proximities derived from attribute data. The symbol δ_{ij} denotes a dissimilarity and s_{ij} a similarity.

	Measure	Formula
P1	Euclidean distance	$\delta_{ij} = \left(\sum_{a=1}^{m}(x_{ia} - x_{ja})^2\right)^{1/2}$
P2	City-block distance	$\delta_{ij} = \sum_{a=1}^{m} \lvert x_{ia} - x_{ja} \rvert$
P3	Dominance distance	$\delta_{ij} = \max_{a=1}^{m} \lvert x_{ia} - x_{ja} \rvert$
P4	Minkowski distance	$\delta_{ij} = \left(\sum_{a=1}^{m}(x_{ia} - x_{ja})^p\right)^{1/p}$ with $p \geq 1$
P5	Canberra distance	$\delta_{ij} = \sum_{a=1}^{m} \frac{\lvert x_{ia} - x_{ja} \rvert}{\lvert x_{ia} + x_{ja} \rvert}$
P6	Bray–Curtis distance	$\delta_{ij} = \frac{\sum_{a=1}^{m} \lvert x_{ia} - x_{ja} \rvert}{\sum_{a=1}^{m}(x_{ia} + x_{ja})}$
P7	Chord distance	$\delta_{ij} = \left(\sum_{a=1}^{m}(x_{ia}^{1/2} - x_{ja}^{1/2})\right)^{1/2}$
P8	Angular separation, congruence coefficient	$s_{ij} = \frac{\sum_{a=1}^{m} x_{ia}x_{ja}}{\left(\sum_{a=1}^{m} x_{ia}^2\right)^{1/2}\left(\sum_{a=1}^{m} x_{ja}^2\right)^{1/2}}$
P9	Correlation	$s_{ij} = \frac{\sum_{a=1}^{m}(x_{ia} - \bar{x}_i)(x_{ja} - \bar{x}_j)}{\left(\sum_{a=1}^{m}(x_{ia} - \bar{x}_i)^2\right)^{1/2}\left(\sum_{a=1}^{m}(x_{ja} - \bar{x}_j)^2\right)^{1/2}}$
P10	Monotonicity coefficient μ_2	$s_{ij} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - x_j)(y_i - y_j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \lvert x_i - x_j \rvert \lvert y_i - y_j \rvert}$

Isosimilarity contours for proximity measures (MMDS)



Isosimilarity contours for proximity measures (cont.)



Gower's coefficient

When collecting object $\times$ attribute data sets in real life, some attributes may be binary; others may be numerical. The general similarity measure of Gower (1971) is particularly suited for this situation. Let s_{ija} be the similarity between objects i and j on variable a. For binary attributes, we assume that only values $x_{ia} = 0$ and $x_{ia} = 1$ occur. In this case, $s_{ija} = 1$ if x_{ia} and x_{ja} fall in the same category and $s_{ija} = 0$ if they do not. If the attribute is numerical, then we compute $s_{ija} = 1 - |x_{ia} - x_{ja}|/r_k$ with r_k being the range of attribute a. This definition ensures again that $0 \leq s_{ija} \leq 1$ for all combinations of i, j, and a. The general similarity measure can be defined by

$$s_{ij} = \frac{\sum_a w_{ija} s_{ija}}{\sum_a w_{ija}},$$

where the w_{ija} are given nonnegative weights. Usually w_{ija} is set to one for all i, j, and a. However, if either x_{ia} or x_{ja} is missing (or both), then w_{ija} should be set to zero so that the missing values do not influence the similarity. Here, too, $0 \leq s_{ij} \leq 1$ so that dissimilarities can be obtained by taking $1 - s_{ij}$. However, Gower (1971) suggests to use $(1 - s_{ij})^{1/2}$ as it can be shown that these values can be perfectly represented in a Euclidean space of high dimensionality.

Mahalanobis distance

- Useful when computing distance from highly correlated variables

$$d_{x,y} = \sqrt{\left(\underline{x} - \underline{y}\right)^T S^{-1}(\underline{x} - \underline{y})}$$

Where $\underline{x}$ and $\underline{y}$ are the vectors of multivariate data for objects x and y, and S is the (sample) variance-covariance matrix.

If the variables are independent, and standardized to have unit variances, then the Mahalanobis distance reduces to the Euclidean distance.

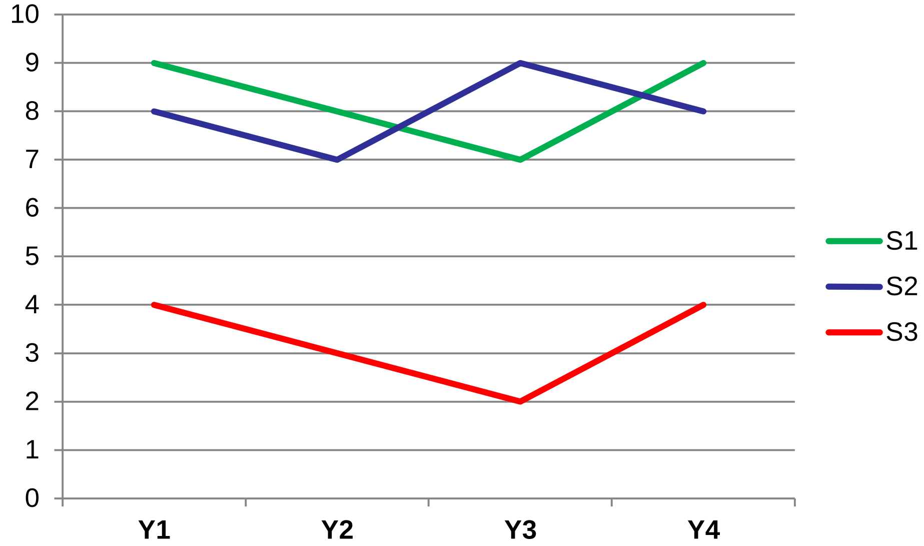
Naturally, the type of data makes a difference in appropriate choice of a measure:

- Nominal, ordinal, interval/ratio, or <u>mixed</u>?
- Discrete or continuous (or a mix)?

In addition, what do we want the measure to be especially sensitive to?

- Profile similarity (→ correlation)
- Level or height of the profile (→ Euclidean distance)
- Common elements / distinctive elements
- Association (how defined?)

Example: Correlation vs. Euclidean Distance as measures of profile similarity



					Pearson Correlation:	
Data:	Y1	Y2	Y3	Y4	R(S1,S2)=	-0.43
S1	9	8	7	9	R(S1,S3)=	1.00
S2	8	7	9	8	Euclidean distance:	
S3	4	3	2	4	D(S1,S2)=	2.65
					D(S1,S3)=	10.00

Some Measures of Association (co-occurrence)

refs: Sneath & Sokal, 1973; Gower & Legendre, 1986

When we have data on the co-occurrence of certain features or aspects across stimuli, we can compute a <u>measure of association</u> to measure the tendency of features to co-occur (or not) in a stimulus

- Perhaps the most familiar such measure is the *phi coefficient*, which is simply the Pearson correlation computed on binary (0,1) data. However, the phi coefficient can also be computed using the cell frequencies in the corresponding *2x2* table (below), as:

$$\emptyset = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

		Y: 1	Y: 0	
X:	1	a	b	(a+b)
	0	c	d	(c+d)
		(a+c)	(b+d)	

- A closely related measure is the *Yule coefficient*:

$$c_Y = \frac{ad-bc}{ad+bc}$$

Effect of base rate of "negative match" on value of ø

	R2: 1	0			phi
R1: 1	4	1	5		0.60
0	1	4	5		
	5	5	10		

	R2: 1	0			phi
R1: 1	4	1	5		0.71
0	1	10	11		
	5	11	16		

	R2: 1	0			phi
R1: 1	4	1	5		0.78
0	1	50	51		
	5	51	56		

	R2: 1	0			phi
R1: 1	4	1	5		0.79
0	1	100	101		
	5	101	106		

	R2: 1	0			phi
R1: 1	4	1	5		0.80
0	1	1000	1001		
	5	1001	1006		

Other measures of association

- A common measure of association is the *simple matching coefficient* (Sokal and Michener,1958):

$$c_m = \frac{a+d}{(a+b+\ c+d)}$$

		Y: 1	0	
X:	1	a	b	(a+b)
	0	c	d	(c+d)
		(a+c)	(b+d)	

- But if "null matches" (d) are less informative than positive matches (a), then another measure such as the *Jaccard coefficient* (Jaccard, 1908) may be more appropriate: $c_J\ = \frac{a}{(a+b+c)}$

Note that for multistate (nominal) variables or attributes, it makes sense to apply the simple matching idea to <u>each</u> value of the attribute, i.e. on *all* the diagonal cells.

(But as the table gets bigger, chance agreement is lower, so *Cohen's kappa* is preferred in contexts of larger tables, where a measure of agreement is desired)

Other measures of association

Gower's coefficient (Gower, 1971) is applicable to binary, to multi-state, and to quantitative attributes. Given *k* attributes or "dimensions", the overall similarity between stimuli i and j is defined as: $S_{ij} = \frac{\sum_k^K w_{ijk} s_{ijk}}{\sum_k^K w_{ijk}}$

where w_{ijk} is a weight that indicates if the comparison is "valid" (e.g., it is normally set to 0 for negative matches of a binary attribute); and $0 \leq S_{ijk} \leq 1$. For quantitative attributes, $s_{ijk} = 1 - (|X_{ij} - X_{ik}|/R_i)$, where R_i is the *range* of the quantitative attribute.

Interestingly, a matrix of similarities computed by this measure is positive semi-definite (Gower, 1971), and has a distance realization defined by $d_{jk} = \sqrt{1 - s_{ij}}$

References: Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27(4), 857-871.*

Gower, J.C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 3,* 5-48.

Informational Measures of Association

- see Sneath & Sokal (1973) or Orloci (1969) for proposed coefficients and discussion
- Example: the *mutual information* between two discrete RVs (asymmetric version: *directed information*)

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right),$$

- Example: the *Goodman-Kruskal lambda* measure can be understood in terms of information. This is true of other probabilistic association measures as well.