HUDM5124 Spring 2020 Assn. 1

Exercises on metric spaces and distance:

In the posted notes for Lecture 1, the following axioms are given. If you are interested, further discussion of the axioms is given in the two posted extracts from the book *What is Distance?* by Yuri Shreider, and in the cited chapter section in the *Modern Multidimensional Scaling* book.

These axioms are constraints that must be satisfied in order for a set of pairwise measures on a set to be considered as distances in a metric space. We will use the following set of axioms (note this is not the minimal axiomatization, but it is clear). The distances must satisfy:

a. symmetry ($d_{xy} = d_{yx}$) *(the distance of x to y must equal the distance of y to x)*

b. positivity ($d_{xy} > 0$) *(the distance between any two distinct points must be positive)*

c. minimality ($d_{xx} = 0$) *(the distance of a point to itself is = 0)*

d. the triangle inequality ($d_{xy} + d_{yz} \geq d_{xz}$) *(for the distances among any three points, the sum of any two of these distances must be equal to or greater than the third)*

In order for a geometric space (or any other kind of distance model) to make sense as a model for proximity data, the data ought to "look like" distances, i.e. the dissimilarities ought to satisfy the metric axioms as well.

Please answer the following questions 1-3 for next week.

DISTANCES IN AN ACTUAL SPACE.

1. Consider five points (labeled A-E) on a line. Discuss if each of the axioms above is satisfied for the actual interpoint distances.

PROXIMITY DATA. In the first course session, I have posted three example proximity data sets: the Rothkopf Morse code data (confusions among pairs of Morse code signals), Ekman's data on similarity ratings of color chips, and a set of ratings of the dissimilarity among emotion terms that I gather some years ago.

2. For the emotions dissimilarities, discuss which of these axioms seem to be satisfied and which do not.

COMMENT: to prove that a property is satisfied, you would need to check the property for EVERY case. So for minimality, you would need to check every diagonal entry of the matrix, if the diagonal entries are given. For the triangle inequality, you would need to check EVERY triple of objects i,j,k to see if dij + djk <= dik. Furthermore, you need to *check also that dij + dik <= djk and that dik + djk <= dij. This is the sort of task that you could write code to check. But no coding is required for this assignment.]*

3. For the Rothkopf Morse code confusions, discuss which of these axioms seem to be satisfied and which do not. Note that confusions data may be considered as similarities. Thus, you will have to "translate" the distance axioms as appropriate for similarity data; symmetry still applies, but "minimality" becomes "maximality" for the diagonal elements; the triangle inequality, translated for similarities, would seem to imply that if S(a,b) and S(b,c) are both very large, then S(a,c) must also be large.

COMMENT: So one can't really find VIOLATIONS of the triangle inequality for similarities, just "suspicious" triples.

COMMENT 2: If one transformed the similarity data into dissimilarities, for example by subtracting every element in the similarity matrix from a large constant (e.g. from 100 for the Morse code data), then one could check the metric axioms directly on the transformed numbers. But the assignment does not ask you to do that. And it turns out that what constant you pick makes a difference.

BTW, the task that was used to gather these data was a "same-different" task, i.e. a pair of Morse code signals was presented (very rapidly) to the subject, and he or she had to say if the two tones were the same signal or different signals. The table shows the percentage of "same" responses (which is correct for the diagonal elements, and incorrect (a confusion) for the off-diagonal elements.

OPTIONAL EXERCISES (more mathematical):

4. Prove that the triangle inequality generalizes to more than three points; i.e. that for points x,y,z,w, $d(x,w) \le d(x,y) + d(y,z) + d(z,w)$

5. Imagine a flat landscape with circular lakes of varying sizes scattered through it. You cannot travel through the lakes, only over land. Would the metric axioms hold for distances in this space? You can offer formal proofs, or merely arguments.