Abstract:
This is an entry for The Encyclopedia of Statistics in Behavioral Science, to be published by Wiley in 2005.

UNIDIMENSIONAL SCALING

JAN DE LEEUW

Abstract. This is an entry for The Encyclopedia of Statistics in Behavioral Science, to be published by Wiley in 2005.

Unidimensional scaling is the special one-dimensional case of **multidimensional scaling** [5]. It is often discussed separately, because the unidimensional case is quite different from the general multidimensional case. It is applied in situations where we have a strong reason to believe there is only one interesting underlying dimension, such as time, ability, or preference. And unidimensional scaling techniques are very different from multidimensional scaling techniques, because they use very different algorithms to minimize their loss functions.

The classical form of unidimensional scaling starts with a symmetric and non-negative matrix $\Delta = \{\delta_{ij}\}$ of *dissimilarities* and another symmetric and non-negative matrix $W = \{w_{ij}\}$ of *weights*. Both W and Δ have a zero diagonal. Unidimensional scaling finds *coordinates* x_i for n points on the

Date: March 27, 2004.

Key words and phrases. fitting distances, multidimensional scaling.

line such that

$$\sigma(x)=\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n w_{ij}(\delta_{ij}-|x_i-x_j|)^2$$

is minimized. Those n coordinates in x define the scale we were looking for.

To analyze this unidimensional scaling problem in more detail, let us start with the situation in which we know the order of the x_i, and we are just looking for their scale values. Now $|x_i-x_j|=s_{ij}(x)(x_i-x_j)$, where $s_{ij}(x)=\textbf{sign}(x_i-x_j)$. If the order of the x_i is known, then the $s_{ij}(x)$ are known numbers, equal to either -1 or $+1$ or 0, and thus our problem becomes minimization of

$$\sigma(x)=\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n w_{ij}(\delta_{ij}-s_{ij}(x_i-x_j))^2$$

over all x such that $s_{ij}(x_i-x_j)\geq 0$. Assume, without loss of generality, that the weighted sum of squares of the dissimilarities is one. By expanding the sum of squares we see that

$$\sigma(x)=1-t't+(x-t)'V(x-t).$$

Here V is the matrix with off-diagonal elements $v_{ij}=-w_{ij}$ and diagonal elements $v_{ii}=\sum_{j=1}^n w_{ij}$. Also, $t=V^+r$, where r is the vector with elements $r_i=\sum_{j=1}^n w_{ij}\delta_{ij}s_{ij}$, and V^+ is the generalized inverse of V. If all the off-diagonal weights are equal we simply have $t=r/n$.

Thus the unidimensional scaling problem, with a known scale order, requires us to minimize $(x-t)'V(x-t)$ over all x satisfying the order restrictions. This is a **monotone regression** problem [2], which can be solved quickly and uniquely by simple quadratic programming methods.

Now for some geometry. The vectors x satisfying the same set of order constraints form a polyhedral convex cone $\mathcal{K}$ in $\mathbb{R}^n$. Think of $\mathcal{K}$ as an ice cream cone with its apex at the origin, except for the fact that the ice cream cone is not round, but instead bounded by a finite number hyperplanes. Since there are $n!$ different possible orderings of x, there are $n!$ cones, all with their apex at the origin. The interior of the cone consists of the vectors without ties, intersections of different cones are vectors with at least one tie. Obviously the union of the $n!$ cones is all of $\mathbb{R}^n$.

Thus the unidimensional scaling problem can be solved by solving $n!$ monotone regression problems, one for each of the $n!$ cones [3]. The problem has a solution which is at the same time very simple and prohibitively complicated. The simplicity comes from the $n!$ subproblems, which are easy to solve, and the complications come from the fact that there are simply too many different subproblems. Enumeration of all possible orders is impractical for $n > 10$, although using combinatorial programming techniques makes it possible to find solutions for n as large as 20 [7].

Actually, the subproblems are even simpler than we suggested above. The geometry tells us that we solve the subproblem for cone $\mathcal{K}$ by finding the closest vector to t in the cone, or, in other words, by projecting t on the cone. There are three possibilities. Either t is in the interior of its cone, or on the boundary of its cone, or outside its cone. In the first two cases t is equal to its projection, in the third case the projection is on the boundary. The general result in [1] tells us that the loss function σ cannot have a local minimum at a point in which there are ties, and thus local minima can only occur in the interior of the cones. This means that we can only have a local minimum if t is in the interior of its cone, and it also means that we actually never have to compute the monotone regression. We just have to verify if t is in the interior, if it is not then σ does not have a local minimum in this cone.

There have been many proposals to solve the combinatorial optimization problem of moving through the $n!$ until the global optimum of σ has been found. A recent review is [8].

We illustrate the method with a simple example, using the vegetable paired comparison data from [6, page 160]. Paired comparison data are usually given in a matrix P of proportions, indicating how many times stimulus i is preferred over stimulus j. P has 0.5 on the diagonal, while corresponding elements p_{ij} and p_{ji} on both sides of the diagonal add up to 1.0. We transform

the proportions to dissimilarities by using using the probit transformation $z_{ij} = \Phi^{-1}(p_{ij})$ and then defining $\delta_{ij} = |z_{ij}|$. There are 9 vegetables in the experiment, and we evaluate all $9! = 362880$ permutations. Of these cones 14354 or 4% have a local minimum in their interior. This may be a small percentage, but the fact that σ has 14354 isolated local minima indicates how complicated the unidimensional scaling problem is. The global minimum is obtained for the order given in Guilford's book, which is Turnips $<$ Cabbage $<$ Beets $<$ Asparagus $<$ Carrots $<$ Spinach $<$ String Beans $<$ Peas $<$ Corn. Since there are no weights in this example, the optimal unidimensional scaling values are the row averages of the matrix with elements $s_{ij}(x)\delta_{ij}$. Except for a single sign change of the smallest element (the Carrots and Spinach comparison), this matrix is identical to the probit matrix Z. And because the Thurstone Case V scale values are the row averages of Z, they are virtually identical to the unidimensional scaling solution in this case.

The second example is quite different. It has weights and incomplete information. We take it from a paper by Fisher [4], in which he studies crossover percentages of eight genes on the sex chromosome of *Drosophila willistoni*. He takes the crossover percentage as a measure of distance, and supposes the number n_{ij} of crossovers in N_{ij} observations is binomial. Although there are eight genes, and thus $\binom{8}{2} = 28$ possible dissimilarities, there are only 15 pairs that are actually observed. Thus 13 of the off-diagonal weights

are zero, and the other weights are set to the inverses of the standard errors of the proportions. We investigate all $8! = 40320$ permutations, and we find 78 local minima. The solution given by Fisher, computed by solving linearized likelihood equations, has Reduced $<$ Scute $<$ Peach $<$ Beaded $<$ Rough $<$ Triple $<$ Deformed $<$ Rimmed. This order corresponds with a local minimum of σ equal to 40.16. The global minimum is obtained for the permutation that interchanges Reduced and Scute, with value 35.88. In Figure 1 we see the scales for the two local minima, one corresponding with Fisher's order and the other one with the optimal order.

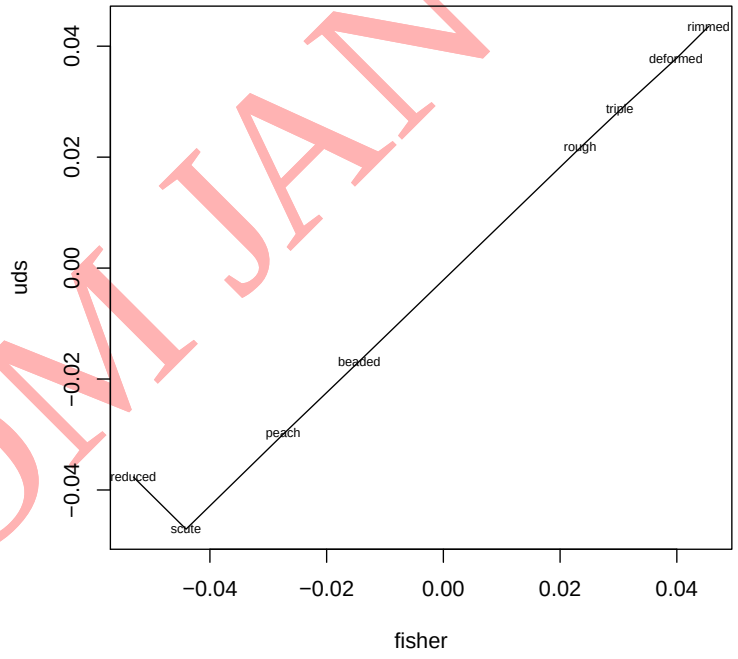

Figure 1. Genes on Chromosome

In this entry we have only discussed least squares metric unidimensional scaling. The first obvious generalizations are to replace the least squares loss function, for example by the least absolute value or ℓ_1 loss function. The second generalization is to look at nonmetric unidimensional scaling. These generalizations have not been studied in much detail, but in both we can continue to use the basic geometry we have discussed. The combinatorial nature of the problem remains intact.

References

[1] J. De Leeuw. Differentiability of Kruskal's Stress at a Local Minimum. *Psychometrika*, 49:111–113, 1984.

[2] J. De Leeuw. Monotone Regression. In B. Everitt and D. Howell, editors, *The Encyclopedia of Statistics in Behavioral Science*. Wiley, 2005.

[3] J. De Leeuw and W. J. Heiser. Convergence of Correction Matrix Algorithms for Multidimensional Scaling. In J.C. Lingoes, editor, *Geometric representations of relational data*, pages 735–752. Mathesis Press, Ann Arbor, Michigan, 1977.

[4] R.A. Fisher. The Systematic Location of Genes by Means of Crossover Observations. *American Naturalist*, 56:406–411, 1922.

[5] P. Groenen. Multidimensional scaling. In B. Everitt and D. Howell, editors, *The Encyclopedia of Statistics in Behavioral Science*. Wiley,

2005.

[6] J.P. Guilford. *Psychometric Methods*. McGrawHill, second edition, 1954.

[7] L.J. Hubert and P. Arabie. Unidimensional Scaling and Combinatorial Optimization. In J. De Leeuw, W. Heiser, J. Meulman, and F. Critchley, editors, *Multidimensional data analysis*, Leiden, The Netherlands, 1986. DSWO-Press.

[8] L.J. Hubert, P. Arabie, and J.J. Meulman. Linear Unidimensional Scaling in the L_2-Norm: Basic Optimization Methods Using MATLAB. *Journal of Classification*, 19:303–328, 2002.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu