LOGISTIC UNFOLDING

JAN DE LEEUW

1. Problem

Suppose the data are categorical and coded as indicator matrices, also known as dummies. The indicator matrix G_j for variable j has n rows and k_j columns. G_j is a binary matrix, and its rows all sum to one. As in other forms of unfolding we represent both the n objects and the k_j categories of variable j as points a_i and b_{jl} in low-dimensional Euclidean space.

Minimize

$$\Delta(A, B) = -2 \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{\ell=1}^{k_j} g_{ij\ell} \log \pi_{ij\ell}(A, B), \tag{1a}$$

where

$$\pi_{ijl}(A, B) = \frac{\exp(-\|a_i - b_{j\ell}\|)}{\sum_{\nu=1}^{k_j} \exp(-\|a_i - b_{j\nu}\|)}. \tag{1b}$$

2. Algorithm

To minimize the loss function we use quadratic majorization [Böhning and Lindsay, 1988; De Leeuw, in press]. We need the first and the second derivatives of the deviance with respect to the $d_{ij\ell}(A, B) = \|a_i - b_{j\ell}\|$. Simple computation gives

$$\frac{\partial \pi_{ij\ell}}{\partial d_{ij\nu}} = -(\pi_{ij\ell}\delta^{\ell\nu} - \pi_{ij\ell}\pi_{ij\nu}) \tag{2a}$$

Date: July 19, 2005.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

with $\delta^{\ell\nu}$ the Kronecker Delta. Thus

$$\frac{\partial\Delta}{\partial d_{ij\ell}} = 2(g_{ij\ell} - \pi_{ij\ell}) \tag{2b}$$

and

$$\frac{\partial^2\Delta}{\partial d_{ij\ell}\partial d_{ij\nu}} = 2(\pi_{ij\ell}\delta^{\ell\nu} - \pi_{ij\ell}\pi_{ij\nu}). \tag{2c}$$

Now lets look at any matrix of the form $V = \Pi - \pi\pi'$, with π a vector of probabilities and with Π the diagonal matrix with these probabilities on the diagonal. The largest eigenvalue λ_{max} of this matrix is bounded above by any matrix norm, and thus

$$\lambda_{max} \leq \max_{i=1}^{n} \sum_{j=1}^{n} |v_{ij}| = \max_{i=1}^{n} 2\pi_i(1 - \pi_i) \leq \frac{1}{2}.$$

Thus from a Taylor expansion at $d(\tilde{A}, \tilde{B})$, using the bound for the second derivatives,

$$\begin{aligned} \Delta(A, B) \leq \Delta(\tilde{A}, \tilde{B}) + &\\ + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\ell=1}^{k_j}(g_{ij\ell} - \pi_{ij\ell}(\tilde{A}, \tilde{B}))(d_{ij\ell}(A, B) - d_{ij\ell}(\tilde{A}, \tilde{B})) + &\\ + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\ell=1}^{k_j}(d_{ij\ell}(A, B) - d_{ij\ell}(\tilde{A}, \tilde{B}))^2. \end{aligned} \tag{3a}$$

If we define the *target*

$$\tau_{ij\ell}(A, B) = d_{ij\ell}(A, B) - 2(g_{ij\ell} - \pi_{ij\ell}(A, B))$$

then completing the square in (3a) gives

$$\begin{aligned} \Delta(A, B) \leq \Delta(\tilde{A}, \tilde{B}) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\ell=1}^{k_j}(d_{ij\ell}(A, B) - \tau_{ij\ell}(\tilde{A}, \tilde{B}))^2 + &\\ - 2\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\ell=1}^{k_j}(g_{ij\ell} - \pi_{ij\ell}(\tilde{A}, \tilde{B}))^2 \end{aligned} \tag{3b}$$

Iteration (k) of the majorization algorithm updates $(A^{(k)}, B^{(k)})$ by minimizing the least squares auxilary function

$$\sigma^{(k)}(A,B) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\ell=1}^{k_j}(d_{ij\ell}(A,B) - \tau_{ij\ell}(\tilde{A},\tilde{B}))^2 \tag{4}$$

3. Metric Unfolding

Minimizing the least squares loss function in (4) is a metric multidimensional unfolding problem [De Leeuw, 2005]. It is somewhat non-standard, because the target values in τ may be negative. We solve the unfolding problem, using majorization, following Heiser [1990].

Decompose $\tilde{\tau}_{ij\ell} = \tau_{ij\ell}(\tilde{A},\tilde{B})$ in its positive and negative parts. Thus $\tilde{\tau}_{ij\ell} = \tilde{\tau}^{+}_{ij\ell} - \tilde{\tau}^{-}_{ij\ell}$. We also stack A and B on top of each other in a matrix Z, and we define the matrices

$$d_{ij}(Z) = \sqrt{\mathbf{tr}\, Z'E_{ij}Z}$$

$$\frac{1}{d_{ij}(\tilde{Z})}\mathbf{tr}\, Z'E_{ij}\tilde{Z} \leq d_{ij}(Z) \leq \frac{1}{2}\frac{1}{d_{ij}(\tilde{Z})}(d_{ij}^2(Z) + d_{ij}^2(\tilde{Z})$$

$$\sigma(Z) \leq \mathbf{tr}Z'VZ - 2\mathbf{tr}Z'\frac{\tau^{+}_{ij}}{d_{ij}(\tilde{Z})}E_{ij}\tilde{Z}+$$

References

D. Böhning and B.G. Lindsay. Monotonicity of Quadratic-approximation Algorithms. *Annals of the Institute of Statistical Mathematics*, 40(4): 641–663, 1988.

J. De Leeuw. Principal Component Analysis of Binary Data by Iterated Singular Value Decomposition. *Computational Statistics and Data Analysis*, in press.

J. De Leeuw. Multidimensional Unfolding. In B.S. Everitt and D.C. Howell, editors, *Encyclopedia of Statistics in Behavioral Science*, volume 3, pages 1289–1294. Wiley, 2005.

W.J. Heiser. A Generalized Majorization Method for Least Squares Multidimensional Scaling of Pseudodistances that May Be Negative. *Psychometrika*, 56:7–27, 1990.

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`