Multidimensional Scaling ~ Cox & Cox

1. Introduction

This suite of programs carries out multidimensional scaling techniques described in the research monograph, "Multidimensional Scaling", by Cox, TF and Cox, MAA (2001), CRC/Chapman and Hall. Pages of the monograph are referred to in the description of the programs below. The monograph is not needed for using the programs, but would be useful for descriptions of multidimensional scaling techniques and their applications. The files **licence.doc**, **licence.ps** and **licence.txt** contain the user licence. An extended version of this document is available as **readme.doc**, **readme.ps** and **readme.txt**. The software may be obtained from the authors for £350.

2. Contact Details

The authors may be contacted via mds.research@newcastle.ac.uk. A web site associated with this software is under development at http://www.ncl.ac.uk/mds.

3. The Menu Program

There are three main groups of programs within the **menu**: *Data Input/Preparation*, *Data Analysis and Data Presentation.* Users would usually pre-process raw data using the Data Input/Preparation menu. This data may then be evaluated in the Data Analysis menu. Finally graphical output may be obtained from the Data Presentation menu. *Control* contains basic information about the program.

4. Data Input/Preparation

Various types of data can be analysed using **MDS**. Files containing data and other information are of the following types. Those with extension:

- **VEC** These contain data, which are observations on random vectors or contain co-ordinates of points in a configuration.
- **DIS** These contain dissimilarities measured between objects or individuals.
- **MAT** These contain integer data, either vector binary observations or contingency table entries.
- **IND** These contain data for a contingency table with each row giving the cell count and the corresponding category for each of the variables. Note that this is different from the usual meaning of an indicator matrix.
- **DEG** These contain the position angles for objects placed on the surface of a sphere.
- **SIM** These contain similarity data for generalised inner product scaling.
- **SHE** These files are created for Shepard plots.
- **UNF** These files are created for unfolding analysis.

The subdirectory ***data*** contains various data sets that can be used for analysis. Users can prepare their own files of data for analysis using a file editor, or word processing package (e.g. Word or Notepad), but files produced in this manner must be saved as ASCII files with the format described below. Alternatively data can be entered or edited using Excel. (See 5.13 and 5.14. Excel is also used in 5.9.)

5. Data Input/Preparation Programs

5.1. Contingency Table To Unfolding (pp165-179)

This program prepares a data file for unfolding analysis. The data can be of real vectors (e.g. nations.vec) or of a matrix of integer values (e.g. whisky.dat).

5.2. Dissimilarities From Vector Data

This program calculates dissimilarities from a data matrix and stores them ready for analysis.

The following dissimilarities are available.

Euclidean distance (2 way)	$\delta_{rs} = \sqrt{\sum_i (x_{ri} - x_{si})^2}$
Weighted Euclidean	$\delta_{rs} = \sqrt{\sum_i w_i (x_{ri} - x_{si})^2}$
Mahalanobis distance	$\delta_{rs} = \sqrt{(\mathbf{x}_r - \mathbf{x}_s)^T \sum^{-1} (\mathbf{x}_r - \mathbf{x}_s)}$
City block metric	$\delta_{rs} = \sum_i \lvert x_{ri} - x_{si} \rvert$
Minkowski metric	$\delta_{rs} = \left(\sum_i w_i \lvert x_{ri} - x_{si} \rvert^{\lambda} \right)^{\frac{1}{\lambda}};\ \lambda \geq 1$
Canberra metric	$\delta_{rs} = \sum_i \frac{\lvert x_{ri} - x_{si} \rvert}{x_{ri} + x_{si}}$
Divergence	$\delta_{rs} = \frac{1}{p} \sum_i \frac{(x_{ri} - x_{si})^2}{(x_{ri} + x_{si})^2}$
Bray-Curtis	$\delta_{rs} = \frac{\sum_i \lvert x_{ri} - x_{si} \rvert}{\sum_i (x_{ri} + x_{si})}$
Soergel	$\delta_{rs} = \frac{\sum_i \lvert x_{ri} - x_{si} \rvert}{\sum_i \max(x_{ri}, x_{si})}$
Bhattacharyya distance	$\delta_{rs} = \sqrt{\sum_i \left(\sqrt{x_{ri}} - \sqrt{x_{si}} \right)^2}$
Wave-Hedges	$\delta_{rs} = \sum_i \left(1 - \frac{\min(x_{ri}, x_{si})}{\max(x_{ri}, x_{si})} \right)$
Angular separation	$\delta_{rs} = 1 - \frac{\sum_i x_{ri} x_{si}}{\sqrt{\sum_i x_{ri}^2 \sum_i x_{si}^2}}$
Correlation	$\delta_{rs} = 1 - \frac{\sum_i (x_{ri} - \bar{x}_r)(x_{si} - \bar{x}_s)}{\sqrt{\sum_i (x_{ri} - \bar{x}_r)^2 \sum_i (x_{si} - \bar{x}_s)^2}}$
Euclidean distance (3 way)	$\delta_{rst} = \sqrt{\sum_i (x_{ri} - x_{si})^2 + (x_{ri} - x_{ti})^2 + (x_{si} - x_{ti})^2}$

5.3. Generate Matching Random Data

This program generates random uniform vector data, either matching the number of observations, the number of dimensions and the acronyms to those in an existing *.vec* file, or matching the number of objects and their acronyms in a *.dis* file of dissimilarities. The program is useful for simulating data for comparison purposes.

5.4. Gower's General Dissimilarity (pp14-16)

This program calculates Gower's general dissimilarity coefficients from a data matrix.

5.5. History Data To Dissimilarities (pp240-241)

This program calculates dissimilarities for the historical data (Maidstone.68)

5.6. Indicator Matrix To Contingency Table (pp200-202)

This program converts an indicator matrix (.ind) based on a contingency table to a contingency table.

5.7. Joins Vectors

This program combines two files containing vector data and stores results in another file. This program is useful for plotting two sets of points, e.g. those from an unfolding analysis.

5.8. Matrix To Dissimilarity (pp12-13)

This program calculates dissimilarities from a binary data matrix (.mat file). The following similarities are available.

The identification of the cells for the dissimilarity between objects r and s.

	Object s	
Object r	1	0
1	a	b
0	c	d

Choices are

Braun, Blanque	$s_{rs} = \frac{a}{\max\{(a+b),(a+c)\}}$
Czekanowski, Sorensen, Dice	$s_{rs} = \frac{2a}{2a+b+c}$
Hamman	$s_{rs} = \frac{a-(b+c)+d}{a+b+c+d}$
Jaccard coefficient	$s_{rs} = \frac{a}{a+b+c}$
Kulezynski	$s_{rs} = \frac{a}{b+c}$
Kulezynski	$s_{rs} = \frac{1}{2}\left(\frac{a}{a+b}+\frac{a}{a+c}\right)$
Michael	$s_{rs} = \frac{4(ad-bc)}{(a+d)^2+(b+c)^2}$
Mountford	$s_{rs} = \frac{2a}{a(b+c)+2bc}$
Mozley, Margalef	$s_{rs} = \frac{a(a+b+c+d)}{(a+b)(a+c)}$
Ochiai	$s_{rs} = \frac{a}{\sqrt{(a+b)(a+c)}}$
Phi	$s_{rs} = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$
Rogers, Tanimoto	$s_{rs} = \frac{a+d}{a+2b+2c+d}$
Russell, Rao	$s_{rs} = \frac{a}{a+b+c+d}$
Simple matching coefficient	$s_{rs} = \frac{a+d}{a+b+c+d}$
Simpson	$s_{rs} = \frac{a}{\min\{(a+b),(a+c)\}}$
Sokal, Sneath, Anderberg	$s_{rs} = \frac{a}{a+2(b+c)}$
Yule	$s_{rs} = \frac{ad-bc}{ad+bc}$

5.9. Random Categories For A Vector Via Excel

This program generates random categorical data, which may be useful for MDS or other analyses. The random categorical data are generated from an initial spatial pattern of points, which have to be generated. These are placed in a *.VEC* file in the usual way. Then random hyperplanes are placed within the space containing the configuration of points and used to allocate categories for the random variables. See Cox and Cox (1998) for further details. Note Excel is necessary to run the program, in order to enter weights for the random hyperplanes.

5.10. Random Subset Of Existing Data

This program chooses a random subset of existing data. It is useful when dealing with very large data sets that are too big for some of the MDS analyses.

5.11. Reciprocal Averaging Dissimilarity (pp198-200)

This program calculates dissimilarities from a data matrix as proposed by Cox and Cox (2000).

Data types are:-

Bin	Binary [0,1]
Cts	Continuous
Ord	Ordinal (ranked, first, second, ...)
Cat	Categorical (classification, say colour ...)

5.12. Transpose Matrix Or Vector

This program transposes a matrix or vector and outputs to another file.

5.13. Two-Way Dissimilarity Input Via Excel

This program allows the user to input dissimilarity data using Excel. (Note: this program has been verified on Excel version 7.0 and Excel97). It may not work for earlier versions of Excel.

5.14. Vector Input Via Excel

This program allows the user to input vector data using Excel. (Note: this program has been verified on Excel version 7.0 and Excel97. It may not work for earlier versions of Excel.)

5.15. Vector To Comma Separated Values

This program will read data prepared for input to MDS or data generated by MDS and output the same data to a file using comma-separated values. This is useful for transfering data to other packages.

6. Data Analysis

The following programs analyse data using various multidimensional scaling techniques.

6.1. Biplot (pp153-163)

This program produces biplots from a data matrix.

6.2. Classical Scaling (pp31-49)

This program carries out classical scaling on dissimilarity data.

6.3. Individual Differences Scaling (pp205-211)

This program carries out individual differences scaling.

6.4. Generalized Inner Product Scaling (pp235-236)

This program carries out general inner product scaling (GIPSCAL).

6.5. Least Squares Scaling (pp49-52)

This program carries out metric least squares scaling.

6.6. Metric Unfolding (pp173-179)

This program carries out metric unfolding on two-mode, two-way dissimilarity data. The user may now wish to join the **X** solution vector and the **Y** solution vector files together using "**joins vectors**" (5.7), before plotting.

6.7. Multidimensional Scaling (Two-Way, Three-Way, Spherical)

This program carries out multidimensional scaling.

6.7.1. Two-Way Multidimensional Scaling (pp61-92)

This program carries out nonmetric multidimensional scaling for dissimilarities between pairs of objects.

6.7.2. Three Way Multidimensional Scaling (pp238-243)

This program carries out nonmetric multidimensional scaling for dissimilarities between triples of objects.

6.7.3. Spherical Multidimensional Scaling (pp105-107)

This program carries out nonmetric scaling for dissimilarities defined between pairs of objects, placing the configuration of points on the surface of a sphere.

6.8. Procrustes Analysis (pp123-140)

This program matches one configuration of points to another.

6.9. Reciprocal Averaging (pp193-198)

This program carries out reciprocal averaging.

6.10. Unidimensional Scaling (pp53-57)

This program carries out unidimensional scaling.

7. Data Presentation

These programs display results of multidimensional scaling analyses.

7.1. Linear Biplot (pp153-159)

This program calculates and plots the classical biplot. It is the same as the biplot in 6.1 but includes the plot.

7.2. Non-Linear Biplot (pp159-163)

This program calculates and plots non-linear biplots.

7.3. Movie - Multidimensional Scaling (pp61-74)

This program carries out nonmetric scaling showing the configuration at every step of the algorithm. The user can see the configuration converge towards one with minimum stress.

7.4. Shepard Plot (pp72-73)

This program plots a Shepard plot for nonmetric scaling.

7.5. Spherical Plot (pp105-107)

This program produces a plot of a spherical configuration generated by the spherical nonmetric scaling program.

7.6. Two Dimensional Plot

This program plots a configuration of points.

8. Contents Of The Cdrom

Mds\biplot.exe	mds\clscal.exe	mds\dat2tran.exe
Mds\dat2unf.exe	mds\gipscal.exe	mds\history.exe
Mds\ind2con.exe	mds\indscal.exe	mds\least_sq.exe
Mds\linear.exe	mds\mat2diss.exe	mds\mdscal_2.exe
Mds\mdscal_3.exe	mds\mdscal_t.exe	mds\mds_inpu.exe
Mds\menu.exe	mds\menu_dat.exe	mds\menu_gol.ico
Mds\movie_md.exe	mds\nonlin.exe	mds\procrust.exe
Mds\rand_cat.exe	mds\ran_dats.exe	mds\ran_vecg.exe
Mds\recavdis.exe	mds\recipeig.exe	mds\shep_plo.exe
Mds\theta_pl.exe	mds\unfoldin.exe	mds\uni_scal.exe
Mds\vec2csv.exe	mds\vec2diss.exe	mds\vec2gowe.exe
Mds\vec_inpu.exe	mds\vec_join.exe	mds\vec_plot.exe
Mds\data\air_expe.asc	mds\data\air_expe.vec	mds\data\air_novi.asc
Mds\data\air_novi.vec	mds\data\birth.asc	mds\data\birth.ind
Mds\data\cancer.asc	mds\data\cancer.mat	mds\data\hans_70.asc
Mds\data\hans_70.vec	mds\data\hans_71.asc	mds\data\hans_71.vec
Mds\data\hans_72.asc	mds\data\hans_72.vec	mds\data\hans_73.asc
Mds\data\hans_73.vec	mds\data\kellog.asc	mds\data\kellog.vec
Mds\data\maidston.68	mds\data\maidston.asc	mds\data\monk_84.asc
Mds\data\monk_84.dis	mds\data\monk_85.asc	mds\data\monk_85.dis
Mds\data\munsinge.asc	mds\data\munsinge.mat	mds\data\nations.asc
Mds\data\nations.vec	mds\data\ord_surv.asc	mds\data\ord_surv.vec
Mds\data\pgeb.asc	mds\data\pgeb.vec	mds\data\pgwc.asc
Mds\data\pgwc.vec	mds\data\plato.asc	mds\data\plato.vec
Mds\data\poq.asc	mds\data\poq.vec	mds\data\scores.asc
Mds\data\scores.vec	mds\data\skulls.asc	mds\data\skulls.vec
Mds\data\speed.asc	mds\data\speed.vec	mds\data\tpo.asc
Mds\data\tpo.vec	mds\data\trd.asc	mds\data\trd.vec
Mds\data\uk_trave.asc	mds\data\uk_trave.dis	mds\data\whisky.asc
Mds\data\whisky.mat	mds\data\world_tr.asc	mds\data\world_tr.deg
Mds\data\world_tr.mat	mds\data\yoghurt.asc	mds\data\yoghurt.vec
Mds\notes\licence.doc	mds\notes\licence.ps	mds\notes\licence.txt
Mds\notes\readme.doc	mds\notes\readme.ps	mds\notes\readme.txt