Multidimensional Scaling

Usage

Multidimensional scaling (MDS) is an extremely popular approach for graphically representing relationships between objects (e.g. plots or samples) in multidimensional space. Dimension reduction via MDS is achieved by taking the original set of samples (e.g. species composition in quadrats) and calculating a dissimilarity (distance) measure for each pairwise comparison of samples. The samples are then usually represented graphically in 2-dimensions such that the distance between them approximates their multivariate dissimilarity as closely as possible. One of the most common applications of MDS is to examine the similarity of different ecological communities based on their species composition.

There are a variety of measures for calculating the dissimilarity amongst samples. The most basic of these is the Euclidean distance (i.e. simply the straight-line distance in multivariate space), however in ecology probably, the most commonly used distance measure is Bray-Curtis. The reason for its popularity is that, compared to other measures, Bray-Curtis is better at handling the large proportion of zeroes (e.g. species absences) commonly found in ecological datasets.

In addition to the choice of distance measure there are a number of different classes of MDS. Here we focus on what is referred to as Nonmetric MDS (NMDS), a non-parametric rank-based method that is comparatively robust to non-linear relationships between the calculated dissimilarity measure and the projected distance between objects.

Running the analysis

Consider a study where the aim is to understand levels of feeding specialisation by seven herbivorous amphipod species on five different host plants.

`metaMDS` in the `vegan` R package combines many of the decisions required to run an MDS analysis into a single function.

```
library(vegan)
herb.spec.mds = metaMDS(comm = community, distance = "bray", trace = FALSE)
```

`distance = bray` indicates you want to use a Bray-Curtis distance measure; `trace = FALSE` indicates you wish to ignore in progress output of results.

Calling the `plot` function on the x-y coordinates of the `metaMDS` output creates an MDS plot. To plot each habitat type (plant species) as a different colour, you need to have a vector identifying which habitat types correspond with the rows (samples) of the community matrix, e.g. `habitat`

```
##  [1] Colpomenia_peregrina       Colpomenia_peregrina
##  [3] Colpomenia_peregrina       Colpomenia_peregrina
##  [5] Colpomenia_peregrina       Colpomenia_peregrina
##  [7] Colpomenia_peregrina       Colpomenia_peregrina
##  [9] Colpomenia_peregrina       Colpomenia_peregrina
## [11] Colpomenia_peregrina       Colpomenia_peregrina
## [13] Colpomenia_peregrina       Colpomenia_peregrina
## [15] Colpomenia_peregrina       Colpomenia_peregrina
## [17] Colpomenia_peregrina       Colpomenia_peregrina
## [19] Colpomenia_peregrina       Colpomenia_peregrina
## [21] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
```

```
##  [23] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [25] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [27] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [29] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [31] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [33] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [35] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [37] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [39] Dictyopteris_acrostichoides Dictyopteris_acrostichoides
##  [41] Sargassum_linearifolium      Sargassum_linearifolium
##  [43] Sargassum_linearifolium      Sargassum_linearifolium
##  [45] Sargassum_linearifolium      Sargassum_linearifolium
##  [47] Sargassum_linearifolium      Sargassum_linearifolium
##  [49] Sargassum_linearifolium      Sargassum_linearifolium
##  [51] Sargassum_linearifolium      Sargassum_linearifolium
##  [53] Sargassum_linearifolium      Sargassum_linearifolium
##  [55] Sargassum_linearifolium      Sargassum_linearifolium
##  [57] Sargassum_linearifolium      Sargassum_linearifolium
##  [59] Sargassum_linearifolium      Sargassum_linearifolium
##  [61] Sargassum_vestitum           Sargassum_vestitum
##  [63] Sargassum_vestitum           Sargassum_vestitum
##  [65] Sargassum_vestitum           Sargassum_vestitum
##  [67] Sargassum_vestitum           Sargassum_vestitum
##  [69] Sargassum_vestitum           Sargassum_vestitum
##  [71] Sargassum_vestitum           Sargassum_vestitum
##  [73] Sargassum_vestitum           Sargassum_vestitum
##  [75] Sargassum_vestitum           Sargassum_vestitum
##  [77] Sargassum_vestitum           Sargassum_vestitum
##  [79] Sargassum_vestitum           Sargassum_vestitum
##  [81] Zonaria_diesingiana          Zonaria_diesingiana
##  [83] Zonaria_diesingiana          Zonaria_diesingiana
##  [85] Zonaria_diesingiana          Zonaria_diesingiana
##  [87] Zonaria_diesingiana          Zonaria_diesingiana
##  [89] Zonaria_diesingiana          Zonaria_diesingiana
##  [91] Zonaria_diesingiana          Zonaria_diesingiana
##  [93] Zonaria_diesingiana          Zonaria_diesingiana
##  [95] Zonaria_diesingiana          Zonaria_diesingiana
##  [97] Zonaria_diesingiana          Zonaria_diesingiana
##  [99] Zonaria_diesingiana          Zonaria_diesingiana
## 5 Levels: Colpomenia_peregrina ... Zonaria_diesingiana
```

```r
plot(herb.spec.mds$points, col = habitat, pch = 16)
# pch = 16 denotes the plotting symbol used

# add a legend
habitat.uni = unique(habitat)
legend(x = -1.9, y = 1.3, habitat.uni, pch=16, col = 1:5, cex = 0.8)
```
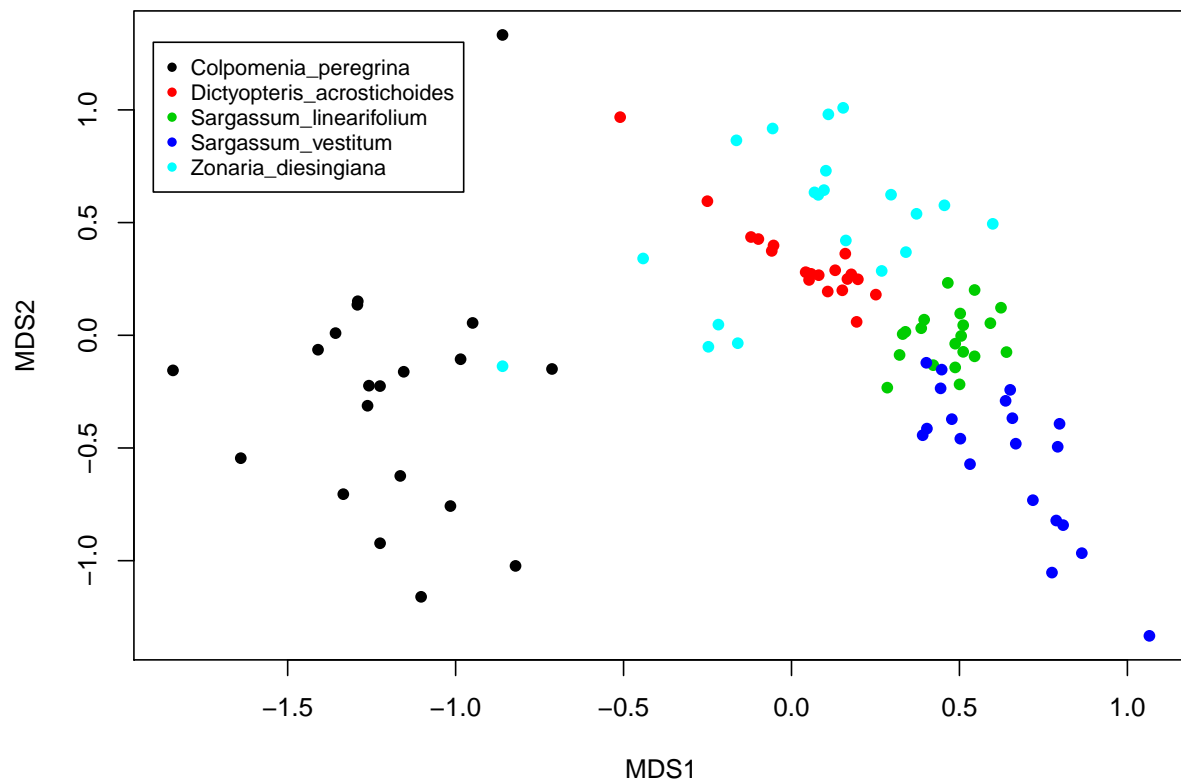
Interpreting the results

Interpreting an MDS plot is reasonably straightforward; objects that are closer together are more alike than those further apart.

Probably the most important factor to consider after generating an MDS plot is the 'stress'. The stress provides a measure of the degree to which the distance between samples in reduced dimensional space (usually 2-dimensions) corresponds with actual multivariate distance between the samples. Lower stress values indicate greater conformity and therefore are desirable. A rule of thumb is that stress values should ideally be less than 0.2 or even 0.1.

```
herb.spec.mds$stress
```

```
## [1] 0.1260515
```

If the stress value is greater than 0.2, it is advisable to include an additional dimension, but remember that human brains are not very well equipped to visualise objects in more than 3-dimensions.

Assumptions to check

As a rank-based approach NMDS, is comparatively robust to non-linearity between inter-object distance in space and their multivariate dissimilarity.

Communicating the results

Visual

MDS results are best presented visually as a 2- or 3-dimensional ordination plot (see above).

Writing

In a formal analysis, MDS plots are usually accompanied by some multivariate statistical test of dissimilarity between treatment/observational groups, e.g. via the `adonis` function in `vegan`.

Further help

Quinn and Keough (2002) *Experimental design and data analysis for biologists.* Cambridge University Press. Chapter 18 Multidimensional scaling and cluster analysis.

McKillup (2012) *Statistics explained. An introductory guide for life scientists.* Cambridge University Press.Chapter 22 Introductory concepts of multivariate analysis.

For a more in depth analysis of the data presented in this module, see Poore et al. 2000. Patterns of host use among alga- and sponge-associated amphipods. *Marine Ecology Progress Series*, 208: 183-196. http://www.int-res.com/abstracts/meps/v208/p183-196/