5.12 Canonical correlation: multiple X’s

In some cases, interest lies in finding axes through the cloud of points so as to maximise correlation with not just one X variable, but with linear combinations of multiple **X** variables simultaneously. In such cases, neither of these two sets of variables (i.e. the PCO’s arising from the resemblance matrix, on the one hand, and the **X** variables, on the other) have a specific role in the analysis – neither set is considered to be either predictors or responses. Rather, when there are several **X** variables, CAP can be conceptually described as sphericising *both* sets of variables, and then rotating them simultaneously against one another in order to find axes with maximum inter-correlations between these two sets.

As the agenda here is neither to explain nor predict one set using the other set, canonical correlation analysis with multiple **X** variables is a method for simply exploring relationships between two sets of variables. As such, its utility is perhaps rather limited for ecological applications, but certainly can be useful for generating hypotheses. CAP does canonical correlation between the PCO axes **Q**$_ m$ ($N \times m$) and **X** ($N \times q$), where m is generally chosen so as to minimise the leave-one-out residual sum of squares (see the section on Diagnostics), and the number of canonical axes generated will be min(m, q, (N – 1)). If $p < (N – 1)$ and the measure being used is Euclidean embeddable (e.g., see the section on Negative eigenvalues in chapter 3 herein and Gower & Legendre (1986) regarding the geometric properties of dissimilarity measures), then it makes sense to manually set $m = p$ in the CAP dialog *a priori*, as the dimensionality of PCO’s in such cases is known to be equal to p.

Note also that, as CAP will search for linear combinations of **X** variables that are maximally correlated with the PCO’s, it therefore makes sense to spend a little time examining the distributions of the **X** variables first (just as in dbRDA) to ensure that they have reasonably symmetric distributions and even scatter (using a draftsman plot for example), to transform them if necessary and also to consider eliminating redundant (very highly correlated) variables. The two sets of variables are treated symmetrically here, so CAP also *simultaneously* searches for linear combinations for the PCO’s that are maximally correlated with the **X** variables. Thus, one might also consider examining the distributions of the PCO’s (in scatterplot ordinations or a draftsman plot), to ensure that they, too, have fairly even scatter (although certainly no formal assumptions in this regard are brought to bear on the analysis).

Revision #5
Created 16 August 2022 15:08:14 by Arden
Updated 11 December 2024 02:59:23 by Abby Miller