5.15 Comparison of methods using SVD

The relationship between dbRDA and CAP can also be seen if we consider their formulation using singular value decomposition (SVD). For simplicity, but without loss of generality, suppose that each of the variables in matrices **Y** and **X** are centred on their means. A classical RDA is then obtained by the SVD of matrix $\mathbf{Y}^\prime\mathbf{X}^0$, namely:

$$ {\bf Y} ^\prime {\bf X} _ 0 = {\bf U} _R {\bf W} _R {\bf V} _R ^ \prime \tag{5.7} $$

where $\mathbf{W}_R$ = ${\bf \Gamma}$, a diagonal matrix of eigenvalues ($\gamma _1$, $\gamma _2$, …, $\gamma _s$)[114]. The RDA coordinate scores are then linear combinations of the fitted values, $\hat{\bf Y} = {\bf HY}$, where the coefficients are contained in the left-singular vectors $\mathbf{U}_R$. That is,

$$ {\bf Z} = \hat{\bf Y} {\bf U} _R \tag{5.8} $$

Now, a classical CCorA is obtained by SVD of matrix $\mathbf{Y}^{0 \prime}\mathbf{X}^0$, namely:

$$ {\bf Y}^{0 \prime} {\bf X} ^0 = {\bf U} _C {\bf W} _C {\bf V} _C ^ \prime \tag{5.9} $$

where $\mathbf{W}_C$ = ${\bf \Delta}$, a diagonal matrix of eigenvalues ($\delta _1$, $\delta _2$, …, $\delta _s$) that are the classical canonical correlations[115]. As CCorA is symmetric, we can plot coordinates either in the space of **X**, or in the space of **Y**. Canonical coordinate scores in the space of X are:

$$ {\bf B} = {\bf X} ^0 {\bf V} _C {\bf \Delta} \tag{5.10} $$

while canonical coordinate scores in the space of Y are:

$$ {\bf C} = {\bf Y} ^0 {\bf U} _C {\bf \Delta} \tag{5.11} $$

The correlation between the variables B_1 and C_1 will be equal to $\delta _1$, between B_2 and C_2 will be $\delta _2$, and so on. Note also that the scaling of each of these new sets of variables by their corresponding eigenvalues (i.e., multiplying by ${\bf \Delta}$ as shown in equations 5.10 and 5.11) is optional.

Next, we can generalise the above to any resemblance measure by replacing **Y** with **Q**, thus dbRDA is obtained by an SVD as follows:

$$ {\bf Q} ^ \prime {\bf X} ^ 0 = {\bf U} _R {\bf \Gamma V} _R ^\prime \tag{5.12} $$

with fitted values **HQ**, dbRDA coordinate scores **Z** = $\mathbf{HQU}_R$ and with an appropriate vector overlay for the plot (eigenvector coefficients for normalised **X** variables) being contained in $\mathbf{V}_R$.

Similarly, CAP is obtained by an SVD of:

$$ {\bf Q} ^ {0 \prime} _m {\bf X} ^ 0 = {\bf U} _C {\bf \Delta V} _C ^\prime \tag{5.13} $$

with canonical coordinate scores $\mathbf{Q}^{0 \prime}_m\mathbf{U}_C {\bf \Delta}$ and an appropriate vector overlay for the plot (eigenvector coefficients for normalised **X** variables) being contained in $\mathbf{V}_C$.

The purpose of this section is not to throw some scary matrix algebra around and generate fear! Rather, it is intended to further highlight the conceptual differences and similarities between these two approaches (as outlined in Table 5.1) and also to provide some algebraic conventions for formulating and discussing these methods which (hopefully) complements existing literature describing the classical (Euclidean-based) versions of them.

---

[114] Note: these are the square root of the eigenvalues that would be obtained from running the dbRDA routine on the data using Euclidean distances.

[115] Note: these are the square root of the eigenvalues that would be obtained from running the CAP routine on the data using Euclidean distances.

Revision #11
Created 17 August 2022 11:02:32 by Arden
Updated 17 August 2022 11:46:58 by Arden