Slides for 2023 honours multivariate canonical correlation analysis lecture

Effect of standardisation on canonical correlation analysis

- Let $Z_k^{(i)} = \frac{X_k^{(1)} - \mu_k^{(1)}}{\sqrt{\sigma}_{kk}}$ with $\mathbf{Z}^{(i)}$ the associated vector of standardised variables.

Interpreting canonical variables using correlations

- To interpret the canonical variables, we can calculate correlations between the canonical variables and the original variables (as we did with PCA).
- Let $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_p\end{bmatrix}'$ be the matrix with canonical coefficient vectors along the rows.
- Then let $\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)}$ be the vector of standardised coefficients.
- We are then interested in the matrix of correlations between $\mathbf{U}$ (the the canonical variables) and $\mathbf{X}^{(1)}$ (the original variables).

$$
\begin{aligned}
{}_{\mathbf{U},\mathbf{X}^{(1)}} &= \mathrm{Cor}[\mathbf{U}, \mathbf{X}^{(1)}], \\
&= \mathrm{Cov}[\mathbf{A}\mathbf{U}, \mathbf{V}^{-\frac{1}{2}}\mathbf{X}] \text{ for } \mathbf{V} = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)
\end{aligned}
$$

- We now consider the canonical correlation matrix $\mathbf{R} = \mathbf{U}'\mathbf{Z}^{(2)}$.