Topic 1: Canonical correlations

Welcome to Week 3

Dr Pavel Krivitsky gives you a brief overview of topics and concepts we'll be covering in this week.

Transcript

Weekly learning outcomes

- Provide the mathematical definition of a canonical correlation.
- Evaluate and interpret the canonical correlations between two datasets.
- Test for the significance of a canonical correlation.
- Determine the optimal number of canonical correlations.
- Perform the regular linear model inference on multivariate linear models.
- Explain the assumptions underlying the above inferential procedures and check them.

Topics we will cover are:

- Topic 1: Canonical correlations
- Topic 2: Linear models with multivariate response

- Topic 3: Multivariate ANOVA: testing canonical correlations and multivariate linear models

Optional reading

Härdle W.K., Simar L. (2012) Regression Models. In: Applied Multivariate Statistical Analysis. Springer, Berlin, Heidelberg

- Section 8.1.2

All readings are available from the course Leganto reading list. Please keep in mind that you will need to be logged into Moodle to access the Leganto reading list.

Questions about this week's topics?

This week's topics were prepared by Dr P. Krivitsky. If you have any questions or comments, please post them under Discussion or email directly: p.krivitsky@unsw.edu.au

Canonical correlation analysis

Introduction

Assume we are interested in the association between two **sets** of random variables. Typical examples include: relation between a set of governmental policy variables and a set of economic goal variables; relation between college "performance" variables (like grades in courses in five different subject matter areas) and pre-college "achievement" variables (like high-school grade-point averages for junior and senior years, number of high-school extracurricular activities) etc.

The way the above problem of measuring association is solved in Canonical Correlation Analysis, is to consider the largest possible correlation between a *linear combination of the variables in the first set* and a *linear combination of the variables in the second set*. The pair of linear combinations obtained through this maximisation process is called **first canonical variables** and their correlation is called **first canonical correlation**. The process can be continued (similarly to the principal components procedure) to find a second pair of linear combinations having the largest correlation among all pairs that are uncorrelated with the initially selected pair. This would give us the second set of canonical variables with their second canonical correlation etc. The maximisation process that we are performing at each step reflects our wish (again like in principal components analysis) to concentrate the initially high dimensional relationship between the 2 sets of variables into a few pairs of canonical variables only. Often, even only **one** pair is considered. The rationale in canonical correlation analysis is that when the number of variables is large, interpreting the **whole set** of correlation coefficients between pairs of variables from each set is hopeless and in that case, one should concentrate on a *few* carefully chosen representative correlations. Finally, we should note that the traditional (simple) correlation coefficient and the multiple correlation coefficient (Topic 2 of Week 2) are *special cases* of canonical correlation in which one or both sets contain a single variable.

Application in testing for independence of sets of variables

Besides being interesting in its own right (See Introduction section above), calculating canonical correlations turns out to be important for the sake of **testing independence of sets of random variables**. Let us remember that testing for independence and for uncorrelatedness in the case of multivariate normal are equivalent problems. Assume now that that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. Furthermore, let $\boldsymbol{X}$ be partitioned into r, q components ($r + q = p$) with $\boldsymbol{X}^{(1)} \in \mathbb{R}^r, \boldsymbol{X}^{(2)} \in \mathbb{R}^q$ and correspondingly, the covariance matrix

$$\Sigma = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \in \mathcal{M}_{p,p}$$

has been also partitioned into $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, accordingly. We shall assume for simplicity that the matrices Σ, Σ_{11}, and Σ_{22} are nonsingular. To test $H_0 : \Sigma_{12} = \mathbf{0}$ against a general alternative, a sensible way to go would be the following: for fixed vectors $\boldsymbol{a} \in \mathbb{R}^r, \boldsymbol{b} \in \mathbb{R}^q$ let $Z_1 = \boldsymbol{a}^\top \boldsymbol{X}^{(1)}$ and $Z_2 = \boldsymbol{b}^\top \boldsymbol{X}^{(2)}$ giving $\rho_{\boldsymbol{a},\boldsymbol{b}} = \mathrm{Cor}(Z_1, Z_2) = \frac{\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b}}{\sqrt{\boldsymbol{a}^\top \Sigma_{11} \boldsymbol{a} \boldsymbol{b}^\top \Sigma_{22} \boldsymbol{b}}}$. H_0 is equivalent to H_0 : $\rho_{\boldsymbol{a},\boldsymbol{b}} = 0$ for all $\boldsymbol{a} \in \mathbb{R}^r, \boldsymbol{b} \in \mathbb{R}^q$. For a particular pair $\boldsymbol{a}, \boldsymbol{b}$, H_0 would be accepted if $|r_{\boldsymbol{a},\boldsymbol{b}}| = \frac{|\boldsymbol{a}^\top \boldsymbol{S}_{12} \boldsymbol{b}|}{\sqrt{\boldsymbol{a}^\top \boldsymbol{S}_{11} \boldsymbol{a} \boldsymbol{b}^\top \boldsymbol{S}_{22} \boldsymbol{b}}} \leq k$ for certain positive constant k. (Here $\boldsymbol{S}_{ij}$ are the corresponding data based estimators of Σ_{ij}.) Hence an appropriate acceptance region for H_0 would be given in the form $\left\{ \boldsymbol{X} \in \mathcal{M}_{p,n} : \max_{\boldsymbol{a},\boldsymbol{b}} r^2_{\boldsymbol{a},\boldsymbol{b}} \leq k^2 \right\}$. But maximising $r^2_{\boldsymbol{a},\boldsymbol{b}}$ means to find the maximum of $\left(a^\top S_{12} b\right)^2$ under constraints $\boldsymbol{a}^\top \boldsymbol{S}_{11} \boldsymbol{a} = 1 \text{ and } \boldsymbol{b}^\top \boldsymbol{S}_{22} \boldsymbol{b} = 1$, and this is exactly the data-based version of the optimisation problem to be solved in the above introduction.

For the goals in the above to be achieved, we need to solve the type of problems presented in the next slide.

Precise mathematical formulation and solution to the problem

Canonical variables are the variables $Z_1 = \boldsymbol{a}^\top \boldsymbol{X}^{(1)}$ and $Z_2 = \boldsymbol{b}^\top \boldsymbol{X}^{(2)}$ where $\boldsymbol{a} \in \mathbb{R}^r, \boldsymbol{b} \in \mathbb{R}^q$ are obtained by maximising $(\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b})^2$ under the constraints $\boldsymbol{a}^\top \Sigma_{11} \boldsymbol{a} = \boldsymbol{b}^\top \Sigma_{22} \boldsymbol{b} = 1$.

To do so, we use the method of Lagrange Multipliers, and the derivation is not dissimilar to the one in PCA, albeit more complex. It is given below, but is **not** assessable. The key results are that:

1. The first canonical correlation is the square root of the largest eigenvalue of $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ or of $\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1}$. (They turn out to be the same.)
2. $\boldsymbol{b} = \Sigma_{22}^{-\frac{1}{2}} \hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{b}}$ is the eigenvector of $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ corresponding to the largest eigenvalue.
3. $\boldsymbol{a} = \frac{1}{\mu} \Sigma_{11}^{-1} \Sigma_{12} \boldsymbol{b}$, where μ is the square root of the largest eigenvalue of $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$.
4. The second canonical correlation (i.e., relationship between the two sets of variables that remains after accounting for the first) is the square root of the second largest eigenvalue of the above matrices and the corresponding variates are based on its eigenvectors, etc..

Detailed derivation (optional: not assessed)

To solve the above maximisation problem, we construct

$$\operatorname{Lag}(\boldsymbol{a}, \boldsymbol{b}, \lambda_1, \lambda_2) = \left(\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b}\right)^2 + \lambda_1 \left(\boldsymbol{a}^\top \Sigma_{11} \boldsymbol{a} - 1\right) + \lambda_2 \left(\boldsymbol{b}^\top \Sigma_{22} \boldsymbol{b} - 1\right)$$

Partial differentiation with respect to the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ gives:

$$2\left(\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b}\right) \Sigma_{12} \boldsymbol{b} + 2\lambda_1 \Sigma_{11} \boldsymbol{a} = \boldsymbol{0} \in \mathbb{R}^r \tag{3.1}$$

$$2\left(\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b}\right) \Sigma_{21} \boldsymbol{a} + 2\lambda_2 \Sigma_{22} \boldsymbol{b} = \boldsymbol{0} \in \mathbb{R}^q \tag{3.2}$$

We multiply (3.1) by the vector $\boldsymbol{a}^\top$ from left and equation (3.2) by $\boldsymbol{b}^\top$ from left and after subtracting the two equations obtained we get $\lambda_1 = \lambda_2 = -\left(\boldsymbol{a}^\top \Sigma_{12} \boldsymbol{b}\right)^2 = -\mu^2$. Hence:

$$\Sigma_{12} \boldsymbol{b} = \mu \Sigma_{11} \boldsymbol{a} \tag{3.3}$$

and

$$\Sigma_{21} \boldsymbol{a} = \mu \Sigma_{22} \boldsymbol{b} \tag{3.4}$$

Now we first multiply (3.3) by $\Sigma_{21} \Sigma_{11}^{-1}$ from left, then both sides of (3.4) by the scalar μ and after finally adding the two equations we get:

$$\left(\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \mu^2 \Sigma_{22}\right) \boldsymbol{b} = \boldsymbol{0} \tag{3.5}$$

The homogeneous equation system (3.5) having a non-trivial solution w.r.t. $\boldsymbol{b}$ means that

$$\left|\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\mu^2\Sigma_{22}\right|=0 \tag{3.6}$$

must hold. Then, of course,

$$\left|\Sigma_{22}^{-\frac{1}{2}}\right|\left|\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\mu^2\Sigma_{22}\right|\left|\Sigma_{22}^{-\frac{1}{2}}\right|=\left|\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}-\mu^2 I_q\right|=0$$

must hold. This means that μ^2 has to be an eigenvalue of the matrix $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$. Also, $\boldsymbol{b}=\Sigma_{22}^{-\frac{1}{2}}\hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{b}}$ is the eigenvector of $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ corresponding to this eigenvalue.

(Note, however, that this representation is good mainly for theoretical purposes, the main advantage being that one is dealing with eigenvalues of a symmetric matrix. If doing calculations by hand, it is usually easier to calculate $\boldsymbol{b}$ directly as the solution of the linear equation (3.5), i.e., find the largest eigenvalue of the (non-symmetric) matrix $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and then find the eigenvector $\boldsymbol{b}$ that corresponds to it. Besides, we also see from the definition of μ that $\mu^2=\left(\boldsymbol{a}^\top\Sigma_{12}\boldsymbol{b}\right)^2$ holds.)

Since we wanted to **maximise** the right hand side, it is obvious that μ^2 must be chosen to be the **largest eigenvalue** of the matrix $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ (or, which is the same thing, the largest eigenvalue of the matrix $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}$). Finally, we can obtain the vector $\boldsymbol{a}$ from (3.3): $\boldsymbol{a}=\frac{1}{\mu}\Sigma_{11}^{-1}\Sigma_{12}\boldsymbol{b}$. That way, the **first** canonical variables $Z_1=\boldsymbol{a}^\top\boldsymbol{X}^{(1)}$ and $Z_2=\boldsymbol{b}^\top\boldsymbol{X}^{(2)}$ are determined and the value of the first canonical correlation is just μ. The orientation of the vector $\boldsymbol{b}$ is chosen such that the sign of μ should be positive.

Now, it is easy to see that if we want to extract a second pair of canonical variables we need to repeat the same process by starting with the **second largest** eigenvalue μ^2 of the matrix $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ (or of the matrix $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$). This will automatically ensure that the second pair of canonical variables is uncorrelated with the first pair. The process can theoretically be continued until the number of pairs of canonical variables equals the number of variables in the smaller group. But in practice, much fewer canonical variables will be needed. Each canonical variable is uncorrelated with all the other canonical variables of either set except for the one corresponding canonical variable in the opposite set.

Note. It is important to point out that already by definition the canonical correlation is at least as large as the multiple correlation between any variable and the opposite set of variables. It is in fact possible for the first canonical correlation to be *very large* while all the multiple correlations of each separate variable with the opposite set of canonical variables are small. This once again underlines the importance of Canonical Correlation analysis.

Estimating and testing canonical correlations

The way to estimate the canonical variables and canonical correlation coefficients is based on the plug-in technique: one follows the steps outlined in the previous section, by each time substituting $\boldsymbol{S}_{ij}$ in place of Σ_{ij}:

1. From the sample variance--covariance matrix $\boldsymbol{S}$, calculate the $\boldsymbol{A} = \boldsymbol{S}_{22}^{-\frac{1}{2}}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-\frac{1}{2}}$.
2. Calculate μ^2 as the greatest eigenvalue of $\boldsymbol{A}$ and $\hat{\boldsymbol{b}}$ as the corresponding eigenvector.
3. Calculate $\boldsymbol{b} = \boldsymbol{S}_{22}^{-\frac{1}{2}}\hat{\boldsymbol{b}}$ and $\boldsymbol{a} = \frac{1}{\mu}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{b}$.

Then, $\mu = \sqrt{\mu^2}$ is the first canonical correlation, and $\boldsymbol{a}$ and $\boldsymbol{b}$ represent the linear combinations of the data vectors that produce it.

Let us now discuss the independence testing issue outlined earlier. The acceptance region of the independence test of H_0 would be $\{\boldsymbol{X} \in \mathcal{M}_{p,n} : \text{ largest eigenvalue of } A \leq k_\alpha\}$ where k_α is some critical value. This value depends on three parameters: $s = \min(r, q)$, $m = \frac{|r-q|-1}{2}$, and $N = \frac{n-r-q-2}{2}$, n being the sample size. One can also use good F-distribution-based approximations for a (transformations of) this distribution like Wilk's lambda, Pillai's trace, Hotelling trace, and Roy's greatest root. *Permutation tests* that are not as sensitive to these assumptions are also possible. We discuss this in more detail in the demonstration below and Topic 3.

Some important computational issues

Note that calculating $X^{-\frac{1}{2}}$ and $X^{\frac{1}{2}}$ for a symmetric positive definite matrix X according to the theoretically attractive spectral decomposition method may be numerically unstable. This is especially the case when some of the eigenvalues are close to zero (or, more precisely, when the ratio of the greatest eigenvalue and the least eigenvalue —the *condition number*— is high).

We can use the **Cholesky decomposition** described in slide "Numerical stability and Cholesky decomposition". Looking back at (3.5), we see that if $U^\top U = \Sigma_{22}^{-1}$ gives the Cholesky decomposition of the matrix Σ_{22}^{-1} then μ^2 is an eigenvalue of the matrix $A = U\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}U^\top$. Indeed, by multiplying from left by U and from right by $U^\top$ in (3.6) we get:

$$\left|A - \mu^2 U\Sigma_{22}U^\top\right| = 0$$

But $U\Sigma_{22}U^\top = U\left(U^\top U\right)^{-1}U^\top = UU^{-1}\left(U^\top\right)^{-1}U^\top = I$ holds.

Check your understanding

Complete the below exercises to check your understanding of concepts presented so far.

1. Let the components of X correspond to scores on tests in arithmetic speed (X_1), arithmetic power (X_2), memory for words (X_3), memory for meaningful symbols (X_4), and memory for meaningless symbols (X_5). The observed correlations in a sample of 140 are

$$\begin{bmatrix} 1.0000 & 0.4248 & 0.0420 & 0.0215 & 0.0573 \\ & 1.0000 & 0.1487 & 0.2489 & 0.2843 \\ & & 1.0000 & 0.6693 & 0.4662 \\ & & & 1.0000 & 0.6915 \\ & & & & 1.0000 \end{bmatrix}$$

Find the canonical correlations and canonical variates between the first two variates and the last three variates. Comment. Write R code to implement the required calculations.

2. Students sit 5 different papers, two of which are closed book and the rest open book. For the 88 students who sat these exams the sample covariance matrix is

$$S = \begin{bmatrix} 302.3 & 125.8 & 100.4 & 105.1 & 116.1 \\ & 170.9 & 84.2 & 93.6 & 97.9 \\ & & 111.6 & 110.8 & 120.5 \\ & & & 217.9 & 153.8 \\ & & & & 294.4 \end{bmatrix}$$

Find the canonical correlations and canonical variates between the first two variates (closed book exams) and the last three variates (open book exams). Comment.

3. Consider a random vector $\boldsymbol{X} \sim N_4(\boldsymbol{\mu}, \Sigma)$ with $\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2\rho & \rho & \rho \\ 2\rho & 1 & \rho & \rho \\ \rho & \rho & 1 & 2\rho \\ \rho & \rho & 2\rho & 1 \end{pmatrix}$

where ρ is a small enough positive constant.

a) Find the two canonical correlations between $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $\begin{pmatrix} X_3 \\ X_4 \end{pmatrix}$

Comment.

b) Find the first pair of canonical variables.

4. Consider the following covariance matrix Σ of a four dimensional normal vector:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \left(\begin{array}{cc|cc} 100 & 0 & 0 & 0 \\ 0 & 1 & 0.95 & 0 \\ \hline 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{array}\right)$$. Verify that the first pair of canonical variates are just the second and the third component of the vector and the canonical correlation equals 0.95.

Demonstration: Canonical correlation examples

Watch the following demonstration by Dr Krivitsky before proceeding to the task.

Transcript

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete this task select the 'Cancor_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

Cancor_Examples.demo.Rmd

The output of the RMD file is also displayed below:

Challenge: Canonical correlation examples

If you choose to complete this task in your own RStudio, upload the following file:

Cancor_Examples.challenge.Rmd

Click on the 'Cancor_Examples.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The solution will also be available here on Friday of Week 1 by clicking on the 'Solution' tab.

The output of the RMD file is also displayed below: