

Canonical Correlation Analysis

Canonical Correlation analysis (CCA) aims to quantify the associations between two sets of variables, each set consists of two or more variables. The association is measured by correlation, thus is on linear relationships of the two groups of variables.

Examples

- Do last year's labor data relate or explain this year's labor situation?
- Are the math science performances of the students related to their reading performances?
- Are there associations between government policies and economic variables?
- For a decathlon athlete, is his track event performance predictive of his performance in field events?

CCA data and objectives

The data for canonical correlation analysis are of n observations, each observation is a vector with $p+q$ components. The vector components form two groups of interest.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} & y_{11} & y_{12} & \cdots & y_{1q} \\ x_{21} & x_{22} & \cdots & x_{2p} & y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jp} & y_{j1} & y_{j2} & \cdots & y_{jq} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} & y_{n1} & y_{n2} & \cdots & y_{nq} \end{bmatrix} \begin{matrix} \leftarrow \text{1st observation} \\ \leftarrow \text{2nd observation} \\ \vdots \\ \leftarrow j\text{th observation} \\ \vdots \\ \leftarrow n\text{th observation} \end{matrix}$$

$\underbrace{\hspace{10em}}_{p \text{ variables}} \quad \underbrace{\hspace{10em}}_{q \text{ variables}}$

In the data matrix, each row can be viewed as the measurement of a sample point in \mathbb{R}^{p+q} from a random vector

$$[X_1 \cdots X_p \ Y_1 \cdots Y_q]$$

We can think of the data as two data matrices of dimensions $n \times p$ and $n \times q$, observed together.

The objective of canonical correlation analysis is to quantify the association between two variable groups $\{X_1, \dots, X_p\}$ and $\{Y_1, \dots, Y_q\}$, in terms of linear correlation.

1 Population canonical variates and canonical correlations

Notations

In theoretical development of the method, we will consider the corresponding population random vectors $\begin{bmatrix} X \\ Y \end{bmatrix}$, where

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = [X_1 \ X_2 \ \cdots \ X_p]', \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix} = [Y_1 \ Y_2 \ \cdots \ Y_q]'$$

Remarks on notations

Another common notation for $\begin{bmatrix} X \\ Y \end{bmatrix}$ is $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, as in Johnson and Wichern, using double subscripts indexing component variables,

$$\mathbf{X}_1 = (X_{11}, \dots, X_{1p})', \text{ } p\text{-variate random vector,}$$

$$\mathbf{X}_2 = (X_{21}, \dots, \dots, X_{2q})', \text{ } q\text{-variate random vector.}$$

The formulation of CCA

We investigate the canonical correlation structure on the two population vectors $\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_q \end{bmatrix}$.

Goal of CCA:

The first and foremost objective for CCA is to find a pair of variables, $\mathbf{a}'\mathbf{X}$ and $\mathbf{b}'\mathbf{Y}$, each a linear combination of the original variables in each group,

$$(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}) = (a_1X_1 + \cdots + a_pX_p, \ b_1Y_1 + \cdots + b_qY_q),$$

so that the correlation of the new pair of variables represent the maximum correlation between the two groups of original variables. (Notes: Here a_i, b_j simply denote the scalar components of arbitrary vectors \mathbf{a} and \mathbf{b} .)

In other words, find

$$\operatorname{argmax}_{\mathbf{a}, \mathbf{b}} \operatorname{Corr}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}) \quad \text{among all } \mathbf{a} \in \mathbb{R}^p, \ \mathbf{b} \in \mathbb{R}^q.$$

After finding such a pair, the derivation process may continue to find the next maximum-correlated pair of linear combinations among all pairs uncorrelated with the first selected pair, and so on.

Conventions

These successively chosen pairs of linear combinations are called **canonical variables** or **canonical variates**, the correlations between canonical variate pairs are termed **canonical correlations**.

Without loss of generality, $p \leq q$ is often used as a convenient assumption on the sizes of the two sets of variables.

Treating the two sets of random variables jointly as

$$\begin{bmatrix} X \\ Y \end{bmatrix} = [\mathbf{X}' \ \mathbf{Y}']' = [X_1 \ \cdots \ X_p \ Y_1 \ \cdots \ Y_q]'$$

Their expectation vector and covariance matrix are formulated as

$$\mathbb{E} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}_{(p+q) \times 1}, \quad \operatorname{Cov} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}_{(p+q) \times (p+q)}$$

Note that Σ_{12} is a p -by- q matrix, Σ_{21} is a q -by- p matrix, $\Sigma'_{21} = \Sigma_{12}$.

Population canonical variates

Consider a pair of linear combinations of the variables:

$$U = \mathbf{a}'\mathbf{X}, \quad V = \mathbf{b}'\mathbf{Y}, \quad \mathbf{a} \in \mathbb{R}^p, \quad \mathbf{b} \in \mathbb{R}^q.$$

Then U and V are univariate random variables, with

$$\begin{aligned} \text{cov}(U, V) &= \mathbf{a}' \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b} = \mathbf{a}' \Sigma_{12} \mathbf{b} \\ \text{var}(U) &= \mathbf{a}' \text{Cov}(\mathbf{X}) \mathbf{a} = \mathbf{a}' \Sigma_{11} \mathbf{a} \\ \text{var}(V) &= \mathbf{b}' \text{Cov}(\mathbf{Y}) \mathbf{b} = \mathbf{b}' \Sigma_{22} \mathbf{b} \end{aligned}$$

Canonical correlation analysis seeks \mathbf{a} and \mathbf{b} to maximize the correlation

$$\text{corr}(U, V) = \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)} \sqrt{\text{var}(V)}} = \frac{\mathbf{a}' \Sigma_{12} \mathbf{b}}{\sqrt{\mathbf{a}' \Sigma_{11} \mathbf{a}} \sqrt{\mathbf{b}' \Sigma_{22} \mathbf{b}}} \quad (1)$$

Note that replacing \mathbf{a} and \mathbf{b} by their constant multiples will not change the correlation (1). Hence for convenience of this derivation, we normalize \mathbf{a} and \mathbf{b} by the constraints

$$\text{var}(U) = \mathbf{a}' \Sigma_{11} \mathbf{a} = 1, \quad \text{var}(V) = \mathbf{b}' \Sigma_{22} \mathbf{b} = 1.$$

Maximizing correlation in (1) can be described as

$$\text{Maximize } \rho = \rho_{a,b} = \mathbf{a}' \Sigma_{12} \mathbf{b} \quad \text{under constraints} \quad \mathbf{a}' \Sigma_{11} \mathbf{a} = 1, \quad \mathbf{b}' \Sigma_{22} \mathbf{b} = 1.$$

Just like the way we derived principal components, Lagrange Multiplier method can be used to find the solutions.

In the following we will go through the procedure to find canonical variates \mathbf{a} and \mathbf{b} which achieve the maximization of canonical correlation in (1). The steps utilize the Lagrange Multiplier method, properties of symmetric and positive semidefinite matrices, and the singular value decomposition of matrices.

Proof. Derivation of canonical variates (The leading pair)

The Lagrange multiplier setup

The optimization problem of CCA can be formulated by the method of Lagrange Multipliers by setting the Lagrangian function

$$L = \mathbf{a}' \Sigma_{12} \mathbf{b} - \frac{1}{2} \lambda (\mathbf{a}' \Sigma_{11} \mathbf{a} - 1) - \frac{1}{2} \gamma (\mathbf{b}' \Sigma_{22} \mathbf{b} - 1).$$

Optimal \mathbf{a} and \mathbf{b} that maximize $\rho = \mathbf{a}' \Sigma_{12} \mathbf{b}$ must occur at a critical point at which $\frac{\partial L}{\partial \mathbf{a}} = \mathbf{0}_p$, $\frac{\partial L}{\partial \mathbf{b}} = \mathbf{0}_q$.

Differentiating L and equating the derivatives to zero vectors,

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}} &= \Sigma_{12} \mathbf{b} - \lambda \Sigma_{11} \mathbf{a} = \mathbf{0}_p \\ \frac{\partial L}{\partial \mathbf{b}} &= \Sigma_{21} \mathbf{a} - \gamma \Sigma_{22} \mathbf{b} = \mathbf{0}_q \end{aligned} \quad (2)$$

In the second equation, we used $\Sigma_{21} = \Sigma'_{12}$.

The values of Lagrange multipliers

To find out the values of λ and γ satisfying (2), left multiply \mathbf{a}' and \mathbf{b}' to the two equations in (2) respectively. Notice that, the scalar

$$\rho = \mathbf{a}' \Sigma_{12} \mathbf{b} = \mathbf{b}' \Sigma_{21} \mathbf{a},$$

we obtain

$$\begin{aligned} \mathbf{a}' \Sigma_{12} \mathbf{b} - \lambda \mathbf{a}' \Sigma_{11} \mathbf{a} &= \rho - \lambda = 0 & \Rightarrow & \lambda = \rho & (\text{since } \mathbf{a}' \Sigma_{11} \mathbf{a} = 1) \\ \mathbf{b}' \Sigma_{21} \mathbf{a} - \gamma \mathbf{b}' \Sigma_{22} \mathbf{b} &= \rho - \gamma = 0 & \Rightarrow & \gamma = \rho & (\text{since } \mathbf{b}' \Sigma_{22} \mathbf{b} = 1) \end{aligned}$$

So (2) is simplified as

$$\begin{aligned} \Sigma_{12} \mathbf{b} - \rho \Sigma_{11} \mathbf{a} &= \mathbf{0}_p \\ \Sigma_{21} \mathbf{a} - \rho \Sigma_{22} \mathbf{b} &= \mathbf{0}_q \end{aligned} \quad (3)$$

Next we need to solve for \mathbf{a} and \mathbf{b} . Note that the two equations above are in different dimensions.

Vectors \mathbf{a} and \mathbf{b} as eigenvectors of matrices

First, we solve for \mathbf{a} in (3). To cancel out the \mathbf{b} terms (then simplify the \mathbf{a} term in the first equation), left multiplying the first equation in (3) by $\rho \Sigma_{11}^{-1}$, then left multiply the second equation by $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1}$,

$$\begin{aligned} \rho \Sigma_{11}^{-1} \Sigma_{12} \mathbf{b} - \rho^2 \mathbf{a} &= \mathbf{0}_p \\ \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{a} - \rho \Sigma_{11}^{-1} \Sigma_{12} \mathbf{b} &= \mathbf{0}_p \end{aligned}$$

Now the two equations are of the same dimensions, we can add the two equations to cancel out the \mathbf{b} terms.

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{a} - \rho^2 \mathbf{a} = \mathbf{0}_p.$$

Write the equation as

$$(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \rho^2 \mathbf{I}_p) \mathbf{a} = \mathbf{0}_p$$

If there are ρ^2 and $\mathbf{a} \neq \mathbf{0}_p$ satisfying the equation, then ρ^2 is an eigenvalue of $p \times p$ matrix $\mathbf{A} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$, and \mathbf{a} is an eigenvector of \mathbf{A} with eigenvalue ρ^2 . Analogous derivation will yield

$$(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 \mathbf{I}_q) \mathbf{b} = \mathbf{0}_q$$

Then \mathbf{b} is an eigenvector of the $q \times q$ matrix $\mathbf{B} = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ with eigenvalue ρ^2 , if there are ρ^2 and $\mathbf{b} \neq \mathbf{0}_p$ satisfying the equation.

To obtain exact expressions for our objective vectors \mathbf{a}, \mathbf{b} , we need to obtain properties of eigenvalues and eigenvectors of matrices \mathbf{A} and \mathbf{B} .

Eigenvalues of matrices \mathbf{A} and \mathbf{B}

We have obtained

$$\begin{aligned} (\mathbf{A} - \rho^2 \mathbf{I}_p) \mathbf{a} &= \mathbf{0}_p \\ (\mathbf{B} - \rho^2 \mathbf{I}_q) \mathbf{b} &= \mathbf{0}_q \end{aligned}$$

What are possible eigenvalues of the matrices \mathbf{A} and \mathbf{B} , and what are their relationships?

Not that matrices \mathbf{A} and \mathbf{B} are not necessarily symmetric. There existence of real eigenvalues and their ranges are not readily established.

In the above, the eigenvalue is of the form ρ^2 , hinting the non-negativeness of the eigenvalues, which is a property of positive semidefinite matrices. To relate \mathbf{A} and \mathbf{B} with symmetric, positive semidefinite matrices, first recall that any symmetric positive (semi)definite matrix can be written as the square of a symmetric positive (semi)definite matrix, which can be denoted as the "square root" matrix of the original one. Assume that Σ_{11}, Σ_{22} are positive definite, consequently so are their inverse matrices. We may write

$$\Sigma_{11} = \Sigma_{11}^{1/2} \Sigma_{11}^{1/2}, \quad \Sigma_{22} = \Sigma_{22}^{1/2} \Sigma_{22}^{1/2}, \quad \Sigma_{11}^{-1} = \Sigma_{11}^{-1/2} \Sigma_{11}^{-1/2}, \quad \Sigma_{22}^{-1} = \Sigma_{22}^{-1/2} \Sigma_{22}^{-1/2}$$

where $\Sigma_{ii}^{1/2}$ is the symbol for a symmetric positive definite that its square is Σ_{ii} , etc. Note that such matrices decompositions are not necessarily unique. Now rewrite

$$\mathbf{A} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Sigma_{11}^{-1/2} (\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}) (\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1/2}) \Sigma_{11}^{1/2} = \Sigma_{11}^{-1/2} (\mathbf{C} \mathbf{C}') \Sigma_{11}^{1/2} \quad (4)$$

with

$$\mathbf{C} = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$$

a p -by- q matrix. Similarly

$$\mathbf{B} = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \Sigma_{22}^{-1/2} (\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1/2}) (\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}) \Sigma_{22}^{1/2} = \Sigma_{22}^{-1/2} (\mathbf{C}' \mathbf{C}) \Sigma_{22}^{1/2} \quad (5)$$

Recall from matrix algebra that if a matrix $M_1 = Q^{-1} M_2 Q$ for some invertible matrix Q , then matrices M_1 and M_2 are "similar", often denoted as $M_1 \sim M_2$. Similar matrices share the same eigenvalues.

Expression (4) shows that $\mathbf{A} \sim \mathbf{C} \mathbf{C}'$, thus \mathbf{A} has the same eigenvalues as $\mathbf{C} \mathbf{C}'$.

Similarly, (5) implies that $\mathbf{B} \sim \mathbf{C}' \mathbf{C}$, thus \mathbf{B} has the same eigenvalues as $\mathbf{C}' \mathbf{C}$.

It is efficient and convenient to focus on the properties of $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$ and their eigenvalues and eigenvectors. The eigenvectors of $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$ will lead to eigenvectors of \mathbf{A} and \mathbf{B} , which are the objectives of CCA.

Common non-zero eigenvalues of $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$: ρ_i^{*2}

Matrices $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$ have several nice algebraic properties:

- They are symmetric matrices (while \mathbf{A}, \mathbf{B} are not necessarily symmetric). Therefore all of their eigenvalues are real, and the eigenvectors can be chosen to be mutually orthogonal.
- They are positive semi-definite matrices. Therefore the eigenvalues are non-negative.
- In addition, $\mathbf{C}' \mathbf{C}$ and $\mathbf{C} \mathbf{C}'$ share the same nonzero eigenvalues.

Label and sort the p eigenvalues of the $p \times p$ matrix $\mathbf{C} \mathbf{C}'$ in descending order as

$$\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_r^{*2} \geq 0, \quad r = \min\{p, q\} \quad (\rho_i^* \geq 0)$$

The top r eigenvalues $\rho_i^{*2}, i = 1, \dots, r$, equal to the r largest eigenvalues of the $q \times q$ matrix $\mathbf{C}' \mathbf{C}$. Other eigenvalues of $\mathbf{C}' \mathbf{C}$ or $\mathbf{C} \mathbf{C}'$ must be zero.

By the similarity relationship in (4) and (5), eigenvalues $\rho_1^{*2}, \dots, \rho_r^{*2}$, are ordered r largest eigenvalues of \mathbf{A} and \mathbf{B} as well. All other eigenvalues of \mathbf{A} and \mathbf{B} must be zero.

Orthonormal eigenvectors of $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$: \mathbf{e}_i and \mathbf{f}_i

Below we obtain related eigenvectors of $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$, which will lead to eigenvectors of \mathbf{A} and \mathbf{B} .

Let \mathbf{e}_i be the p -variate unit eigenvector of $\mathbf{C} \mathbf{C}'$ corresponding to eigenvalue $\rho_i^{*2}, i = 1, \dots, r$.

$$\mathbf{C} \mathbf{C}' \mathbf{e}_i = \rho_i^{*2} \mathbf{e}_i.$$

Then

$$\mathbf{C}' \mathbf{C} (\mathbf{C}' \mathbf{e}_i) = \mathbf{C}' (\mathbf{C} \mathbf{C}' \mathbf{e}_i) = \mathbf{C}' (\rho_i^{*2} \mathbf{e}_i) = \rho_i^{*2} (\mathbf{C}' \mathbf{e}_i)$$

which means the q -variate vector $\mathbf{C}' \mathbf{e}_i$ and its multiples are eigenvectors of $\mathbf{C}' \mathbf{C}$ with eigenvalue ρ_i^{*2} . This also shows that $\mathbf{C} \mathbf{C}'$ and $\mathbf{C}' \mathbf{C}$ share the same non-zero eigenvalues.

In particular, we choose \mathbf{f}_i to be the q -variate unit eigenvector of $\mathbf{C}' \mathbf{C}$ corresponding to eigenvalue ρ_i^{*2} . Since \mathbf{f}_i and $\mathbf{C}' \mathbf{e}_i$ are eigenvectors of the same eigenvalue, we may let $\mathbf{f}_i = c_i \mathbf{C}' \mathbf{e}_i$ for some constant c_i , and $\mathbf{f}_i' \mathbf{f}_i = 1$.

By the symmetric, positive semi-definiteness of $\mathbf{C} \mathbf{C}'$, $\{\mathbf{e}_i\}_{i=1}^p$ can be obtained and can be chosen to be orthonormal:

$$\mathbf{e}_i' \mathbf{e}_j = \delta_{ij} = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad i, j = 1, \dots, p.$$

To determine the constant c_i in \mathbf{f}_i , use the desired orthogonal condition

$$\mathbf{f}_i' \mathbf{f}_j = \delta_{ij} = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

That is,

$$\mathbf{f}_i' \mathbf{f}_j = (c_i \mathbf{C}' \mathbf{e}_i)' (c_j \mathbf{C}' \mathbf{e}_j) = c_i c_j \mathbf{e}_i' \mathbf{C} \mathbf{C}' \mathbf{e}_j = c_i c_j \mathbf{e}_i' \rho_j^{*2} \mathbf{e}_j = c_i c_j \rho_j^{*2} \delta_{ij} = \delta_{ij} \iff c_i = 1/\rho_i^*$$

which leads to

$$\mathbf{f}_i = (1/\rho_i^*) \mathbf{C}' \mathbf{e}_i$$

Orthonormal eigenvectors can be established for zero eigenvalues.

Now both $\{\mathbf{e}_i\}_{i=1}^p$ $\{\mathbf{f}_i\}_{i=1}^q$ are chosen to be orthonormal:

$$\mathbf{e}_i' \mathbf{e}_j = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases} \quad i, j = 1, \dots, p; \quad \mathbf{f}_i' \mathbf{f}_j = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases} \quad i, j = 1, \dots, q.$$

Eigenvectors of \mathbf{A} and \mathbf{B} : \mathbf{a}_i and \mathbf{b}_i

Let

$$\mathbf{a}_i = \Sigma_{11}^{-1/2} \mathbf{e}_i, \quad \mathbf{b}_i = \Sigma_{22}^{-1/2} \mathbf{f}_i.$$

Then

$$\mathbf{A} \mathbf{a}_i = \mathbf{A} (\Sigma_{11}^{-1/2} \mathbf{e}_i) = \Sigma_{11}^{-1/2} \mathbf{C} \mathbf{C}' \Sigma_{11}^{1/2} (\Sigma_{11}^{-1/2} \mathbf{e}_i) = \Sigma_{11}^{-1/2} \mathbf{C} \mathbf{C}' \mathbf{e}_i = \Sigma_{11}^{-1/2} (\rho_i^{*2} \mathbf{e}_i) = \rho_i^{*2} (\Sigma_{11}^{-1/2} \mathbf{e}_i) = \rho_i^{*2} \mathbf{a}_i$$

Therefore $\mathbf{a}_i = \Sigma_{11}^{-1/2} \mathbf{e}_i$ is an eigenvector of \mathbf{A} with eigenvalue ρ_i^{*2} for $i = 1, \dots, p$.

Similarly we can show that $\mathbf{b}_i = \Sigma_{22}^{-1/2} \mathbf{f}_i$ is an eigenvector of \mathbf{B} with eigenvalue ρ_i^{*2} for $i = 1, \dots, q$.

Optimal \mathbf{a} and \mathbf{b} maximizing canonical correlation

The canonical variate pairs $\mathbf{a}_i, \mathbf{b}_i$ are eigenvectors of matrices \mathbf{A}, \mathbf{B} with the same eigenvalue ρ_i^{*2} . The pair of linear combinations $\mathbf{a}_i' \mathbf{X}$ and $\mathbf{b}_i' \mathbf{Y}$ has correlation

$$\begin{aligned} \rho &= \text{corr}(\mathbf{a}_i' \mathbf{X}, \mathbf{b}_i' \mathbf{Y}) = \text{cov}(\mathbf{a}_i' \mathbf{X}, \mathbf{b}_i' \mathbf{Y}) = \mathbf{a}_i' \Sigma_{12} \mathbf{b}_i = (\Sigma_{11}^{-1/2} \mathbf{e}_i)' \Sigma_{12} (\Sigma_{22}^{-1/2} \mathbf{f}_i) \\ &= \mathbf{e}_i' (\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}) (\mathbf{f}_i) \\ &= \mathbf{e}_i' \mathbf{C} \mathbf{f}_i \\ &= \mathbf{e}_i' \mathbf{C} (1/\rho_i^*) \mathbf{C}' \mathbf{e}_i \\ &= (1/\rho_i^*) \mathbf{e}_i' \mathbf{C} \mathbf{C}' \mathbf{e}_i \\ &= (1/\rho_i^*) \mathbf{e}_i' \rho_i^{*2} \mathbf{e}_i = \rho_i^* \mathbf{e}_i' \mathbf{e}_i \\ &= \rho_i^* \end{aligned}$$

There are at most $r = \min\{p, q\}$ such pairs, up to a factor of (-1) of \mathbf{a}_i and \mathbf{b}_i . Conventionally $\rho \geq 0$.

Choose $\mathbf{a}_1, \mathbf{b}_1$, the first pair of canonical variables. Now we can achieve $\rho = \rho_1^*$, the common largest eigenvalue of \mathbf{A} and \mathbf{B} , which maximizes (1), by a pair of linear combinations $U = \mathbf{a}_1' \mathbf{X}$ and $V = \mathbf{b}_1' \mathbf{Y}$.

This concludes the derivation of the leading pair of canonical variates. □

The above procedure produces

$$U_1 = \mathbf{a}_1' \mathbf{X}, \quad V_1 = \mathbf{b}_1' \mathbf{Y}$$

capturing maximum correlation

$$\text{corr}(U_1, V_1) = \rho_1^*$$

among all possible linear combinations $\mathbf{a}' \mathbf{X}$ and $\mathbf{b}' \mathbf{Y}$ (under constraints $\mathbf{a}' \Sigma_{11} \mathbf{a} = 1$, $\mathbf{b}' \Sigma_{22} \mathbf{b} = 1$).

If one wants to explore the rest correlations between \mathbf{X} and \mathbf{Y} , the procedure can go on to find the second pair $U_2 = \mathbf{a}_2' \mathbf{X}$ and $V_2 = \mathbf{b}_2' \mathbf{Y}$ with maximum $\text{corr}(U_2, V_2)$. We may define the Lagrangian function

$$L_2 = \mathbf{a}_2' \Sigma_{12} \mathbf{b}_2 - \frac{1}{2} \lambda (\mathbf{a}_2' \Sigma_{11} \mathbf{a}_2 - 1) - \frac{1}{2} \gamma (\mathbf{b}_2' \Sigma_{22} \mathbf{b}_2 - 1) - \delta \mathbf{a}_2' \Sigma_{11} \mathbf{a}_1 - \nu \mathbf{b}_2' \Sigma_{22} \mathbf{b}_1 - \beta \mathbf{a}_1' \Sigma_{12} \mathbf{b}_2 - \alpha \mathbf{b}_1' \Sigma_{21} \mathbf{a}_2$$

Setting

$$\begin{aligned} \frac{\partial L_2}{\partial \mathbf{a}_2} &= \mathbf{0}_p \\ \frac{\partial L_2}{\partial \mathbf{b}_2} &= \mathbf{0}_q \end{aligned}$$

we can obtain the second pair of canonical covariates (U_2, V_2) under the normalization $\text{var}(U_2) = \text{var}(V_2) = 1$, achieving the strongest correlation $\text{Cov}(U_2, V_2) = \rho_2^*$, and are uncorrelated to the first canonical pair (U_1, V_1) by satisfying the constraints

$$\text{cov}(U_1, U_2) = 0, \quad \text{cov}(V_1, V_2) = 0, \quad \text{cov}(U_1, V_2) = 0, \quad \text{cov}(U_2, V_1) = 0$$

The process may continue, if desired, to obtain all pairs of canonical variates.

- The first pair is the most desirable one. In all normalized (variance = 1) linear combination variables

$$U_1 = \mathbf{a}_1' \mathbf{X}, \quad V_1 = \mathbf{b}_1' \mathbf{Y}, \quad \text{var}(U_1) = 1, \quad \text{var}(V_1) = 1.$$

the (U_1, V_1) pair that maximizes $\text{Corr}(U_1, V_1)$ in (1) among all $\mathbf{a}_1, \mathbf{b}_1$ are attained at

$$\mathbf{a}_1 = \Sigma_{11}^{-1/2} \mathbf{e}_1, \quad \mathbf{b}_1 = \Sigma_{22}^{-1/2} \mathbf{f}_1,$$

so

$$U_1 = \mathbf{e}_1' \Sigma_{11}^{-1/2} \mathbf{X}, \quad V_1 = \mathbf{f}_1' \Sigma_{22}^{-1/2} \mathbf{Y}, \quad \text{Cov}(U_1, V_1) = \rho_1^*.$$

It verifies that

$$\text{var}(U_1) = \mathbf{e}_1' \Sigma_{11}^{-1/2} \text{var}(\mathbf{X}) \Sigma_{11}^{-1/2} \mathbf{e}_1 = \mathbf{e}_1' \mathbf{e}_1 = 1, \quad \text{var}(V_1) = \dots = \mathbf{f}_1' \mathbf{f}_1 = 1.$$

- The second pair of normalized canonical variables are

$$U_2 = \mathbf{a}_2' \mathbf{X}, \quad V_2 = \mathbf{b}_2' \mathbf{Y}, \quad \text{var}(U_2) = \text{var}(V_2) = 1, \quad \text{Cov}(U_2, V_2) = \rho_2^*.$$

Additional constraints are

$$\mathbf{a}_2' \Sigma_{11} \mathbf{a}_1 = 0, \quad \mathbf{b}_2' \Sigma_{22} \mathbf{b}_1 = 0, \quad \mathbf{a}_1' \Sigma_{12} \mathbf{b}_2 = 0, \quad \mathbf{a}_2' \Sigma_{12} \mathbf{b}_1 = 0,$$

which imply that

$$\text{cov}(U_1, U_2) = 0, \quad \text{cov}(V_1, V_2) = 0, \quad \text{cov}(U_1, V_2) = 0, \quad \text{cov}(U_2, V_1) = 0.$$

- The k th pair of canonical variables are

$$U_k = \mathbf{e}_k' \Sigma_{11}^{-1/2} \mathbf{X}, \quad V_k = \mathbf{f}_k' \Sigma_{22}^{-1/2} \mathbf{Y} \quad \text{with} \quad \text{var}(U_k) = \text{var}(V_k) = 1, \quad \text{Cov}(U_k, V_k) = \rho_k^*.$$

All satisfies the condition that, for $\ell < k$,

$$\mathbf{a}_k' \Sigma_{11} \mathbf{a}_\ell = 0, \quad \mathbf{b}_k' \Sigma_{22} \mathbf{b}_\ell = 0, \quad \mathbf{a}_\ell' \Sigma_{12} \mathbf{b}_k = 0, \quad \mathbf{a}_k' \Sigma_{12} \mathbf{b}_\ell = 0.$$

- The process stops at $k = \min\{p, q\}$.

Summary of properties of canonical variables $U_i = \mathbf{a}_i' \mathbf{X}$, $V_i = \mathbf{b}_i' \mathbf{Y}$

- $\text{var}(U_k) = \text{var}(V_k) = 1$
- $\text{cov}(U_k, U_\ell) = 0, \text{cov}(V_k, V_\ell) = 0$ for $k \neq \ell$.
- $\text{cov}(U_k, V_\ell) = 0$ for $k \neq \ell$
- $\text{cov}(U_i, V_i) = \rho_i^*$,
where ρ_i^{*2} is the common i th largest eigenvalue of the following matrices:

$$\begin{aligned} &- p \times p \text{ matrix } \mathbf{C}\mathbf{C}' = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \\ &- q \times q \text{ matrix } \mathbf{C}'\mathbf{C} = \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} \\ &- p \times p \text{ matrix } \mathbf{A} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Sigma_{11}^{-1/2} \mathbf{C}\mathbf{C}' \Sigma_{11}^{1/2} \\ &- q \times q \text{ matrix } \mathbf{B} = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \Sigma_{22}^{-1/2} \mathbf{C}'\mathbf{C} \Sigma_{22}^{1/2} \end{aligned}$$

and

$$\mathbf{A}\mathbf{a}_i = \rho_i^{*2} \mathbf{a}_i, \quad \mathbf{B}\mathbf{b}_i = \rho_i^{*2} \mathbf{b}_i$$

The proofs are a repeated use of extended Cauchy-Schwarz inequality and its corollary the maximization lemma.

Covariance matrix of canonical variables

The covariance matrix of the canonical variables has neat form.

$$\text{Cov}(U_1, \dots, U_r, V_1, \dots, V_r) = \begin{bmatrix} I_r & R_r \\ R_r & I_r \end{bmatrix}, \quad R_r = \text{diag}[\rho_1^*, \dots, \rho_r^*]$$

Sort in canonical variate pairs,

$$\text{Cov}(U_1, V_1, \dots, U_r, V_r) = \text{diag}[D_1, \dots, D_r], \quad D_k = \begin{bmatrix} 1 & \rho_k^* \\ \rho_k^* & 1 \end{bmatrix}$$

To scale or not to scale

If there are large variations in the magnitudes of component variable in multivariate data, we know that principal component analysis on the original variables and that on scaled data with $\text{var}(X_i) = 1$ may yield very different results.

Does scaling matter in canonical correlation analysis?

Recall the relationship between covariance matrix Σ and correlation matrix \mathbf{R} ,

$$\Sigma = \mathbf{D}\mathbf{R}\mathbf{D}, \quad \mathbf{R} = \mathbf{D}^{-1} \Sigma \mathbf{D}^{-1}$$

where \mathbf{D} is the diagonal matrix with i th diagonal element $\sqrt{\sigma_{ii}}$. We may write

$$\mathbf{D} = [\text{diag}(\Sigma)]^{1/2}$$

Let (\mathbf{X}, \mathbf{Y}) be the pair of the original random vectors, $(\mathbf{X}^*, \mathbf{Y}^*)$ be the scaled pair with each component variable with variance = 1. The relationship can be expressed as

$$(\mathbf{X}^*, \mathbf{Y}^*) = (D_1^{-1}\mathbf{X}, D_2^{-1}\mathbf{Y})$$

with

$$D_1 = [\text{diag}(\Sigma_{11})]^{1/2}, \quad D_2 = [\text{diag}(\Sigma_{22})]^{1/2}.$$

If $(U_k, V_k) = (\mathbf{a}'_k \mathbf{X}, \mathbf{b}'_k \mathbf{Y})$ is the k th pair of canonical correlation variates of the original variables, with

$$\text{cov}(U_k, V_k) = \text{cov}(\mathbf{a}'_k \mathbf{X}, \mathbf{b}'_k \mathbf{Y}) = \rho_k^*$$

and if $(U_k^*, V_k^*) = (\mathbf{a}^{*'}_k \mathbf{X}^*, \mathbf{b}^{*'}_k \mathbf{Y}^*)$ is the k th pair of canonical correlation variates of the scaled variance-one variables, then from the derivation process, we have

$$\mathbf{a}^*_k = D_1 \mathbf{a}_k, \quad \mathbf{b}^*_k = D_2 \mathbf{b}_k$$

and

$$\begin{aligned} \text{cov}(U_k^*, V_k^*) &= \text{cov}(\mathbf{a}^{*'}_k \mathbf{X}^*, \mathbf{b}^{*'}_k \mathbf{Y}^*) \\ &= \text{cov}((D_1 \mathbf{a}_k)' D_1^{-1} \mathbf{X}, (D_2 \mathbf{b}_k)' D_2^{-1} \mathbf{Y}) \\ &= \text{cov}(\mathbf{a}_k' D_1' D_1^{-1} \mathbf{X}, \mathbf{b}_k' D_2' D_2^{-1} \mathbf{Y}) \\ &= \text{cov}(\mathbf{a}'_k \mathbf{X}, \mathbf{b}'_k \mathbf{Y}) \\ &= \text{cov}(U_k, V_k) = \rho_k^* \end{aligned}$$

Therefore, the canonical coefficients for the scaled variance-one variables are simply related to the canonical coefficients attached to the original variables, by component-wise scaling.

The canonical correlations, the ρ_k^* 's, are unchanged by the scaling. If the strength of the correlation between the two groups of variables is the main focus (as it often is), the scaling does not matter.

Of course there are the usual non-uniqueness of $\mathbf{a}_k, \mathbf{b}_k$ when $\rho_k^* = \rho_{k+1}^*$ happens.

Remarks on canonical variables

- Matrix $\mathbf{C} = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$, sometimes called the canonical correlation matrix, can be viewed as a generalization of the correlation coefficient of two univariate random variables to two multivariate random vectors.
- The pair of matrices

$$\mathbf{C}\mathbf{C}' = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \text{ and } \mathbf{C}'\mathbf{C} = \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2}$$

called the matrices of multivariate coefficients of determination, are natural generalizations of population coefficient of determination (square of the correlation between Y and its best linear predictor, may not be covered in this course) in linear regression.

The derivation of canonical variates shows that the ρ_{*i} 's are correlations, thus matrices \mathbf{A}, \mathbf{B} , positive semi-definite matrices $\mathbf{C}\mathbf{C}'$ and $\mathbf{C}'\mathbf{C}$ all have eigenvalues in $[0, 1]$.

Comparisons

Canonical correlation analysis (CCA) can be viewed as a multivariate statistical technique similar in spirit to principal component analysis (PCA).

In Principal Component Analysis, from $\mathbf{X} = [X_1 \ X_2 \ \cdots \ X_p]'$, we derived principal components $\mathbf{a}_i \mathbf{X}$, $i = 1, \dots, p$. The uncorrelated PC variables capture the variation in the data and reduce the dimension of the data effectively. However certain structures of the data can be obscured or lost in the process.

CCA evaluates a particular structure in $[\mathbf{X}' \ \mathbf{Y}']'$ = $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$, which can be described as the correlations between two random vectors \mathbf{X} and \mathbf{Y} .

While PCA works with a single random vector and maximizes the variance of projections of the data, CCA works with a pair of random vectors (or generalized to a set of m random vectors) and maximizes correlation between sets of projections.

2 Sample CCA

Similar to the practical usage of principal component analysis, in applications, the observed data are often treated as the population to carry out canonical correlation analysis directly, without probability distribution assumptions on the data generating mechanism.

- In the place of population random vectors, we work with sample data

$$[\mathbf{X}_{n \times p} \ \mathbf{Y}_{n \times q}]$$

- In the place to Σ_{ij} are sample covariance matrices. The joint sample covariance matrix of all $p + q$ variables can be in terms of sample covariance matrices of x and y variables, as well as their covariance.

$$\mathbf{S} = \begin{bmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{bmatrix}$$

- The first pair of sample canonical-variate vectors are given be

$$(\hat{\mathbf{a}}_1, \hat{\mathbf{b}}_1) = \arg\max_{\mathbf{a} \in \mathbb{R}^p, \mathbf{b} \in \mathbb{R}^q} \{ \mathbf{a}' S_{xy} \mathbf{b} : \mathbf{a}' S_x \mathbf{a} = 1, \mathbf{b}' S_y \mathbf{b} = 1 \}$$

- The second pair of sample canonical-variate vectors are given be

$$(\hat{\mathbf{a}}_2, \hat{\mathbf{b}}_2) = \arg\max_{\mathbf{a}, \mathbf{b}} \{ \mathbf{a}' S_{xy} \mathbf{b} : \mathbf{a}' S_x \mathbf{a} = \mathbf{b}' S_y \mathbf{b} = 1, \mathbf{a}' S_x \mathbf{a}_1 = \mathbf{a}' S_{xy} \mathbf{b}_1 = \mathbf{b}' S_{yx} \mathbf{a}_1 = \mathbf{b}' S_y \mathbf{b}_1 = 0 \}$$

- For $k \leq \min\{p, q\}$, the k th pair of sample canonical-variate vectors are given be

$$(\hat{\mathbf{a}}_k, \hat{\mathbf{b}}_k) = \arg\max_{\mathbf{a}, \mathbf{b}} \{ \mathbf{a}' S_{xy} \mathbf{b} : \mathbf{a}' S_x \mathbf{a} = \mathbf{b}' S_y \mathbf{b} = 1, \mathbf{a}' S_x \mathbf{a}_i = \mathbf{a}' S_{xy} \mathbf{b}_i = \mathbf{b}' S_{yx} \mathbf{a}_i = \mathbf{b}' S_y \mathbf{b}_i = 0, i = 1, \dots, k-1 \}$$

- The achieved k th sample canonical correlation is

$$\hat{\rho}_k^* = \hat{\mathbf{a}}_k' S_{xy} \hat{\mathbf{b}}_k, \quad k = 1, \dots, \min\{p, q\}$$

where

$$\hat{\rho}_1^{*2} \geq \hat{\rho}_2^{*2} \geq \dots \geq \hat{\rho}_r^{*2} \geq 0, \quad r = \min\{p, q\} \quad (\hat{\rho}_i^* \geq 0)$$

are common eigenvalues of the matrices of $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}\hat{\mathbf{C}}', \hat{\mathbf{C}}, \hat{\mathbf{C}}'$.