

FINM 331: DATA ANALYSIS FOR FINANCE AND STATISTICS
FALL 2015
CANONICAL CORRELATION ANALYSIS

1. POPULATION CANONICAL CORRELATION ANALYSIS

- in population PCA, we deal with one set of random variables
- in population CCA, we deal with two sets of random variables
- given random variables X_1, \dots, X_p and Y_1, \dots, Y_q , we assemble them into two random vectors

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

- without loss of generality, let $p \leq q$ throughout this handout
- we introduce the following notations

$$\begin{aligned} \boldsymbol{\mu}_X &= E(\mathbf{X}) \in \mathbb{R}^p, & \boldsymbol{\mu}_Y &= E(\mathbf{Y}) \in \mathbb{R}^q, \\ \Sigma_X &= \text{Cov}(\mathbf{X}) \in \mathbb{R}^{p \times p}, & \Sigma_Y &= \text{Cov}(\mathbf{Y}) \in \mathbb{R}^{q \times q}, \\ \Sigma_{XY} &= \text{Cov}(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{p \times q}, & \Sigma_{YX} &= \text{Cov}(\mathbf{Y}, \mathbf{X}) \in \mathbb{R}^{q \times p} \end{aligned}$$

- note that

$$\Sigma_{XY}^\top = \Sigma_{YX}$$

- now write

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \\ Y_1 \\ \vdots \\ Y_q \end{bmatrix}$$

and denote

$$\begin{aligned} \boldsymbol{\mu} &= E(\mathbf{W}) = \begin{bmatrix} E(\mathbf{X}) \\ E(\mathbf{Y}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} \in \mathbb{R}^{p+q} \\ \Sigma &= \text{Cov}(\mathbf{W}) = E(\mathbf{W} - \boldsymbol{\mu})(\mathbf{W} - \boldsymbol{\mu})^\top \\ &= \begin{bmatrix} E(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^\top & E(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^\top \\ E(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{X} - \boldsymbol{\mu}_X)^\top & E(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{Y} - \boldsymbol{\mu}_Y)^\top \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \in \mathbb{R}^{(p+q) \times (p+q)} \end{aligned}$$

- consider two linear combinations

$$U = \mathbf{a}^\top \mathbf{X}, \quad V = \mathbf{b}^\top \mathbf{Y}$$

with $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^q$ constant vectors (note that U and V are random variables)

- by results in the Handout 4, the variances and covariances of U and V are given by

$$\begin{aligned}\text{Var}(U) &= \mathbf{a}^\top \text{Cov}(\mathbf{X}) \mathbf{a} = \mathbf{a}^\top \Sigma_X \mathbf{a} \in \mathbb{R}, \\ \text{Var}(V) &= \mathbf{b}^\top \text{Cov}(\mathbf{Y}) \mathbf{b} = \mathbf{b}^\top \Sigma_Y \mathbf{b} \in \mathbb{R}, \\ \text{Cov}(U, V) &= \mathbf{a}^\top \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b} = \mathbf{a}^\top \Sigma_{XY} \mathbf{b} \in \mathbb{R}, \\ \text{Corr}(U, V) &= \frac{\mathbf{a}^\top \Sigma_{XY} \mathbf{b}}{\sqrt{\mathbf{a}^\top \Sigma_X \mathbf{a}} \sqrt{\mathbf{b}^\top \Sigma_Y \mathbf{b}}} \in \mathbb{R}\end{aligned}$$

- note that

$$\text{Cov}(U, V) = \mathbf{a}^\top \Sigma_{XY} \mathbf{b} = \mathbf{b}^\top \Sigma_{XY} \mathbf{a} = \text{Cov}(V, U)$$

- the first pair of *population canonical variables* is defined by

$$(U_1, V_1) = \text{argmax}\{\text{Corr}(U, V) : \text{Var}(U) = 1 = \text{Var}(V)\}$$

- the second pair of population canonical variables is defined by

$$\begin{aligned}(U_2, V_2) &= \text{argmax}\{\text{Corr}(U, V) : \text{Var}(U) = \text{Var}(V) = 1, \\ &\quad \text{Cov}(U, U_1) = \text{Cov}(V, V_1) = \text{Cov}(U, V_1) = \text{Cov}(V, U_1) = 0\}\end{aligned}$$

- the k th pair of population canonical variables is defined by

$$\begin{aligned}(U_k, V_k) &= \text{argmax}\{\text{Corr}(U, V) : \text{Var}(U) = \text{Var}(V) = 1, \\ &\quad \text{Cov}(U, U_j) = \text{Cov}(V, V_j) = \text{Cov}(U, V_j) = \text{Cov}(V, U_j) = 0, \quad j = 1, \dots, k-1\}\end{aligned}$$

for $k = 1, \dots, p$

- if we replace argmax by \max , the corresponding maximal values are called the k th *population canonical correlation* and denoted

$$\rho_k = \text{Corr}(U_k, V_k)$$

Theorem 1. Suppose $p \leq q$. Let

$$\Gamma_{XY} = \Sigma_X^{-1/2} \Sigma_{XY} \Sigma_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

and the condensed SVD of Γ_{XY} be

$$\Gamma_{XY} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_p^\top \end{bmatrix},$$

where $\mathbf{u}_1, \dots, \mathbf{u}_p \in \mathbb{R}^p$ are the left singular vectors, $\sigma_1 \geq \dots \geq \sigma_p$ are the singular values, and $\mathbf{v}_1, \dots, \mathbf{v}_p \in \mathbb{R}^q$ are the right singular vectors. Then

$$U_k = \mathbf{u}_k^\top \Sigma_X^{-1/2} \mathbf{X}, \quad V_k = \mathbf{v}_k^\top \Sigma_Y^{-1/2} \mathbf{Y}$$

are the k th population canonical variables and

$$\text{Corr}(U_k, V_k) = \sigma_k$$

is the k th population canonical correlation for $k = 1, \dots, p$.

- the proof will be in the homework
- the vectors

$$\mathbf{a}_k = \Sigma_X^{-1/2} \mathbf{u}_k \in \mathbb{R}^p, \quad \mathbf{b}_k = \Sigma_Y^{-1/2} \mathbf{v}_k \in \mathbb{R}^q$$

are called the *population canonical vectors* and we have

$$U_k = \mathbf{a}_k^\top \mathbf{X}, \quad V_k = \mathbf{b}_k^\top \mathbf{Y}$$

- every symmetric positive semidefinite matrix $A \in \mathbb{R}^{p \times p}$ has a *square root* $A^{1/2} \in \mathbb{R}^{p \times p}$ that is also symmetric positive semidefinite
- this can be defined via the EVD of A : if

$$A = Q\Lambda Q^\top$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ is the diagonal matrix of nonnegative (since A is symmetric positive semidefinite) eigenvalues, then

$$A^{1/2} = Q\Lambda^{1/2}Q^\top$$

where $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$

- you can check that the population canonical variables in Theorem 1 satisfy

$$\text{Cov}(U_i, U_j) = \text{Cov}(V_i, V_j) = \delta_{ij}, \quad \text{Cov}(U_i, V_j) = \text{Cov}(V_i, U_j) = 0$$

for $i, j = 1, \dots, p$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

- recall from the properties of SVD that $\sigma_1^2 \geq \dots \geq \sigma_p^2$ are the eigenvalues of the matrix

$$\Gamma_{XY}\Gamma_{XY}^\top = \Sigma_X^{-1/2}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1/2} \in \mathbb{R}^{p \times p}$$

with corresponding eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$

- $\sigma_1^2 \geq \dots \geq \sigma_p^2$ are also the p largest eigenvalues of the matrix

$$\Gamma_{XY}^\top\Gamma_{XY} = \Sigma_Y^{-1/2}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1/2} \in \mathbb{R}^{q \times q}$$

with corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_p$

- CCA has a nice property not found in PCA and FA

Theorem 2. Let $\mathbf{X} = [X_1, \dots, X_p]^\top$ and $\mathbf{Y} = [Y_1, \dots, Y_q]^\top$ be random vectors. If

$$\mathbf{X}' = M\mathbf{X} + \mathbf{c}, \quad \mathbf{Y}' = N\mathbf{Y} + \mathbf{d}$$

for some nonsingular matrices $M \in \mathbb{R}^{p \times p}$, $N \in \mathbb{R}^{q \times q}$ and some vectors $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{d} \in \mathbb{R}^q$, then

- the canonical correlations between \mathbf{X}' and \mathbf{Y}' are the same as the canonical correlations between \mathbf{X} and \mathbf{Y} ;
- the canonical correlation vectors of \mathbf{X}' and \mathbf{Y}' are

$$\mathbf{a}'_k = M^{-\top}\mathbf{a}_k \in \mathbb{R}^p, \quad \mathbf{b}'_k = N^{-\top}\mathbf{b}_k \in \mathbb{R}^q$$

where $\mathbf{a}_k \in \mathbb{R}^p$ and $\mathbf{b}_k \in \mathbb{R}^q$ are the canonical vectors between \mathbf{X} and \mathbf{Y} .

- note that if M is nonsingular, then M^\top must also be nonsingular and

$$(M^\top)^{-1} = (M^{-1})^\top$$

we denote this matrix (inverse of transpose or transpose of inverse) by $M^{-\top}$

- just as PCA and FA apply with correlation matrix in place of covariance matrix, the same is true for CCA, we may use

$$P_X^{-1/2}P_{XY}P_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

in place of

$$\Sigma_X^{-1/2}\Sigma_{XY}\Sigma_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

in Theorem 1

- thanks to Theorem 2, there is a straightforward relation between using correlation and covariance matrices

- if $\mathbf{a}'_k \in \mathbb{R}^p$ and $\mathbf{b}'_k \in \mathbb{R}^q$ are the canonical vectors of \mathbf{X} and \mathbf{Y} coming the correlation matrix version of CCA and $\mathbf{a}_k \in \mathbb{R}^p$ and $\mathbf{b}_k \in \mathbb{R}^q$ are the canonical vectors of \mathbf{X} and \mathbf{Y} coming the covariance matrix version, then

$$\mathbf{a}'_k = V_X^{-1/2} \mathbf{a}_k, \quad \mathbf{b}'_k = V_Y^{-1/2} \mathbf{b}_k \quad (1.1)$$

where $V_X^{-1/2} \in \mathbb{R}^{p \times p}$ and $V_Y^{-1/2} \in \mathbb{R}^{q \times q}$ are the population standard deviation matrices, i.e.,

$$V_X = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_p)) \in \mathbb{R}^{p \times p}, \quad V_Y = \text{diag}(\text{Var}(Y_1), \dots, \text{Var}(Y_q)) \in \mathbb{R}^{q \times q}$$

- this follows from applying Theorem 2 with $M = V_X^{-1/2}$ and $N = V_Y^{-1/2}$
- note that there is no equivalent for PCA — there is no simple relationship in general between the principal components obtained from the covariance matrix and those obtained from the correlation matrix
- Theorem 2 and the discussion that follows apply also to sample CCA

2. SAMPLE CANONICAL CORRELATION ANALYSIS

- in sample PCA, we deal with one data matrix
- in sample CCA, we deal with two data matrices

$$X = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times p}, \quad Y = \begin{bmatrix} \mathbf{y}_1^\top \\ \mathbf{y}_2^\top \\ \vdots \\ \mathbf{y}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times q}$$

- the relation between population and sample CCA, again, is the following: the (i, j) th entry of the data matrix

$$\begin{aligned} x_{ij} &= i\text{th observation of the random variable } X_j, \\ y_{ik} &= i\text{th observation of the random variable } Y_k, \end{aligned}$$

where $i = 1, \dots, n$, $j = 1, \dots, p$, $k = 1, \dots, q$

- so each row vector $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]^\top \in \mathbb{R}^p$ is an observation of the random vector $\mathbf{X} = [X_1, \dots, X_p]^\top$, each row vector $\mathbf{y}_i = [y_{i1}, \dots, y_{iq}]^\top \in \mathbb{R}^q$ is an observation of the random vector $\mathbf{Y} = [Y_1, \dots, Y_q]^\top$
- note that in CCA, the observations must come in pairs, so that the number of rows of the two data matrices X and Y must be the same
- we will assemble the data matrices into a single one

$$W = [X, Y] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} & y_{11} & y_{12} & \cdots & y_{1q} \\ x_{21} & x_{22} & \cdots & x_{2p} & y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} & y_{n1} & y_{n2} & \cdots & y_{nq} \end{bmatrix} \in \mathbb{R}^{n \times (p+q)}$$

- the sample mean of W is

$$\bar{\mathbf{w}} = \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix} \in \mathbb{R}^{p+q}, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \in \mathbb{R}^p, \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \in \mathbb{R}^q$$

- the sample mean of W is

$$S = \begin{bmatrix} S_X & S_{XY} \\ S_{YX} & S_Y \end{bmatrix} \in \mathbb{R}^{(p+q) \times (p+q)},$$

$$S_X = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \in \mathbb{R}^{p \times p}, \quad S_Y = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top \in \mathbb{R}^{q \times q},$$

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top = S_{YX}^\top \in \mathbb{R}^{p \times q}$$

- for any $\mathbf{a} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$, we form

$$\alpha = \mathbf{a}^\top \mathbf{x}, \quad \beta = \mathbf{b}^\top \mathbf{y}$$

where $\mathbf{x} = [x_1, \dots, x_p]^\top$ and $\mathbf{y} = [y_1, \dots, y_q]^\top$ are vectors of (usual, not random) variables that represent unspecified observations of the random vectors \mathbf{X} and \mathbf{Y} , then α and β have sample correlation

$$r_{\alpha\beta} = \frac{\mathbf{a}^\top S_{XY} \mathbf{b}}{\sqrt{\mathbf{a}^\top S_X \mathbf{a}} \sqrt{\mathbf{b}^\top S_Y \mathbf{b}}} \in \mathbb{R}$$

- note that the $r_{\alpha\beta}$ does not depend on \mathbf{x} and \mathbf{y} , which are just ‘dummy variables’
- a linear combination of the form

$$\mathbf{a}^\top \mathbf{x} = a_1 x_1 + \dots + a_p x_p$$

where x_1, \dots, x_p are variables (in the sense of indeterminates) is just an alternative way to specify the vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} \in \mathbb{R}^p$$

- statisticians like this way of specifying a vector because they could write things like

$$123\text{IQ} + 234\text{GRE} + 3.45\text{GPA}$$

for which it is clear what the numbers measure instead of

$$\begin{bmatrix} 123 \\ 234 \\ 3.45 \end{bmatrix} \in \mathbb{R}^3$$

for which one would have to specify what the meaning each coordinate on the side

- we have the following sample analogue of Theorem 1

Theorem 3. Suppose $p \leq q$. Let

$$G_{XY} = S_X^{-1/2} S_{XY} S_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

and the condensed SVD of G_{XY} be

$$G_{XY} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_p^\top \end{bmatrix},$$

where $\mathbf{u}_1, \dots, \mathbf{u}_p \in \mathbb{R}^p$ are the left singular vectors, $\sigma_1 \geq \dots \geq \sigma_p$ are the singular values, and $\mathbf{v}_1, \dots, \mathbf{v}_p \in \mathbb{R}^q$ are the right singular vectors. Then

$$\hat{\mathbf{a}}_k = S_X^{-1/2} \mathbf{u}_k \in \mathbb{R}^p, \quad \hat{\mathbf{b}}_k = S_Y^{-1/2} \mathbf{v}_k \in \mathbb{R}^q$$

are the k th sample canonical vectors,

$$\alpha_k = \mathbf{a}_k^\top \mathbf{x}, \quad \beta_k = \mathbf{b}_k^\top \mathbf{y}$$

are the k th sample canonical variables,

$$\text{Corr}(U_k, V_k) = \sigma_k$$

is the k th sample canonical correlation for $k = 1, \dots, p$.

- again $\mathbf{x} = [x_1, \dots, x_p]^\top$ and $\mathbf{y} = [y_1, \dots, y_q]^\top$ in the theorem are vectors of dummy variables
- when \mathbf{x} and \mathbf{y} take on actual values in \mathbb{R} coming from an (say, i th) observation of \mathbf{X} and \mathbf{Y} , then we get $\mathbf{x}_i \in \mathbb{R}^p$ and $\mathbf{y}_i \in \mathbb{R}^q$, row vectors of the data matrices $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$
- in this case α_k and β_k becomes actual values in \mathbb{R} (as opposed to a linear combination of dummy variables) and are called *scores*
- the *score vectors* in the context of CCA are

$$X\mathbf{a} \in \mathbb{R}^n \quad \text{and} \quad Y\mathbf{b} \in \mathbb{R}^n$$

the entries $\mathbf{a}^\top \mathbf{x}_1, \dots, \mathbf{a}^\top \mathbf{x}_n$ and $\mathbf{b}^\top \mathbf{y}_1, \dots, \mathbf{b}^\top \mathbf{y}_n$ are scores

- sample CCA could of course also be used with sample correlation matrices in place of sample covariance matrices, i.e.,

$$R_X^{-1/2} R_{XY} R_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

in place of

$$S_X^{-1/2} S_{XY} S_Y^{-1/2} \in \mathbb{R}^{p \times q}$$

in Theorem 3

- Theorem 2 also holds for sample CCA and so does (1.1)
- as in the case of population CCA, $\sigma_1^2 \geq \dots \geq \sigma_p^2$ are the eigenvalues of the matrix

$$G_{XY} G_{XY}^\top = S_X^{-1/2} S_{XY} S_Y^{-1} S_{YX} S_X^{-1/2} \in \mathbb{R}^{p \times p}$$

with corresponding eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$

- $\sigma_1^2 \geq \dots \geq \sigma_p^2$ are also the p largest eigenvalues of the matrix

$$G_{XY}^\top G_{XY} = S_Y^{-1/2} S_{YX} S_X^{-1} S_{XY} S_Y^{-1/2} \in \mathbb{R}^{q \times q}$$

with corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_p$

- sometimes, especially in instances when you need calculate things by hand, it can be more convenient to work with the matrices

$$H_X = S_X^{-1/2} G_{XY} G_{XY}^\top S_X^{1/2} = S_X^{-1} S_{XY} S_Y^{-1} S_{YX} \in \mathbb{R}^{p \times p}$$

and

$$H_Y = S_Y^{-1/2} G_{XY}^\top G_{XY} S_Y^{1/2} = S_Y^{-1} S_{YX} S_X^{-1} S_{XY} \in \mathbb{R}^{q \times q}$$

- the eigenvalues of H_X are exactly those of $G_{XY} G_{XY}^\top$ and the eigenvalues of H_Y are exactly those of $G_{XY}^\top G_{XY}$ since these matrices are related by similarity transformations (recall that the eigenvalues of $B^{-1}AB$ are always the same as the eigenvalues of A)
- hence the sample canonical correlations can be calculated without having to compute square roots of matrices, i.e., from the eigenvalues of

$$S_X^{-1} S_{XY} S_Y^{-1} S_{YX} \in \mathbb{R}^{p \times p}, \quad S_Y^{-1} S_{YX} S_X^{-1} S_{XY} \in \mathbb{R}^{q \times q}$$

- however you should only do this if you're calculating things by hand (like in the first example in the next section)
- as in the case of PCA, when you do sample CCA on a computer, you should always use SVD of G_{XY} and avoid EVD altogether
- note that if you find the sample canonical vectors from the eigenvectors of H_X and H_Y , you would still need to involve square roots of matrices (why?)

- what we say above apply to population CCA as well, i.e., the population canonical correlations are the eigenvalues of the following square-root free expressions

$$\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} \in \mathbb{R}^{p \times p}, \quad \Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} \in \mathbb{R}^{q \times q}$$

- they also apply to using population or sample correlation matrices in place of covariance matrices

3. EXAMPLE: POPULATION CCA

- here is an example where population CCA is useful in yielding insights
- suppose we have $p = q = 2$ and we are interested in population CCA on correlation matrices of a pair of bivariate random vectors \mathbf{X} and \mathbf{Y}
- note that the 2×2 correlation matrices must necessarily have the forms

$$P_X = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \quad P_Y = \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}, \quad P_{XY} = \begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix} = P_{YX}$$

- we will just assume, for non-triviality, that $|\alpha| < 1$ and $|\gamma| < 1$
- in which case $\det(P_X) = 1 - \alpha^2 \neq 0$ and so P_X is nonsingular, likewise for P_Y
- in this example, we will calculate everything by hand, starting with

$$P_X^{-1}P_{XY} = \frac{\beta}{1 - \alpha^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{\beta}{1 + \alpha} \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{2 \times 2}$$

where $\mathbf{1} = [1, 1]^\top \in \mathbb{R}^2$

- likewise

$$P_Y^{-1}P_{YX} = \frac{\beta}{1 + \gamma} \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{2 \times 2}$$

- since $(\mathbf{1}\mathbf{1}^\top)^2 = 2\mathbf{1}\mathbf{1}^\top$, we obtain

$$H_X = P_X^{-1}P_{XY}P_Y^{-1}P_{YX} = \frac{2\beta^2}{(1 + \alpha)(1 + \gamma)} \mathbf{1}\mathbf{1}^\top$$

- since the eigenvalues of the matrix $\mathbf{1}\mathbf{1}^\top$ are 2 and 0 (check it), the eigenvalues of H_X are

$$\frac{4\beta^2}{(1 + \alpha)(1 + \gamma)} \quad \text{and} \quad 0$$

- thus the first canonical correlation coefficient is

$$\rho_1 = \frac{2\beta}{\sqrt{(1 + \alpha)(1 + \gamma)}}$$

4. EXAMPLE: SAMPLE CCA

- this is the famous Frets's head data set (`fretsheads` in R) comprising measurements of head length and head breath for the first and second sons of 25 families
- so we regard the measurements as observations of the random vectors

$$\mathbf{X} = \begin{bmatrix} \text{head length of first son} \\ \text{head breath of first son} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \text{head length of second son} \\ \text{head breath of second son} \end{bmatrix}$$

- so $p = q = 2$, $n = 25$ and with data matrices $X, Y \in \mathbb{R}^{25 \times 2}$ and

$$W = [X, Y] = \begin{matrix} & l_1 & b_1 & l_2 & b_2 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 25 \end{matrix} & \begin{bmatrix} 191 & 155 & 179 & 145 \\ 195 & 149 & 201 & 152 \\ 181 & 148 & 185 & 149 \\ 183 & 153 & 188 & 149 \\ 176 & 144 & 171 & 142 \\ 208 & 157 & 192 & 152 \\ 189 & 150 & 190 & 149 \\ 197 & 159 & 189 & 152 \\ 188 & 152 & 197 & 159 \\ 192 & 150 & 187 & 151 \\ 179 & 158 & 186 & 148 \\ 183 & 147 & 174 & 147 \\ 174 & 150 & 185 & 152 \\ 190 & 159 & 195 & 157 \\ 188 & 151 & 187 & 158 \\ 163 & 137 & 161 & 130 \\ 195 & 155 & 183 & 158 \\ 186 & 153 & 173 & 148 \\ 181 & 145 & 182 & 146 \\ 175 & 140 & 165 & 137 \\ 192 & 154 & 185 & 152 \\ 174 & 143 & 178 & 147 \\ 176 & 139 & 176 & 143 \\ 197 & 167 & 200 & 158 \\ 190 & 163 & 187 & 150 \end{bmatrix} \end{matrix} \in \mathbb{R}^{25 \times 4}$$

where l_1, b_1, l_2, b_2 denote respectively the head length of 1st son, head breath of 1st son, head length of 2nd son, head breath of 2nd son; the rows are indexed by the family number (arbitrary)

- the sample correlation matrices are

$$R_X = \begin{bmatrix} 1 & 0.7346 \\ 0.7346 & 1 \end{bmatrix}, \quad R_Y = \begin{bmatrix} 1 & 0.8392 \\ 0.8392 & 1 \end{bmatrix}, \quad R_{XY} = \begin{bmatrix} 0.7108 & 0.7040 \\ 0.6932 & 0.7086 \end{bmatrix} = R_{YX}^T$$

- so

$$G_{XY} = R_X^{-1/2} R_{XY} R_Y^{-1/2} = \begin{bmatrix} 0.4217 & 0.3761 \\ 0.3586 & 0.4206 \end{bmatrix}$$

and its singular values are

$$r_1 = 0.7886, \quad r_2 = 0.0539$$

- if you want to do this by hand, you can compute the eigenvalues of

$$H_X = R_X^{-1} R_{XY} R_Y^{-1} R_{YX} \in \mathbb{R}^{2 \times 2},$$

which are 0.6218 and 0.0029, taking square roots then gives you r_1 and r_2

- the left and right singular vectors of G_{XY} are

$$\mathbf{u}_1 = \begin{bmatrix} 0.7149 \\ 0.6992 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0.6992 \\ -0.7149 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 0.7003 \\ 0.7138 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0.7138 \\ 0.7003 \end{bmatrix}$$

- so the first pair of sample canonical vectors are

$$\mathbf{a}_1 = R_X^{-1/2} \mathbf{u}_1 = \begin{bmatrix} 0.552 \\ 0.522 \end{bmatrix}, \quad \mathbf{b}_1 = R_Y^{-1/2} \mathbf{v}_1 = \begin{bmatrix} 0.505 \\ 0.538 \end{bmatrix}$$

- the second pair of sample canonical vectors are

$$\mathbf{a}_2 = R_X^{-1/2} \mathbf{u}_2 = \begin{bmatrix} 1.367 \\ -1.378 \end{bmatrix}, \quad \mathbf{b}_2 = R_Y^{-1/2} \mathbf{v}_2 = \begin{bmatrix} 1.767 \\ -1.757 \end{bmatrix}$$

- equivalently we may express these as pairs of sample canonical variables and interpret them as follows:

$$\alpha_1 = \mathbf{a}_1^T \mathbf{x} = 0.552x_1 + 0.522x_2$$

is roughly the sum¹ of head length and breadth of the first brother since $0.552 \approx 0.522$

$$\beta_1 = \mathbf{b}_1^T \mathbf{y} = 0.505y_1 + 0.538y_2$$

is roughly the sum of head length and breadth of the second brother since $0.505 \approx 0.538$

$$\alpha_2 = \mathbf{a}_2^T \mathbf{x} = 1.367x_1 - 1.378x_2$$

is roughly the difference² of head length and breadth of the first brother since $1.367 \approx 1.378$

$$\beta_2 = \mathbf{b}_2^T \mathbf{y} = 1.767y_1 - 1.757y_2$$

is roughly the difference of head length and breadth of the second brother since $1.767 \approx 1.757$

- we will call the sum of length and breadth *girth* and the difference of length and breadth *shape*
- since $r_1 = 0.7886 \approx 80\%$, the head girths of first and second brothers are highly correlated
- since $r_2 = 0.0539 \approx 5\%$, the head shapes of first and second brothers has little correlation

5. REAL-WORLD EXAMPLE FROM LUIS MOREIRA DA COSTA

- this is the example Luis mentioned in class that came from his work (many thanks to him for sharing it)

We wanted to calculate a proxy for the commodities part of Chinese GDP - a Chinese commodities GDP, if you will. We tried to do this using CCA. On one side, we gathered a variety of Chinese economic activity numbers related to commodities (industrial production, some surveys, construction figures, etc...). This would be our X_1, X_2, \dots, X_n . On the other side, we had the prices of five commodities. This was our Y_1, Y_2, \dots, Y_n . The variables in both X and Y are time-series, naturally. They were both monthly in frequency, and were log-differenced to ensure stationarity. We then ran a CCA on these two sets. The first canonical covariate related to X (U_1) was our proxy.

- very nice and not as surprising as it first appears — my favorite example along the same lines is `garbage.pdf` in the handouts folder

¹Strictly speaking a constant multiple of the sum.

²Strictly speaking a constant multiple of the difference.