Canonical Correlation Analysis (CCA)

- Recall: Regression analysis --- concerned with the relationship between a *single* response and a set of predictors

$$\underset{(n\times p)}{\mathbf{X}} = \left[\begin{array}{c|cccc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{array}\right]$$

  - $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$
  - connection between regression model and conditional normal distribution
  - $R^2 = cor(Y, \hat{Y})^2$

- **Q**: What if we have *more than one* response?

$$\underset{(n\times p)}{\mathbf{X}} = \left[\begin{array}{cc|cc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{array}\right]$$

  - Examples
    - variables related to arithmetic power & variables related to reading power
    - variables related to governmental policy & variables related to economic goal
    - variables related to college performance & variables related to precollege achievement
  - CCA seeks to identify and quantify the linear associations between *two sets* of variables

- Population Canonical Variables and Canonical Correlations
  - Data: two groups of variables

    The first group, of p variables, is represented by the $(p \times 1)$ random vector $\mathbf{X}^{(1)}$. The second group, of q variables, is represented by the $(q \times 1)$ random vector $\mathbf{X}^{(2)}$. We assume, in the theoretical development, that $\mathbf{X}^{(1)}$ represents the *smaller* set, so that $p \leq q$.

  - Some notations
    - $$\underset{((p+q)\times 1)}{\mathbf{X}} = \left[\begin{array}{c} \mathbf{X}^{(1)} \\ \hline \mathbf{X}^{(2)} \end{array}\right] = \left[\begin{array}{c} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{array}\right]$$
    - $$\underset{((p+q)\times 1)}{\boldsymbol{\mu}} = E(\mathbf{X}) = \left[\begin{array}{c} E(\mathbf{X}^{(1)}) \\ \hline E(\mathbf{X}^{(2)}) \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hline \boldsymbol{\mu}^{(2)} \end{array}\right]$$
    - $$\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \left[\begin{array}{c|c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$
  - the pq elements of $\boldsymbol{\Sigma}_{12}$ measure the association between the two sets.
    - p and q are relatively large, hopeless to interpret elements of Σ_{12} collectively
    - moreover, it's often linear combinations of variables that are interesting and useful for predictive or comparative purposes

➢ CCA finds linear functions of one set of variables that maximally correlated with linear function of the other set of variables

- Set $U = \mathbf{a}'\mathbf{X}^{(1)}$ and $V = \mathbf{b}'\mathbf{X}^{(2)}$ for some pair of coefficient vectors $\mathbf{a}$ and $\mathbf{b}$.
- $$\mathrm{Var}(U) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)})\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}$$
$$\mathrm{Var}(V) = \mathbf{b}'\,\mathrm{Cov}(\mathbf{X}^{(2)})\mathbf{b} = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}$$
$$\mathrm{Cov}(U, V) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}$$
- We shall seek coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ such that
$$\mathrm{Corr}(U, V) = \frac{\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\,\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}}} \qquad (❖)$$
is as large as possible.
- The *first pair of canonical variables*, or *first canonical variate pair*, is the pair of linear combinations U_1, V_1 having unit variances, which maximize the correlation (❖)
- The *second pair of canonical variables*, or *second canonical variate pair*, is the pair of linear combinations U_2, V_2 having unit variances, which maximize the correlation (❖) among all choices that are uncorrelated with the first pair of canonical variables.
- The *kth pair of canonical variables*, or *kth canonical variate pair*, is the pair of linear combinations U_k, V_k having unit variances, which maximize the correlation (❖) among all choices uncorrelated with the previous $k - 1$ canonical variable pairs.

- **Result 10.1.** Suppose $p \le q$ and let the random vectors $\underset{(p\times 1)}{\mathbf{X}^{(1)}}$ and $\underset{(q\times 1)}{\mathbf{X}^{(2)}}$ have $\mathrm{Cov}(\mathbf{X}^{(1)}) = \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}}$, $\mathrm{Cov}(\mathbf{X}^{(2)}) = \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}}$ and $\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}}$, where $\boldsymbol{\Sigma}$ has full rank. For coefficient vectors $\underset{(p\times 1)}{\mathbf{a}}$ and $\underset{(q\times 1)}{\mathbf{b}}$, form the linear combinations $U = \mathbf{a}'\mathbf{X}^{(1)}$ and $V = \mathbf{b}'\mathbf{X}^{(2)}$. Then
$$\max_{\mathbf{a},\mathbf{b}} \mathrm{Corr}(U, V) = \rho_1^*$$
attained by the linear combinations (first canonical variate pair)
$$U_1 = \underbrace{\mathbf{e}_1'\boldsymbol{\Sigma}_{11}^{-1/2}}_{\mathbf{a}_1'}\mathbf{X}^{(1)} \quad \text{and} \quad V_1 = \underbrace{\mathbf{f}_1'\boldsymbol{\Sigma}_{22}^{-1/2}}_{\mathbf{b}_1'}\mathbf{X}^{(2)}$$
The kth pair of canonical variates, $k = 2, 3, \ldots, p$,
$$U_k = \mathbf{e}_k'\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{X}^{(1)} \qquad V_k = \mathbf{f}_k'\boldsymbol{\Sigma}_{22}^{-1/2}\mathbf{X}^{(2)}$$
maximizes
$$\mathrm{Corr}(U_k, V_k) = \rho_k^*$$
among those linear combinations uncorrelated with the preceding $1, 2, \ldots, k-1$ canonical variables.

Here $\rho_1^{*2} \ge \rho_2^{*2} \ge \cdots \ge \rho_p^{*2}$ are the eigenvalues of $\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}$, and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ are the associated $(p \times 1)$ eigenvectors. [The quantities $\rho_1^{*2}, \rho_2^{*2}, \ldots, \rho_p^{*2}$ are also the p largest eigenvalues of the matrix $\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}$ with corresponding $(q \times 1)$ eigenvectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_p$. Each $\mathbf{f}_i$ is proportional to $\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_i$.]

- **Q**: why $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$ (or $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$)?

- geometrical interpretation

Let $\underset{(p\times p)}{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p]'$ and $\underset{(q\times q)}{\mathbf{B}} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q]'$, so that the vectors of canonical variables are

$$\underset{(p\times 1)}{\mathbf{U}} = \mathbf{A}\mathbf{X}^{(1)} \qquad \underset{(q\times 1)}{\mathbf{V}} = \mathbf{B}\mathbf{X}^{(2)}$$

$\mathbf{A} = \mathbf{E}'\Sigma_{11}^{-1/2} = \mathbf{E}'\mathbf{P}_1\Lambda_1^{-1/2}\mathbf{P}_1'$ where $\mathbf{E}'$ is an orthogonal matrix with row $\mathbf{e}_i'$

- The canonical variates have the properties

$$\begin{aligned}
\text{Var}(U_k) &= \text{Var}(V_k) = 1 \\
\text{Cov}(U_k, U_\ell) &= \text{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell \\
\text{Cov}(V_k, V_\ell) &= \text{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell \\
\text{Cov}(U_k, V_\ell) &= \text{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell
\end{aligned}$$

for $k, \ell = 1, 2, \ldots, p$.

- $\text{Cov}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) = \mathbf{A}\Sigma_{11}$

$$\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \text{Corr}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{U}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}) = \text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)})$$
$$= \mathbf{A}\Sigma_{11}\mathbf{V}_{11}^{-1/2}$$

Similar calculations for the pairs $(\mathbf{U}, \mathbf{X}^{(2)})$, $(\mathbf{V}, \mathbf{X}^{(2)})$ and $(\mathbf{V}, \mathbf{X}^{(1)})$ yield

$$\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \mathbf{A}\Sigma_{11}\mathbf{V}_{11}^{-1/2} \qquad \underset{(q\times q)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(2)}}} = \mathbf{B}\Sigma_{22}\mathbf{V}_{22}^{-1/2}$$

$$\underset{(p\times q)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(2)}}} = \mathbf{A}\Sigma_{12}\mathbf{V}_{22}^{-1/2} \qquad \underset{(q\times p)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(1)}}} = \mathbf{B}\Sigma_{21}\mathbf{V}_{11}^{-1/2}$$

- To ease the computation burden, many people prefer to get the canonical correlations from

$$|\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \rho^{*2}\mathbf{I}| = 0$$

The coefficient vectors **a** and **b** follow directly from the eigenvector equations

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{a} = \rho^{*2}\mathbf{a}$$

$$\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{b} = \rho^{*2}\mathbf{b}$$

➢ Interpreting the population canonical variables

- identify the canonical variables
  - the linear combinations can be interpreted much as in principal components
  - it is helpful to compute the correlation between the canonical variables and the original variables as way to determine the relative important of original variables to a particular canonical variable
- canonical correlation
  - canonical correlation generalizes the correlation between 2 variables
  - $|\text{Corr}(X_i^{(1)}, X_k^{(2)})| = |\text{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)})| \leq \max_{\mathbf{a},\mathbf{b}} \text{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \rho_1^*$
  - An R^2 type statistic can be obtained to explain the total variance explained by a given set of canonical variables:

Because of its multiple correlation coefficient interpretation, the kth *squared* canonical correlation ρ_k^{*2} is the proportion of the variance of canonical variate U_k "explained" by the set $\mathbf{X}^{(2)}$. It is also the proportion of the variance of canonical variate V_k "explained" by the set $\mathbf{X}^{(1)}$.

- first r canonical variables as a summary of variability
  - remember that the linear combinations are chosen to maximize correlation between the canonical variables
  - these linear combinations may be quite different from those obtained from principal component within a particular set of variables
  - the canonical may not necessarily explain the total variation within a set of original variables

➢ CCA for standardized data

- If the original variables are standardized with $\mathbf{Z}^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_p^{(1)}]'$ and $\mathbf{Z}^{(2)} = [Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_q^{(2)}]'$, from first principles, the canonical variates are of the form

$$U_k = \mathbf{a}_k'\mathbf{Z}^{(1)} = \mathbf{e}_k'\boldsymbol{\rho}_{11}^{-1/2}\mathbf{Z}^{(1)}$$

$$V_k = \mathbf{b}_k'\mathbf{Z}^{(2)} = \mathbf{f}_k'\boldsymbol{\rho}_{22}^{-1/2}\mathbf{Z}^{(2)}$$

Here, $\text{Cov}(\mathbf{Z}^{(1)}) = \boldsymbol{\rho}_{11}$, $\text{Cov}(\mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{22}$, $\text{Cov}(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{12} = \boldsymbol{\rho}_{21}'$, and $\mathbf{e}_k$ and $\mathbf{f}_k$ are the eigenvectors of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ and $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$, respectively. The canonical correlations, ρ_k^*, satisfy

$$\text{Corr}(U_k, V_k) = \rho_k^*, \qquad k = 1, 2, \ldots, p$$

where $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ are the nonzero eigenvalues of the matrix $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ (or, equivalently, the largest eigenvalues of $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$).

- the canonical correlations are unchanged by the standardization (cf. PCA)

❖ **Reading**: Reference, 10.1, 10.2, 10.3