Canonical Correlation Analysis (CCA)

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

Q: how to add X_1^2?

$$\underset{(n\times p)}{\mathbf{X}} = \begin{array}{c} \begin{array}{cccc} Y & X_1 & \cdots & X_k \end{array} \\ \left[\begin{array}{c|ccc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{array}\right] \end{array}$$

  - $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$
  - connection between regression model and conditional normal distribution → LN p.3-14~16.
  - $R^2 = cor(Y, \hat{Y})^2$ $\quad X\hat{\beta}$: a linear combination of $X_1, \cdots, X_k$.
    ↳ proportion of variation in Y explained by the variation of predictors.

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

$$\underset{(n\times p)}{\mathbf{X}} = \begin{array}{c} \begin{array}{cccc} Y_1 & Y_2 & X_1 \cdots & X_k \end{array} \\ \left[\begin{array}{cc|cc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{array}\right] \end{array}$$

① multivariate regression.
$Y_1 = \beta_{10} + \beta_{11}X_1 + \cdots + \beta_{1k}X_k + \varepsilon$
$Y_2 = \beta_{20} + \beta_{21}X_1 + \cdots + \beta_{2k}X_k + \varepsilon$
② apply PCA on $Y_1, \cdots, Y_p$ → $PC_1, PC_2, \cdots$
$PC_1 = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$.
③ CCA.
- - -

  - 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 $p \le q$.

  - Some notations
    - $$\underset{((p+q)\times 1)}{\mathbf{X}} = \left[\begin{array}{c} \mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)} \end{array}\right] = \left[\begin{array}{c} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hdashline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{array}\right]$$
    - $$\underset{((p+q)\times 1)}{\boldsymbol{\mu}} = E(\mathbf{X}) = \left[\begin{array}{c} E(\mathbf{X}^{(1)}) \\ \hdashline E(\mathbf{X}^{(2)}) \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{array}\right]$$
    - $$\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \left[\begin{array}{c:c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

information about association between 1st set of variables.

association between 2nd set of variables

  - 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 **a** and **b**.
- $$\mathrm{Var}(U) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)})\,\mathbf{a} = \mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}$$
  $$\mathrm{Var}(V) = \mathbf{b}'\,\mathrm{Cov}(\mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}$$
  $$\mathrm{Cov}(U,V) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}$$

CCA	PC
2 sets	1 set
2 projection	one projection
max cor.	max var.

- We shall seek coefficient vectors **a** and **b** such that

$$\mathrm{Corr}(U,V) = \frac{\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}}\,\sqrt{\mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}}} \qquad (❖)$$

cor²

canonical correlation 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 (❖)

$\mathrm{cor}(U_1,V_1)$

- 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.

$\mathrm{cor}(U_2,V_2)$

$0 = \mathrm{cor}(U_2,U_1) = \mathrm{cor}(U_2,V_1) = \mathrm{cor}(V_2,U_1) = \mathrm{cor}(V_2,V_1)$

- 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.

$\mathrm{cor}(U_k,V_k)$

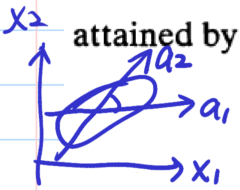
- **Result 10.1.** Suppose $p \le q$ and let the random vectors $\underset{(p\times 1)}{\mathbf{X}^{(1)}}$ and $\underset{(q\times 1)}{\mathbf{X}^{(2)}}$ have $\mathrm{Cov}(\mathbf{X}^{(1)}) = \underset{(p\times p)}{\mathbf{\Sigma}_{11}}$, $\mathrm{Cov}(\mathbf{X}^{(2)}) = \underset{(q\times q)}{\mathbf{\Sigma}_{22}}$ and $\mathrm{Cov}(\mathbf{X}^{(1)},\mathbf{X}^{(2)}) = \underset{(p\times q)}{\mathbf{\Sigma}_{12}}$, where $\mathbf{\Sigma}$ 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)

$$U_1 = \underbrace{\mathbf{e}_1'\mathbf{\Sigma}_{11}^{-1/2}}_{\mathbf{a}_1'}\mathbf{X}^{(1)} \quad \text{and} \quad V_1 = \underbrace{\mathbf{f}_1'\mathbf{\Sigma}_{22}^{-1/2}}_{\mathbf{b}_1'}\mathbf{X}^{(2)}$$

standardizing.



Q: $e_i \perp e_j$
$\Rightarrow e_i^T e_j = 0$
$f_i \perp f_j$
$\Rightarrow f_i^T f_j = 0$

$a_i \not\perp a_j$?
$a_i^T a_j$
$= e_i^T \Sigma_{11}^{-1/2}\Sigma_{11}^{-1/2} e_j$
$= e_i^T \Sigma_{11}^{-1} e_j \neq 0$
$b_i \not\perp b_j$

The kth pair of canonical variates, $k = 2, 3, \ldots, p$,

$$U_k = \underbrace{\mathbf{e}_k'\mathbf{\Sigma}_{11}^{-1/2}}_{a_k'}\mathbf{X}^{(1)} \qquad V_k = \underbrace{\mathbf{f}_k'\mathbf{\Sigma}_{22}^{-1/2}}_{b_k'}\mathbf{X}^{(2)}$$

maximizes

$$\mathrm{Corr}(U_k, V_k) = \rho_k^*$$

among those linear combinations uncorrelated with the preceding $1, 2, \ldots, k-1$ canonical variables.

symmetric

exercise

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

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

* max u^Tx subject to $x^TQx=1$ (Q symmetric) is equivalent to maximizing $\frac{u^Tx}{\sqrt{x^TQx}}$ for all x.

$$\frac{u^Tx}{\sqrt{x^TQx}} = \frac{u^TQ^{-1/2}Q^{1/2}x}{\sqrt{(Q^{1/2}x)^TQ^{1/2}x}} \equiv v,\quad v=(u^TQ^{-1/2})^T = Q^{-1/2}u$$

x, maximum value $=\sqrt{u^TQ^{-1}u}$, $x = Q^{-1/2}v = Q^{-1}u$

* max cor(U,V) over a with b fixed.

$$\text{max. cor} = \frac{a^T\Sigma_{12}b = b^T\Sigma_{21}a}{\sqrt{a^T\Sigma_{11}a}\sqrt{b^T\Sigma_{22}b}},\quad a = Q^{-1}u = \Sigma_{11}^{-1}\Sigma_{12}b,$$

maximum value

$$\frac{b^T\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}b}{\sqrt{b^T\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12}b}\sqrt{b^T\Sigma_{22}b}} = \left(\frac{b^T\Sigma_{22}^{1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{1/2}b}{\underbrace{b^T\Sigma_{22}b}_{(\Sigma_{22}^{1/2}b)^T(\Sigma_{22}^{1/2}b)}}\right)^{1/2}$$

$$= \left(\frac{b^{*T}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}b^*}{b^{*T}b^*}\right)^{1/2},\quad \underbrace{\Sigma_{22}^{1/2}b}_{b^*}$$

c.f. PCA

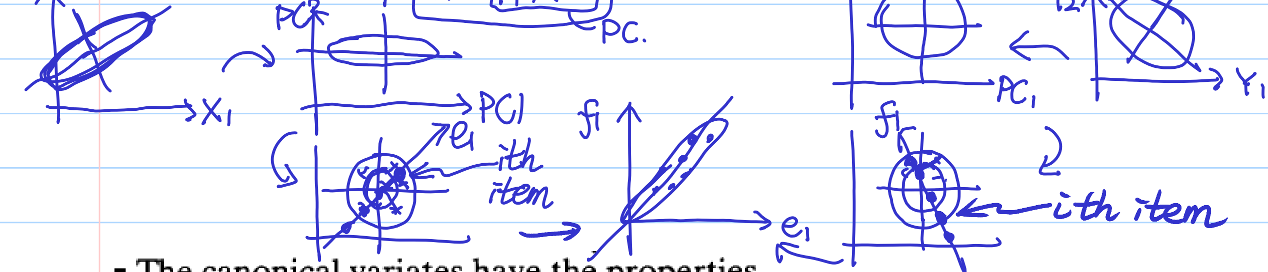
- geometrical interpretation

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

$$\underset{(p\times 1)}{\mathbf{U}} = \mathbf{A}\mathbf{X}^{(1)} \qquad \underset{(q\times 1)}{\mathbf{V}} = \mathbf{B}\mathbf{X}^{(2)} \qquad E=[e_1|e_2|\cdots]$$

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

$$U = AX^{(1)} = E'\underbrace{P_1\Lambda_1^{-1/2}P_1'X^{(1)}}_{PC.}$$



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

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

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

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

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

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