STA437/2005 - Methods for Multivariate Data

Lecture 9

Gun Ho Jang

November 24, 2014

Canonical Correlation Analysis

- Canonical correlation is concerned about the relationship between two data sets.
- As usual, linear relationship is discussed as long as most of the relationship is preserved.
- Similar to principal component analysis, the correlation measures the amount of relationship.

Canonical Correlation Analysis I

- Let $\mathbf{Y}$ and $\mathbf{Z}$ be two random vectors with dimension $p \leq q$ respectively.
- Define $\mathbf{X} = (\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$.
- Then the mean and variance of $\mathbf{X}$ are $\mathbb{E}(\mathbf{X}) = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ and $\mathbb{V}ar(\mathbf{X}) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.
- Let U and V be $U = \mathbf{a}^\top \mathbf{Y}$ and $V = \mathbf{b}^\top \mathbf{Z}$. Then

$$
\begin{aligned}
\mathbb{V}ar(U) &= \mathbb{V}ar(\mathbf{a}^\top \mathbf{Y}) = \mathbf{a}^\top \mathbb{V}ar(\mathbf{Y})\mathbf{a} = \mathbf{a}^\top \Sigma_{11} \mathbf{a} \\
\mathbb{V}ar(V) &= \mathbb{V}ar(\mathbf{b}^\top \mathbf{Z}) = \mathbf{b}^\top \mathbb{V}ar(\mathbf{Z})\mathbf{b} = \mathbf{b}^\top \Sigma_{22} \mathbf{b} \\
\mathrm{Cov}(U, V) &= \mathrm{Cov}(\mathbf{a}^\top \mathbf{Y}, \mathbf{b}^\top \mathbf{Z}) = \mathbf{a}^\top \mathrm{Cov}(\mathbf{Y}, \mathbf{Z})\mathbf{b} = \mathbf{a}^\top \Sigma_{12} \mathbf{b} \\
\mathrm{Cor}(U, V) &= \frac{\mathrm{Cov}(U, V)}{\sqrt{\mathbb{V}ar(U)\mathbb{V}ar(V)}} = \frac{\mathbf{a}^\top \Sigma_{12} \mathbf{b}}{\sqrt{\mathbf{a}^\top \Sigma_{11} \mathbf{a} \times \mathbf{b}^\top \Sigma_{22} \mathbf{b}}}
\end{aligned}
$$

Canonical Correlation Analysis II

- Define the *first canonical variate pair* is $U_1 = \mathbf{a}_1^\top \mathbf{Y}$ and $V_1 = \mathbf{b}_1^\top \mathbf{Z}$ maximizing the correlation $\mathrm{Cor}(U_1, V_1)$ subject to unit variances.
- The *second canonical variate pair* is $U_2 = \mathbf{a}_2^\top \mathbf{Y}$ and $V_2 = \mathbf{b}_2^\top \mathbf{Z}$ having unit variances which maximize correlation $\mathrm{Cor}(U_2, V_2)$ among all choices that are uncorrelated with the first canonical variate pair.
- Sequentially, the kth canonical variate pair is $U_k = \mathbf{a}_k^\top \mathbf{Y}$ and $V_k = \mathbf{b}_k^\top \mathbf{Z}$ having unit variances which maximize correlation $\mathrm{Cor}(U_k, V_k)$ among all choices that are uncorrelated with the previous $k-1$ canonical variate pairs.

Remark

Such canonical variate pairs can be found using the singular value decomposition which is generalized version of the spectral decomposition.

Canonical Correlation Analysis III

Theorem (Singular value decomposition)

Let A be a $p \times q$ matrix with $p \leq q$. Then there exist orthonormal matrix $U \in \mathbb{R}^{p\times p}$, $V \in \mathbb{R}^{q\times q}$ and a non-negative valued matrix $\Lambda \in \mathbb{R}^{p\times q}$ with $\Lambda_{ij} = 0$ for $i \neq j$ so that

$$A = U\Lambda V^\top = U_1\Lambda_{11}V_1^\top + \cdots + U_p\Lambda_{pp}V_p^\top$$

where $U = (U_1, \ldots, U_p)$ and $V = (V_1, \ldots, V_q)$

Canonical Correlation Analysis IV

Proof.

The spectral decomposition theorem for $M^\top M$ gives $M^\top M = V \begin{pmatrix} D & O \\ O & O \end{pmatrix} V^\top$ where $D \in \text{ı} \times \text{ı}$ is a non-negative diagonal matrix. Let $V = (V_1 \ V_2)$ with $V_1 \in \mathbb{R}^{q \times p}$. The orthonormality gives $V_1^\top V_1 = I_p$, $V_2^\top V_2 = I_{q-p}$ and $V_1 V_a^\top + V_2 V_2^\top = I_q$. Then $V^\top M^\top M V = \begin{pmatrix} D & O \\ O & O \end{pmatrix}$ gives $V_1^\top M^\top M V_1 = D$ and $V_2^\top M^\top M V_2 = O$.

Let $U = M V_1 D^{-1/2}$ and $\Lambda = (D^{1/2} \ O)$ so that $U^\top U = D^{-1/2} V_1^\top M^\top M V_1 D^{-1/2} = D^{-1/2} D D^{-1/2} = I_p$. Then

$$\begin{aligned} U \Lambda V^\top &= U (D^{1/2} \ O) \begin{pmatrix} V_1^\top \\ V_2^\top \end{pmatrix} = U (D^{1/2} V_1^\top + O V_2^\top) = U D^{1/2} V_1^\top \\ &= M V_1 D^{-1/2} D^{1/2} V_1^\top = M V_1 V_1^\top = M. \end{aligned}$$

$\square$

Canonical Correlation Analysis I

Let $\widetilde{\mathbf{a}} = \Sigma_{11}^{1/2}\mathbf{a}$ and $\widetilde{\mathbf{b}} = \Sigma_{22}^{1/2}\mathbf{b}$ so that

$$\text{Cor}(U, V) = \frac{\text{Cov}(\mathbf{a}^\top \mathbf{Y}, \mathbf{b}^\top \mathbf{Z})}{\sqrt{\mathbb{V}ar(\mathbf{a}^\top \mathbf{Y})\mathbb{V}ar(\mathbf{b}^\top \mathbf{Z})}} = \frac{\tilde{\mathbf{a}}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\widetilde{\mathbf{b}}}{\sqrt{\widetilde{\mathbf{a}}^\top \widetilde{\mathbf{a}}\widetilde{\mathbf{b}}^\top \widetilde{\mathbf{b}}}}$$

$$= \frac{\widetilde{\mathbf{a}}}{||\tilde{\mathbf{a}}||}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\frac{\widetilde{\mathbf{b}}}{||\widetilde{\mathbf{b}}||}$$

Using the singular value decomposition, let $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} = C(D\ O)E^\top$. Then

$$= \frac{1}{||\widetilde{\mathbf{a}}|| \cdot ||\widetilde{\mathbf{b}}||}\widetilde{\mathbf{a}}^\top C(D\ O)E^\top \widetilde{\mathbf{b}} = \widehat{\mathbf{a}}^\top D\widehat{\mathbf{b}} = \sum_{j=1}^{p} \widehat{\mathbf{a}}_j D_{jj}\widehat{\mathbf{b}}_j$$

where $\widehat{\mathbf{a}} = C^\top \widetilde{\mathbf{a}}/||\widetilde{\mathbf{a}}||$ and $\widehat{\mathbf{b}} = (I_p\ O)E^\top \widetilde{\mathbf{b}}/||\widetilde{\mathbf{b}}||$. then using Cauchy-Schwartz inequality,

$$\leq \max D_{jj}$$

Canonical Correlation Analysis I

Proposition

- *Let* $\mathbb{V}ar(\mathbf{Y}) = \Sigma_{11}$, $\mathbb{V}ar(\mathbf{Z}) = \Sigma_{22}$ *and* $Cov(\mathbf{Y}, \mathbf{Z}) = \Sigma_{12}$.
- *The canonical correlation pairs are* $(U_1, V_1), \ldots, (U_p, V_p)$ *with corresponding correlations* $\rho_1^*, \ldots, \rho_p^*$.
- *Then the correlations* $\rho_1^*, \ldots, \rho_p^*$ *are the diagonal elements of* Λ *which is in the singular value decomposition of* $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} = C\Lambda D^\top$.
- *Besides* $\mathbf{U} = (U_1, \ldots, U_p)^\top = C^\top\Sigma_{11}^{-1/2}\mathbf{Y}$ *and* $\mathbf{V} = (V_1, \ldots, V_p) = (I_p \ O)D^\top\Sigma_{22}^{-1/2}\mathbf{Z}$

Canonical Correlation Analysis II

Proof.

- Note that $\mathbb{V}ar(\mathbf{U}) = \mathbb{V}ar(C^\top \Sigma_{11}^{-1/2}\mathbf{Y}) = C^\top \Sigma_{11}^\top \mathbb{V}ar(\mathbf{Y})\Sigma_{11}^{-1/2} C = C^\top \Sigma_{11}^{-1/2}\Sigma_{11}\Sigma_{11}^{-1/2} C = C^\top C = I_p$
- and $\mathbb{V}ar(\mathbf{V}) = (I_p\ O)D^\top \Sigma_{22}^{-1/2}\mathbb{V}ar(\mathbf{Z})\Sigma_{22}^{-1/2} D(I_p\ O)^\top = (I_p\ O)I_q(I_p\ O)^\top = I_p$.
- Hence $U_1, \ldots, U_p$ are uncorrelated and $V_1, \ldots, V_p$ are uncorrelated too.
- Then $\mathrm{Cor}(\mathbf{U}, \mathbf{V}) = \mathrm{Cov}(\mathbf{U}, \mathbf{V}) = \mathrm{Cov}(C^\top \Sigma_{11}^{-1/2}\mathbf{Y}, (I_p\ O)D^\top \Sigma_{22}^{-1/2}\mathbf{Z}) = C^\top \Sigma_{11}\mathrm{Cov}(\mathbf{Y}, \mathbf{Z})\Sigma_{22}^{-1/2} D(I_p\ O)^\top = C^\top [C\Lambda D^\top] D(I_p\ O)^\top = \Lambda(I_p\ O)^\top = \mathrm{diag}(\Lambda_{11}, \ldots, \Lambda_{pp})$.
- Hence the theorem follows.

$\square$

Canonical Correlation Analysis III

Proposition

In the previous proposition, $(\rho_j^*)^2$ are the eigen values of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$. Also $C_1, \dots, C_p$ are corresponding eigen vectors where $C = (C_1, \dots, C_p)$.

Proof.

- Note that

$$A = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2} = (C\Lambda D^\top)(C\Lambda D^\top)^\top$$
$$= C\Lambda D^\top D\Lambda^\top C^\top = C\Lambda\Lambda^\top C^\top = C\text{diag}(\Lambda_{11}^2, \dots, \Lambda_{pp}^2)C^\top.$$

- Hence the spectral decomposition implies $(\rho_j^*)^2 = \Lambda_{jj}^2$ is an eigen value with eigen vector C_j for each j.

□

Canonical Correlation Analysis IV

Proposition

In the previous proposition, $(\rho_j^*)^2$ are the p largest eigen values of $B = \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$. Also $D_1, \ldots, D_p$ are corresponding eigen vectors where $D = (D_1, \ldots, D_p)$.

Exercise

Prove the above proposition.

Exercise

Show that the symmetric matrix B in the previous proposition has at least $q - p$ zero eigen values.

Large Sample Property I

Suppose $\mathbf{X}_i = (\mathbf{Y}_i^\top, \mathbf{Z}_i^\top)^\top \sim i.i.d.\ N_{p+q}(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \ldots, n$.

Exercise

Show that $|\Sigma| = |\Sigma_{11}| \cdot |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|$.

Under the assumption $H : \Sigma_{12} = O$, the maximum likelihood estimators of $\boldsymbol{\mu}, \Sigma_{11}, \Sigma_{22}$ are $\bar{\mathbf{x}} = (\bar{\mathbf{y}}^\top, \bar{\mathbf{z}}^\top)^\top$, $(1 - 1/n)S_{11}$ and $(1 - 1/n)S_{22}$. Then the maximum likelihood is

$$
\begin{aligned}
|2\pi\widehat{\Sigma}|^{-n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})^\top \widehat{\Sigma}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})) &= |2\pi\widehat{\Sigma}|^{-n/2} \exp(-n(p+q)/2) \\
&= (2\pi)^{-n(p+q)/2}|(1 - 1/n)S_{11}|^{-n/2}|(1 - 1/n)S_{22}|^{-n/2} \exp(-n(p+q)/2) \\
&= (2\pi(1 - 1/n))^{-n(p+q)/2}|S_{11}|^{-n/2}|S_{22}|^{-n/2} \exp(-n(p+q)/2).
\end{aligned}
$$

Large Sample Property II

Without any restriction, the maximum likelihood estimator of Σ is $(1 - 1/n)S$ along with the maximum likelihood

$$
\begin{aligned}
|2\pi\widehat{\Sigma}|^{-n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})^\top \widehat{\Sigma}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})) &= |2\pi\widehat{\Sigma}|^{-n/2} \exp(-n(p+q)/2) \\
&= (2\pi(1 - 1/n))^{-n(p+q)/2}|S|^{-n/2} \exp(-n(p+q)/2).
\end{aligned}
$$

Then the likelihood ratio becomes

$$
\begin{aligned}
\Lambda &= \frac{(2\pi(1 - 1/n))^{-n(p+q)/2}|S_{11}|^{-n/2}|S_{22}|^{-n/2} \exp(-n(p+q)/2)}{(2\pi(1 - 1/n))^{-n(p+q)/2}|S|^{-n/2} \exp(-n(p+q)/2)} \\
&= \Big(\frac{|S_{11}| \cdot |S_{22}|}{|S|}\Big)^{-n/2}.
\end{aligned}
$$

Large Sample Property III

The log likelihood ratio statistic for a hypothesis $H : \Sigma_{12} = O$ is

$$
\begin{aligned}
-2\log\Lambda &= n\log\frac{|S_{11}|\cdot|S_{22}|}{|S|} = n\log\frac{|S_{11}|\cdot|S_{22}|}{|S_{22}|\cdot|S_{11}-S_{12}S_{22}^{-1}S_{21}|} \\
&= -n\log|S_{11}|^{-1}|S_{11}-S_{12}S_{22}^{-1}S_{21}| = -n\log|I_p - S_{11}^{-1/2}S_{12}S_{22}^{-1}S_{21}S_{11}^{-1/2}| \\
&= -n\log|CC^\top - C\mathrm{diag}((\widehat{\rho}_j^*))C^\top| = -n\log|I_p - \mathrm{diag}((\widehat{\rho}_j^*))| \\
&= -n\log\prod_{j=1}^{p}(1-(\widehat{\rho}_j^*)^2)
\end{aligned}
$$

which converges to $\chi^2(pq)$ as $n \to \infty$.
Hence the independence of two group can be assessed using the likelihood ratio statistic.

Large Sample Property IV

In general, Bartlette proposed an assessment for $H : \rho_1^* \geq \rho_2^* \geq \cdots \geq \rho_k^* > 0 = \rho_{k+1}^* = \cdots = \rho_p^*$ based on

$$-(n - 1 - (p + q + 1)/2) \log \prod_{j=k+1}^{p} (1 - (\widehat{\rho}_j^*)^2) \rightarrow \chi^2((p - k)(q - k)).$$

Job Satisfaction Example I

The correlation matrix R is given by

$$
\left(\begin{array}{ccccc|ccccccc}
1.00 & 0.49 & 0.53 & 0.49 & 0.51 & 0.33 & 0.32 & 0.20 & 0.19 & 0.30 & 0.37 & 0.21 \\
0.49 & 1.00 & 0.57 & 0.46 & 0.53 & 0.30 & 0.21 & 0.16 & 0.08 & 0.27 & 0.35 & 0.20 \\
0.53 & 0.57 & 1.00 & 0.48 & 0.57 & 0.31 & 0.23 & 0.14 & 0.07 & 0.24 & 0.37 & 0.18 \\
0.49 & 0.46 & 0.48 & 1.00 & 0.57 & 0.24 & 0.22 & 0.12 & 0.19 & 0.21 & 0.29 & 0.16 \\
0.51 & 0.53 & 0.57 & 0.57 & 1.00 & 0.38 & 0.32 & 0.17 & 0.23 & 0.32 & 0.36 & 0.27 \\
\hline
0.33 & 0.30 & 0.31 & 0.24 & 0.38 & 1.00 & 0.43 & 0.27 & 0.24 & 0.34 & 0.37 & 0.40 \\
0.32 & 0.21 & 0.23 & 0.22 & 0.32 & 0.43 & 1.00 & 0.33 & 0.26 & 0.54 & 0.32 & 0.58 \\
0.20 & 0.16 & 0.14 & 0.12 & 0.17 & 0.27 & 0.33 & 1.00 & 0.25 & 0.46 & 0.29 & 0.45 \\
0.19 & 0.08 & 0.07 & 0.19 & 0.23 & 0.24 & 0.26 & 0.25 & 1.00 & 0.28 & 0.30 & 0.27 \\
0.30 & 0.27 & 0.24 & 0.21 & 0.32 & 0.34 & 0.54 & 0.46 & 0.28 & 1.00 & 0.35 & 0.59 \\
0.37 & 0.35 & 0.37 & 0.29 & 0.36 & 0.37 & 0.32 & 0.29 & 0.30 & 0.35 & 1.00 & 0.31 \\
0.21 & 0.20 & 0.18 & 0.16 & 0.27 & 0.40 & 0.58 & 0.45 & 0.27 & 0.59 & 0.31 & 1.00
\end{array}\right)
$$

Job Satisfaction Example II

Define $R_{11}^{-1/2}, R_{22}^{-1/2}$ using the spectral decomposition theorem. Then the rescaled correlation $R_{11}^{-1/2} R_{12} R_{22}^{-1/2}$ becomes

$$\begin{pmatrix} 0.133 & 0.170 & 0.079 & 0.080 & 0.118 & 0.170 & -0.003 \\ 0.121 & 0.001 & 0.041 & -0.062 & 0.119 & 0.173 & 0.048 \\ 0.127 & 0.044 & 0.014 & -0.088 & 0.056 & 0.205 & 0.012 \\ 0.041 & 0.059 & 0.001 & 0.109 & 0.033 & 0.100 & 0.003 \\ 0.201 & 0.129 & 0.005 & 0.136 & 0.132 & 0.122 & 0.100 \end{pmatrix}$$

Job Satisfaction Example III

Of which singular decomposition is

$$
\begin{pmatrix}
-0.548 & -0.213 & -0.638 & 0.495 & -0.036 \\
-0.396 & 0.529 & 0.307 & 0.136 & -0.672 \\
-0.396 & 0.612 & -0.155 & -0.332 & 0.578 \\
-0.243 & -0.299 & -0.284 & -0.791 & -0.381 \\
-0.572 & -0.459 & 0.628 & -0.003 & 0.262
\end{pmatrix}
\begin{pmatrix}
0.554 & 0 & 0 & 0 & 0 & 0\,0 \\
0 & 0.236 & 0 & 0 & 0 & 0\,0 \\
0 & 0 & 0.119 & 0 & 0 & 0\,0 \\
0 & 0 & 0 & 0.072 & 0 & 0\,0 \\
0 & 0 & 0 & 0 & 0.057 & 0\,0
\end{pmatrix}
$$

$$
\times \begin{pmatrix}
-0.534 & 0.036 & 0.392 & 0.103 & 0.422 \\
-0.360 & -0.360 & -0.423 & 0.314 & 0.521 \\
-0.122 & 0.046 & -0.310 & 0.542 & -0.380 \\
-0.160 & -0.838 & -0.015 & -0.359 & -0.316 \\
-0.393 & 0.007 & 0.222 & 0.407 & -0.523 \\
-0.608 & 0.402 & -0.324 & -0.549 & -0.178 \\
-0.144 & -0.056 & 0.646 & -0.022 & -0.014
\end{pmatrix}^{\top}
$$

Job Satisfaction Example IV

Then the coefficients matrices are

	Coefficients for $\mathbf{Y}$					ρ_j^*
$\mathbf{a}_1^\top$	-0.422	-0.195	-0.168	0.023	-0.460	0.554
$\mathbf{a}_2^\top$	-0.343	0.668	0.853	-0.356	-0.729	0.236
$\mathbf{a}_3^\top$	-0.858	0.443	-0.259	-0.423	0.980	0.119
$\mathbf{a}_4^\top$	0.788	0.269	-0.469	-1.042	0.168	0.072
$\mathbf{a}_5^\top$	-0.031	-0.983	0.914	-0.524	0.439	0.057

ρ_j^*		Coefficients for $\mathbf{Z}$						
0.554	$\mathbf{b}_1^\top$	-0.425	-0.209	0.036	-0.024	-0.290	-0.516	0.110
0.236	$\mathbf{b}_2^\top$	0.088	-0.436	0.093	-0.926	0.101	0.554	0.032
0.119	$\mathbf{b}_3^\top$	0.492	-0.783	-0.478	-0.007	0.283	-0.412	0.928
0.072	$\mathbf{b}_4^\top$	0.128	0.341	0.606	-0.404	0.447	-0.688	-0.274
0.057	$\mathbf{b}_5^\top$	0.482	0.750	-0.346	-0.312	-0.703	-0.180	0.014

Job Satisfaction Example V

- Assessment for $H : \Sigma_{12} = O$: $n = 784$, $p = 5$, $q = 7$ and $\widehat{\rho}_j^*$ are $0.554, 0.236, 0.119, 0.072, 0.057$. Hence

$$-2\log\Lambda = -n\sum_{j=1}^{p}\log(1-(\widehat{\rho}_j^*)^2) = 350.0303 > 49.802 = \chi^2_\gamma(pq).$$

  Hence $\Sigma_{12} = O$ is rejected at the significance level 5%.
- Assessment for $H : \rho_1^* \geq \rho_2^* \geq \cdots \geq \rho_k^* > 0 = \rho_{k+1}^* = \cdots = \rho_p^*$ becomes
  - when $k = 1$, $62.378 > 36.415 = \chi^2_{0.95}(24)$ implies that the hypothesis is rejected
  - when $k = 2$, $17.722 > 24.996 = \chi^2_{0.95}(15)$ implies that the hypothesis cannot be rejected.
- Hence the first two canonical variate pairs are non-zero at the significance level 5%.