Version: August 5, 2012

# Notes on Canonical Correlation

Suppose we have a collection of random variables in a $(q+p) \times 1$ vector $\mathbf{X}$ that we partition in the following form (and supposing without loss of generality that $p \leq q$):

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ --- \\ X_{p+1} \\ \vdots \\ X_{p+q} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ --- \\ \mathbf{X}_2 \end{pmatrix} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ ,$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \ ; \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \ ,$$

and remembering that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$, and

$$\text{Cor}(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2) = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b} / \sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}} \ .$$

Suppose

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}'_{12}\mathbf{a} = \lambda\mathbf{a} \ ,$$

with roots $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and corresponding eigenvectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$. Also, let

$$\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}'_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{b} = \lambda\mathbf{b} \ ,$$

with roots $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $\lambda_{p+1} = \lambda_q = 0$; the corresponding eigenvectors are $\mathbf{b}_1, \ldots, \mathbf{b}_p$.

Looking at the two linear combinations, $\mathbf{a}_i'\mathbf{X}_1$ (called the $i^{th}$ canonical variate in the first set), and $\mathbf{b}_i'\mathbf{X}_2$ (called the $i^{th}$ canonical variate in the second set), the squared correlation between them is $\lambda_i$; the $i^{th}$ canonical correlation is $\sqrt{\lambda_i}$. The maximum correlation between any two linear combinations is $\sqrt{\lambda_1}$, and is obtained for $\mathbf{a}_1$ and $\mathbf{b}_1$. For $\mathbf{a}_i$ and $\mathbf{b}_i$, these are uncorrelated with every canonical variate up to that point, and maximize the correlation subject to that restriction.

Points to make:

a) The matrices $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'$ and $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ are not symmetric and so the standard eigenvector/eigenvalue decompositions are not straightforward. However, the two matrices

$$\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1/2}$$

and

$$\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}$$

are symmetric. Also,

$$\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_i = \lambda_i\mathbf{e}_i \ ,$$

and

$$\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}\mathbf{f}_i = \lambda_i\mathbf{f}_i \ ,$$

where the roots, i.e., the $\lambda_i$s, are the same as before. We can then obtain $\mathbf{a}_i = \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_i$, and $\mathbf{b}_i = \boldsymbol{\Sigma}_{22}^{-1/2}\mathbf{f}_i$. Both $\boldsymbol{\Sigma}_{11}^{-1/2}$ and $\boldsymbol{\Sigma}_{22}^{-1/2}$ are constructed from the spectral decompositions of $\boldsymbol{\Sigma}_{11} = \mathbf{PDP}'$ and $\boldsymbol{\Sigma}_{22} = \mathbf{QFQ}'$ as $\boldsymbol{\Sigma}_{11}^{-1/2} = \mathbf{PD}^{-1/2}\mathbf{P}'$ and $\boldsymbol{\Sigma}_{22}^{-1/2} = \mathbf{QF}^{-1/2}\mathbf{Q}'$. Note the normalizations of $\text{Var}(\mathbf{a}_i'\mathbf{X}_1) = \mathbf{a}_i'\boldsymbol{\Sigma}_{11}\mathbf{a}_i' = \mathbf{e}_i'\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_i = 1$ and $\text{Var}(\mathbf{b}_i'\mathbf{X}_2) = 1$.

b) There are three different normalizations that are commonly used for $\mathbf{a}_i$ and $\mathbf{b}_i$:

(i) leave as unit length so $\mathbf{a}_i'\mathbf{a}_i = \mathbf{b}_i'\mathbf{b}_i = 1$;

(ii) make the largest value 1.0 in both $\mathbf{a}_i$ and $\mathbf{b}_i$;

(iii) do as we did above and make $\mathbf{a}_i'\boldsymbol{\Sigma}_{11}\mathbf{a}_i' = 1 = \mathbf{b}_i'\boldsymbol{\Sigma}_{22}\mathbf{b}_i'$.

(c) Special cases: When $p = 1$ and $q = 1$, $\lambda_1$ is the (simple) squared correlation between two variables; when $p = 1$ and $q > 1$, $\lambda_1$ is a squared multiple correlation. In considering $\mathbf{a}_i'\mathbf{X}_1$ versus $\mathbf{X}_2$, $\lambda_i$ is the squared multiple correlation of $\mathbf{a}_i'\mathbf{X}_1$ with $\mathbf{X}_2$; $\mathbf{b}_i$ gives the regression weights.

(d) When moving to the sample, all items have direct analogues. The one restriction on sample size is $n \geq p + q + 1$.

(e) Suppose the variables $\mathbf{X}_1$ and $\mathbf{X}_2$ are transformed by nonsingular matrices, $\mathbf{A}_{p\times p}$ and $\mathbf{B}_{q\times q}$, as follows:

$$\mathbf{Y}_1 = \mathbf{A}_{p\times p}\mathbf{X}_1 + \mathbf{c}_{p\times 1}$$

$$\mathbf{Y}_2 = \mathbf{B}_{q\times q}\mathbf{X}_2 + \mathbf{d}_{q\times 1}$$

The same canonical variates and correlations using $\mathbf{Y}_1$ and $\mathbf{Y}_2$ would be generated as from $\mathbf{X}_1$ and $\mathbf{X}_2$; the weights in $\mathbf{a}_i$ and $\mathbf{b}_i$ would be on the transformed variables, obviously. In particular, we could work with standardized variables without loss of any generality, and just use the correlation matrix.

(f) To evaluate $H_0 : \mathbf{\Sigma}_{12} = \mathbf{0}$, a likelihood ratio test is available:

$$-(n - 1 - (1/2)(p + q + 1)) \ln \prod_{i=1}^{p} (1 - \lambda_i) \sim \chi^2_{pq} \ .$$

Also, sometimes a sequential process is used to test the remaining roots until nonsignificance is reached:

$$-(n - 1 - (1/2)(p + q + 1)) \ln \prod_{i=k+1}^{p} (1 - \lambda_i) \sim \chi^2_{(p-k)(q-k)} \ .$$

This latter sequential procedure is a little problematic because there is no real control over the overall significance level with this strategy.

Generally, there is some tortuous difficulty in interpreting the canonical weights substantively. I might suggest using a constrained least-squares approach (iteratively moving from one set to a second), where the weights are forced to be nonnegative.