TUTORIAL 9
STA437 WINTER 2015

AL NOSEDAL

Contents

1. Canonical Correlation

In this tutorial, we discuss how the canonical correlation model is developed. We begin with the population model and then discuss how canonical variates are extracted from sample data.

1.1. **Canonical Correlation Analysis: The Population Model.**

Let m be the number of "predictors" and p the number of "criterion" variables, and assume that $m \geq p$. Denote by $\mathbf{X}^{'} = (X_1, X_2, ..., X_m)$ the m dimensional vector of predictor variables, and by $\mathbf{Y}^{'} = (Y_1, Y_2, ..., Y_m)$ the p dimensional vector of criterion measures. Letting μ_X and μ_Y denote the respective mean vectors associated with the set variables X and Y, we can define the following population variance-covariance matrices:

$$\mathbf{\Sigma}_{xx} = E\left[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^{'}\right]$$

$$\mathbf{\Sigma}_{yy} = E\left[(\mathbf{Y} - \mu_Y)(\mathbf{Y} - \mu_Y)^{'}\right]$$

$$\mathbf{\Sigma}_{xy} = E\left[(\mathbf{X} - \mu_X)(\mathbf{Y} - \mu_Y)^{'}\right]$$

If we define an $(m+p)$ dimensional variable $\mathbf{Z} = (\mathbf{X},\ \mathbf{Y})$, then we can view the problem in terms of the partitioned variance-covariance matrix $\mathbf{\Sigma}_{zz}$ shown below

$$\mathbf{\Sigma}_{zz} = \begin{pmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{pmatrix}$$

The objective of canonical correlation analysis is to find a linear combination of the m predictors that maximally correlates with a linear combination of Y's. We will denote the respective linear combinations by

$$U = \mathbf{a}^{'}\mathbf{x} = a_1x_1 + a_2x_2 + ... + a_mx_m$$

and

$$V = \mathbf{b}^{'}\mathbf{x} = b_1x_1 + b_2x_2 + ... + b_mx_m.$$

The correlation (as a function of $\mathbf{a}$ and $\mathbf{b}$) between U and V is given by

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^{'}\mathbf{\Sigma}_{xy}\mathbf{b}}{((\mathbf{a}'(\mathbf{\Sigma}_{xx}\mathbf{a})(\mathbf{b}'(\mathbf{\Sigma}_{yy}\mathbf{b}))^{1/2}}$$

where we use the Greek letter ρ for the correlation coefficient in order to emphasize that we are dealing with the population variance-covariance matrices. Out of the infinite number of linear combinations between the X's and the Y's, we find that set of linear combinations which maximizes the correlation $\rho(\mathbf{a}, \mathbf{b})$. Since $\rho(\mathbf{a}, \mathbf{b})$ is invariant under scaling of $\mathbf{a}$ and $\mathbf{b}$, we can make an arbitrary normalization of $\mathbf{a}$ and $\mathbf{b}$. We will show, later, that this problem is equivalent to solving the following canonical equations:

$$\left(\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx} - \lambda\mathbf{I}\right)\mathbf{a} = \mathbf{0}$$

$$\left(\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy} - \lambda\mathbf{I}\right)\mathbf{a} = \mathbf{0}$$

where $\mathbf{\Sigma}_{xx}, \mathbf{\Sigma}_{yy}, \mathbf{\Sigma}_{xy}$ and $\mathbf{\Sigma}_{yx}$ are defined as before, $\mathbf{I}$ is the identity matrix, and λ is the largest eigenvalue for the characteristic equations

$$det\left(\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx} - \lambda\mathbf{I}\right) = 0$$

and

$$det\left(\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy} - \lambda\mathbf{I}\right) = 0$$

The largest eigenvalue of the product matrix

$$\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}$$

or

$$\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}$$

is the squared canonical correlation coefficient. The eigenvectors associated with the eigenvalue λ - there are two sets of eigenvectors, one for $\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}$ and one for $\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}$ - then become the vector of coefficients $\mathbf{a}$ and $\mathbf{b}$. It can be shown that

$$\mathbf{a} = \frac{\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{b}}{\sqrt{\lambda}}$$

and

$$\mathbf{b} = \frac{\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}\mathbf{a}}{\sqrt{\lambda}}$$

which means that it is not necessary to solve for both characteristic equations defined above, since the eigenvectors **a** and **b** are themselves defined interchangeably.

1.2. **Sample-Based Canonical Correlation Analysis.** The discussion so far has been in terms of $\mathbf{\Sigma}_{xx}$, $\mathbf{\Sigma}_{xy}$, $\mathbf{\Sigma}_{yx}$, and $\mathbf{\Sigma}_{yy}$, the population variance-covariance matrices. In most applications, however, these matrices will not be known. A canonical correlation analysis usually starts with a sample of n responses on the $(m+p)$ dimensional variable $\mathbf{Z} = (\mathbf{X},\ \mathbf{Y})$.

The components of the variance-covariance matrix generated from data are then used to estimate the coefficients of each pair of canonical variates. That is, the two product matrices that drive the analysis correspond to

$$\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}$$

and

$$\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{S}_{xy}$$

where $\mathbf{S}_{xx}$, $\mathbf{S}_{xy}$, $\mathbf{S}_{yx}$, and $\mathbf{S}_{yy}$ are, respectively, the sample-based estimates of $\mathbf{\Sigma}_{xx}$, $\mathbf{\Sigma}_{xy}$, $\mathbf{\Sigma}_{yx}$, and $\mathbf{\Sigma}_{yy}$. Given the necessary inverses, the procedures followed here are precisely the same as those described in the previous subsection. Frequently, the measurements collected have different properties, which means that they are not commensurable. In such case the **X** and **Y** variables making up the data matrix are first standardized to have unit variance so that the variance-covariance matrix is a correlation matrix. Following the previous approach, the two (product) matrices that become the input to the analysis are

$$\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}$$

and

$$\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{R}_{xy}.$$

The same canonical correlations will be obtained from $\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}$ and $\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{R}_{xy}$ as from $\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}$ and $\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{S}_{xy}$. The sample-based estimates of the canonical weights will be denoted by $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$.

1.3. **Example 1.** Consider the sample correlation matrix given by

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}$$

where

$$\mathbf{R}_{11} = \begin{pmatrix} 1 & 0.4 & 0.3 \\ 0.4 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{pmatrix}$$

$$\mathbf{R}_{12} = \begin{pmatrix} 0.3 & 0.4 \\ 0.2 & 0.5 \\ 0.4 & 0.1 \end{pmatrix}$$

$$\mathbf{R}_{21} = \begin{pmatrix} 0.3 & 0.2 & 0.4 \\ 0.4 & 0.5 & 0.1 \end{pmatrix}$$

$$\mathbf{R}_{22} = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix}$$

a) Calculate the canonical correlations r_1 and r_2.
b) Determine the canonical variate pairs (U_1, V_1) and (U_2, V_2).
Solution

```
## Entering matrices

R11<-matrix(c(1,0.4,0.3,0.4,1,0.4,0.3,0.4,1),nrow=3,ncol=3)

R22<-matrix(c(1,0.3,0.3,1),nrow=2,ncol=2)

R12<-matrix(c(0.3,0.2,0.4,0.4,0.5,0.1),nrow=3,ncol=2)

R21<-t(R12)

## Product 1

prod1<-solve(R11)%*%R12%*%solve(R22)%*%R21

prod1

## Note. By hand, you would need to find the solution of
## det(prod1 - lambda*I ) = 0

## Finding eigenvalues and eigenvectors
```

```
e.vec.val.1<-eigen(prod1)

e.vec.val.1

#####################################
## first canonical correlation
#####################################

r1<-sqrt(e.vec.val.1$val[1])

## u1 = first canonical variate for first set of variables
## a1 = coefficients that define u1

a1<-e.vec.val.1$vec[ ,1]

r1

a1

a1<-a1/max(a1)

a1<-matrix(a1,ncol=1)

## Product 2

prod2<-solve(R22)%*%R21%*%solve(R11)%*%R12

prod2

## Note. By hand, you would need to find the solution of
## det(prod2 - lambda*I ) = 0

## Finding eigenvalues and eigenvectors

e.vec.val.2<-eigen(prod2)

e.vec.val.2

## first canonical correlation

r1.star<-sqrt(e.vec.val.2$val[1])
```

```
## v1 = first canonical variate for second set of variables
## b1 = coefficients that define v1

b1<-e.vec.val.2$vec[ ,1]

r1.star

b1

b1<-(-1)*b1

b1<-matrix(b1,ncol=1)

b1<-b1/max(b1)

b1

## What is the correlation between u1 and v1?

t(a1)%*%R12%*%b1/sqrt( t(a1)%*%R11%*%a1*t(b1)%*%R22%*%b1 )

#####################################
## second canonical correlation
#####################################

## second eigenvalue

r2<-sqrt(e.vec.val.1$val[2])

## u2 = second canonical variate for first set of variables
## a2 = coefficients that define u2

a2<-e.vec.val.1$vec[ ,2]

r2

a2

a2<-a2/max(a2)

a2<-matrix(a2,ncol=1)
```

```
a2

## Finding eigenvalues and eigenvectors

e.vec.val.2<-eigen(prod2)

e.vec.val.2

## second canonical correlation

r2.star<-sqrt(e.vec.val.2$val[2])

## v2 = second canonical variate for second set of variables

b2<-e.vec.val.2$vec[ ,2]

r2.star

b2

b2<-(-1)*b2

b2<-b2/max(b2)

b2<-matrix(b2,ncol=1)

b2

## What is the correlation between u2 and v2?

t(a2)%*%R12%*%b2/sqrt( t(a2)%*%R11%*%a2*t(b2)%*%R22%*%b2 )

###########################
### Can we get a from b?
###########################

####################
## first pair
####################

## a1
```

```
solve(R11)%*%R12%*%b1

solve(R11)%*%R12%*%b1/max(solve(R11)%*%R12%*%b1)

a1


## b1

solve(R22)%*%R21%*%a1

solve(R22)%*%R21%*%a1/max(solve(R22)%*%R21%*%a1)

b1

####################
## second pair
####################

## a2

solve(R11)%*%R12%*%b2

solve(R11)%*%R12%*%b2/max(solve(R11)%*%R12%*%b2)

a2

## b2

solve(R22)%*%R21%*%a2

solve(R22)%*%R21%*%a2/max(solve(R22)%*%R21%*%a2)

b2
```

1.4. **Exercise.** In an investigation of the relation of the Wechsler Adult Intelligence Scale to age. Researchers obtained this matrix of correlations among the digit span and vocabulary subsets, chronological age, and years of formal education:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.45 & -0.19 & 0.43 \\ 0.45 & 1 & -0.02 & 0.62 \\ -0.19 & -0.02 & 1 & -0.29 \\ 0.43 & 0.62 & -0.29 & 1 \end{pmatrix},$$

The sample consisted of $N = 933$ men and women aged 25 to 64. Let us find the canonical correlations and covariates for the pair of WAIS subtest variates and age and education variates.