Canonical Correlation Analysis (CCA)

Lecture #13
BIOE 597, Spring 2017,
Penn State University
By Xiao Liu
Agenda

• Review
• CCA Basics
• CCA Solution
• Hypothesis Testing
• Examples
• Final Project
• Midterm Review
Independent Component Analysis (Review)

• What is ICA

“Independent component analysis (ICA) is a method for finding underlying factors or components from multivariate (multi-dimensional) statistical data. What distinguishes ICA from other methods is that it looks for components that are both statistically independent, and nonGaussian.”
Independent Component Analysis (Review)

• **Blind Signal Separation**

  - Blind signal separation (BSS), also known as blind source separation, is the separation of a set of source signals from a set of mixed signals, without the aid of information (or with very little information) about the source signals or the mixing process.
Independent Component Analysis (Review)

• **Mathematical Description**

\[ x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{im}s_m, \text{ for all } i = 1, \ldots, m \]

\[ X_{n \times r} = S_{n \times m}A_{m \times r} \]

• Giving:
  observation “\( X \)"

• Find:
  Original independent components “\( S \)"

\[ S_{n \times m} = X_{n \times r}W_{r \times m} \]
Independent Component Analysis (Review)

- **Identifiability**

  - $s_i$ are statistically independent
  
  - At most one of the sources $s_i$ is Gaussian
  
  - The number of observed mixtures, $r$, must be at least as large as the number of estimated components $m$: $r \geq m$
Independent Component Analysis (Review)

- **PCA versus ICA**
  - **PCA**: Finds directions of maximal variance in gaussian data
  - **ICA**: Finds directions of maximal independence in nongaussian data
Independent Component Analysis (Review)

- **ICA Steps: Whitening**
  - Whitening/Sphering, i.e., PCA

\[
\begin{align*}
Z &= XW \\
C_Z &= W^T C_X W = \text{diag}(\lambda_i) \\
Y &= XW_N \\
C_Y &= W_N^T C_X W_N = I
\end{align*}
\]

SVD:
\[
\begin{align*}
XV &= U\Sigma \\
XV\Sigma^{-1} &= U \\
\Sigma &= \text{diag}(\sigma_i) \\
XW &= Z \\
XW_N &= Y \\
\Sigma^{-1} &= \text{diag}(1/\sigma_i)
\end{align*}
\]

- **rotation**
- **scaling**
Independent Component Analysis (Review)

- **ICA Steps: Whitening**
  
  - Why do we do "whitening/sphering"?

\[
Y = XW_N \quad C_Y = I
\]

for any orthogonal rotation \( R \)

\[
S = YR
\]

\[
C_S = R^T C_Y R = R^T C_Y R = I
\]

- No matter how we rotate the whitened data, the resulting columns will be "uncorrelated"
Independent Component Analysis (Review)

- **ICA Steps: Rotation**
  - Maximize the statistical independence of the estimated components
    - Maximize non-Gaussianity
    - Minimize mutual information
  - Measures of non-Gaussianity and independence
    - Kurtosis: \( kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \)
    - Entropy: \( H(y) = -\int f(y) \log f(y) dy \)
    - Negentropy: \( J(y) = H(y_{gauss}) - H(y) \)
    - Kullback–Leibler divergence (relative entropy)
  - ...
Independent Component Analysis (Review)

- ICA Steps: Rotation

Separated signals after 1 step of FastICA

Separated signals after 2 steps of FastICA

Separated signals after 3 steps of FastICA

Separated signals after 4 steps of FastICA
Independent Component Analysis (Review)
• Examples
  • Original Signals
  • Independent Components

Mother’s Heart Beating
Respiration
Baby’s Heart Beating
Noise
Independent Component Analysis (Review)

- **Examples**
- Clearing up MEG (Magnetoencephalography) data
Introduction

• When we have univariate data there are times when we would like to measure the linear relationship between things
  
  o Simple Linear Regression: we have 2 variables and all we are interested in is measuring their linear relationship.
  
  o Multiple linear regression: we have several independent variables and one dependent variable.

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \cdots + \beta_1 x_{ik} + e_i \quad e_i \sim N(0, \sigma^2)
\]
Introduction

• When we have univariate data there are times when we would like to measure the linear relationship between things
  
  o Simple Linear Regression: we have 2 variables and all we are interested in is measuring their linear relationship.
  
  o Multiple linear regression: we have several independent variables and one dependent variable.

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \cdots + \beta_1 x_{ik} + e_i \quad e_i \sim N(0, \sigma^2) \]

• What if we have several dependent variables and several independent variables?
  
  o Multivariate Regression
  
  o Canonical Correlation Analysis
Introduction

• Canonical correlation analysis (CCA) is a way of measuring the linear relationship between two groups of multidimensional variables.
Introduction

- Canonical correlation analysis (CCA) is a way of measuring the linear relationship between two groups of multidimensional variables.

- Finding two sets of basis vectors such that the correlation between the projections of the variables onto these basis vectors is maximized.
Introduction

- Canonical correlation analysis (CCA) is a way of measuring the linear relationship between two groups of multidimensional variables.

- Finding two sets of basis vectors such that the correlation between the projections of the variables onto these basis vectors is maximized.

- Determine correlation coefficients.
Jargon

- Variables: two sets of variables $X$ and $Y$
Jargon

- Variables: two sets of variables $X$ and $Y$
- Canonical Variates --- Linear combinations of variables
Jargon

- Variables: two sets of variables $X$ and $Y$
- Canonical Variates --- Linear combinations of variables
- Canonical Variates Pair --- Two Canonical Variates with each from one set showing non-zero correlations
Jargon

• Variables: two sets of variables $X$ and $Y$

• Canonical Variates --- Linear combinations of variables

• Canonical Variates Pair --- Two Canonical Variates with each from one set showing non-zero correlations

• Canonical Correlations--- Correlation between Canonical Variate Pairs
CCA Definition

- Two groups of multidimensional variables $X = [x_1, x_2, ..., x_p]$ and $Y = [y_1, y_2, ..., y_q]$

Where

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ \vdots \\ x_{in} \end{bmatrix} \quad y_i = \begin{bmatrix} y_{j1} \\ y_{j2} \\ y_{j3} \\ \vdots \\ y_{jn} \end{bmatrix}$$
CCA Definition

- Two groups of multidimensional variables \( X = [x_1, x_2, \ldots, x_p] \) and \( Y = [y_1, y_2, \ldots, y_q] \)

  where \( x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ \vdots \\ x_{in} \end{bmatrix} \quad y_i = \begin{bmatrix} y_{j1} \\ y_{j2} \\ y_{j3} \\ \vdots \\ y_{jn} \end{bmatrix} \)

- Purpose of CCA: find coefficient vectors \( a_1 = (a_{11}, a_{21}, \ldots, a_{p1})^T \), and \( b_1 = (b_{11}, b_{21}, \ldots, b_{q1})^T \) to maximize the correlation \( \rho = corr(Xa_1, Yb_1) \)
CCA Definition

• Two groups of multidimensional variables $X = [x_1, x_2, ..., x_p]$ and $Y = [y_1, y_2, ..., y_q]$

where $x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ ... \\ x_{in} \end{bmatrix}$ and $y_i = \begin{bmatrix} y_{j1} \\ y_{j2} \\ y_{j3} \\ ... \\ y_{jn} \end{bmatrix}$

• Purpose of CCA: find coefficient vectors $a_1 = (a_{11}, a_{21}, ..., a_{p1})^T$, and $b_1 = (b_{11}, b_{21}, ..., b_{q1})^T$ to maximize the correlation $\rho = corr(Xa_1, Yb_1)$

• $U_1 = Xa_1$ and $V_1 = Yb_1$, i.e., linear combinations of $X$ and $Y$ respectively, are the first pair of canonical variates.
CCA Definition

• Then, the second pair of canonical variates can be found in the same way subject to the constraint that they are uncorrelated with the first pair of variables.
CCA Definition

- Then, the second pair of canonical variates can be found in the same way subject to the constraint that they are uncorrelated with the first pair of variables.
- $r = \min\{p, q\}$ pairs of canonical variate pairs can be found by repeating this procedure.
CCA Definition

• Then, the second pair of canonical variates can be found in the same way subject to the constraint that they are uncorrelated with the first pair of variables.

• \( r = \min\{p, q\} \) pairs of canonical variate pairs can be found by repeating this procedure.

• We will finally get two matrices \( A = [a_1, a_2, ..., a_r] \) and \( B = [b_1, b_2, ..., b_r] \) to transfer the \( X \) and \( Y \) to canonical variates \( U \) and \( V \).

\[
U_{n \times r} = X_{n \times p} A_{p \times r}
\]

\[
V_{n \times r} = Y_{n \times q} B_{q \times r}
\]
Geometric Interpretation
PCA versus CCA

- PCA looks for patterns with a single multivariate dataset that represent maximum amounts of the variation in the data.

- In CCA, the patterns are chosen such that the projected data onto these patterns exhibit maximum correlation – while being uncorrelated with the projections onto any other pattern.

- In other words: CCA identifies new variables that maximize the inter-relationships between two data sets, in contrast to the patterns describing the internal variability within a single dataset from PCA.
Mathematical Description

• IF $X$ and $Y$ are both centered, we can concatenate them and calculate the covariance matrix

\[
C = Cov([X Y]) = \frac{1}{n-1} [X Y]^T [X Y] = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}
\]

where $C_{xx}$ and $C_{xx}$ are within-set covariance matrices, and $C_{xy} = C_{yx}^T$ are between-set covariance matrices

• The first canonical variates $a_1$ and $b_1$ maximizes

\[
\rho_1 = \frac{a_1^T C_{xy} b_1}{\sqrt{a_1^T C_{xx} a_1} \sqrt{b_1^T C_{yy} b_1}}
\]
Mathematical Description

• The subsequent pairs of canonical variates $a_i$ and $b_i$ ($i \geq 2$) maximizes

$$\rho_i = \frac{a_i^T C_{xy} b_i}{\sqrt{a_i^T C_{xx} a_i} \sqrt{b_i^T C_{yy} b_i}}$$

subject to the constraint

$$a_i^T C_{xx} a_j = 0 \quad for \ all \ j < i$$

$$b_i^T C_{yy} b_j = 0 \quad for \ all \ j < i$$
Solution

• The solution for this problem

\[
\begin{align*}
C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} a_i &= \rho_i^2 a_i \\
C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} b_i &= \rho_i^2 b_i
\end{align*}
\]

• So, the \( a_i \) are eigenvectors of \( C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \) corresponding to eigenvalues of \( \rho_i^2 \)

• So, the \( b_i \) are eigenvectors of \( C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} \) corresponding to eigenvalues of \( \rho_i^2 \)

• They are related to each other by

\[
\begin{align*}
C_{xy} b_i &= \rho_i \lambda_x C_{xx} a_i \\
C_{yx} a_i &= \rho_i \lambda_y C_{yy} b_i
\end{align*}
\]

\[
\text{where } \lambda_x = \frac{1}{\lambda_y} = \sqrt{\frac{b_i^T C_{yy} b_i}{a_i^T C_{xx} a_i}}
\]
Steps via Eigendecomposition

• Compute the matrix $C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}$, and then eigendecompose it to get the square root of its eigenvalues $= [\rho_1, \rho_2, ..., \rho_r]$ and eigenvectors $A = [a_1, a_2, ..., a_r]$

• Compute the matrix $C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}$, and then eigendecompose it to get the square root of its eigenvalues $= [\rho_1, \rho_2, ..., \rho_r]$ and eigenvectors $B = [b_1, b_2, ..., b_r]$

• The eigenvalues for both equations are equal and between zero and one. Their square root is the canonical correlation.

• The eigenvectors are weights for constructing the linear combinations of original data, i.e., canonical variates
Hypothesis Testing

• We can also test whether the canonical correlations are significant different from zero

• The test statistic is called Wilks’s Lambda

\[
\Lambda_k = \prod_{i=k}^{\min(p,q)} (1 - \rho_i^2)
\]

\[-\left(n - 1 - \frac{1}{2}(p + q + 1)\right)\ln(\Lambda_k)\] is asymptotically distribute as a chi-squared with \((p - k + 1)(p - k + 1)\) degree of freedom
CCA Properties

• Canonical correlations are invariant.
  o scale changes (such as standardizing) will not change the correlation
  o Actually, they are invariant after nonsingular linear transformations on $X$ and $Y$.

• The first canonical correlation is the best we can do with associations.
  o it is larger than any of the simple correlations or any multiple correlation with the variables under study
Matlab Function

• \([A, B, r, U, V, \text{stat} ] = \text{canoncorr}(x, y)\)
  
  o \(x, y\): set of variables in the form of matrices
    - Each row is an observation
    - Each column is an attribute/feature

  o \(A, B\): Matrices containing the correlation coefficient

  o \(r\): Column matrix containing the canonical correlations
    (Successively decreasing)

  o \(U, V\): Canonical variates/basis vectors for \(A, B\) respectively

  o \(\text{stat}\): statistics for hypothesis testing
Example

- Suppose we have two sets of variables $X$ and $Y$

\[
X = \begin{pmatrix}
1 & 1 & 3 \\
2 & 3 & 2 \\
1 & 1 & 1 \\
1 & 1 & 2 \\
2 & 2 & 3 \\
3 & 3 & 2 \\
1 & 3 & 2 \\
4 & 3 & 5 \\
5 & 5 & 5
\end{pmatrix}, \quad Y = \begin{pmatrix}
4 & 4 & -1.07846 \\
3 & 3 & 1.214359 \\
2 & 2 & 0.307180 \\
2 & 3 & -0.385641 \\
2 & 1 & -0.078461 \\
1 & 1 & 1.61436 \\
1 & 2 & 0.814359 \\
2 & 1 & -0.0641016 \\
1 & 2 & 1.535900
\end{pmatrix}
\]
Example

• Suppose we have two sets of variables $X$ and $Y$

\[
X = \begin{pmatrix}
1 & 1 & 3 \\
2 & 3 & 2 \\
1 & 1 & 1 \\
1 & 1 & 2 \\
2 & 2 & 3 \\
3 & 3 & 2 \\
1 & 3 & 2 \\
4 & 3 & 5 \\
5 & 5 & 5
\end{pmatrix}, \quad
Y = \begin{pmatrix}
4 & 4 & -1.07846 \\
3 & 3 & 1.214359 \\
2 & 2 & 0.307180 \\
2 & 3 & -0.385641 \\
2 & 1 & -0.078461 \\
1 & 1 & 1.61436 \\
1 & 2 & 0.814359 \\
2 & 1 & -0.0641016 \\
1 & 2 & 1.535900
\end{pmatrix}
\]

• **Note:** the third column of $Y$ is a linear combination of $X$:

\[
Y(:, 3) = 0.4 \times X(:, 1) + 0.6 \times X(:, 2) - \sqrt{0.48} \times X(:, 3)
\]
Example

\[ [A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y); \]

\[
A =
\begin{bmatrix}
-0.4324 & -1.4468 & -0.8180 \\
-0.6485 & 1.0610 & 0.6070 \\
0.7489 & 0.2902 & 0.9838
\end{bmatrix}
\]

\[
B =
\begin{bmatrix}
-0.0000 & -0.8487 & -1.5200 \\
0.0000 & 1.3346 & 0.2524 \\
-1.0809 & 0.0216 & -0.9702
\end{bmatrix}
\]
Example

\[
[A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y);
\]

<table>
<thead>
<tr>
<th>A</th>
<th>A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4324</td>
<td>0.4000</td>
</tr>
<tr>
<td>-1.4468</td>
<td>0.7961</td>
</tr>
<tr>
<td>-0.8180</td>
<td>-0.5776</td>
</tr>
<tr>
<td>-0.6485</td>
<td>0.6000</td>
</tr>
<tr>
<td>1.0610</td>
<td>-0.5838</td>
</tr>
<tr>
<td>0.6070</td>
<td>0.4286</td>
</tr>
<tr>
<td>0.7489</td>
<td>-0.6928</td>
</tr>
<tr>
<td>0.2902</td>
<td>-0.1597</td>
</tr>
<tr>
<td>0.9838</td>
<td>0.6947</td>
</tr>
</tbody>
</table>

\[
B = \\
\begin{bmatrix}
-0.0000 & -0.8487 & -1.5200 \\
0.0000 & 1.3346 & 0.2524 \\
-1.0809 & 0.0216 & -0.9702 \\
\end{bmatrix}
\]

\[
B1 = \\
\begin{bmatrix}
0.0000 & -0.8348 & -0.5365 \\
0.0000 & 0.1386 & 0.8438 \\
1.0000 & -0.5329 & 0.0136 \\
\end{bmatrix}
\]
Example

\[ \begin{bmatrix} A, B, r, U, V, \text{stat} \end{bmatrix} = \text{canoncorr}(X, Y); \]

\[
A = \\
\begin{bmatrix}
-0.4324 & -1.4468 & -0.8180 \\
-0.6485 & 1.0610 & 0.6070 \\
0.7489 & 0.2902 & 0.9838
\end{bmatrix}
\]

\[
B = \\
\begin{bmatrix}
-0.0000 & -0.8487 & -1.5200 \\
0.0000 & 1.3346 & 0.2524 \\
-1.0809 & 0.0216 & -0.9702
\end{bmatrix}
\]

\[
A_1 = \\
\begin{bmatrix}
0.4000 & 0.7961 & -0.5776 \\
0.6000 & -0.5838 & 0.4286 \\
-0.6928 & -0.1597 & 0.6947
\end{bmatrix}
\]

\[
B_1 = \\
\begin{bmatrix}
0.0000 & -0.8348 & -0.5365 \\
-0.0000 & 0.1386 & 0.8438 \\
1.0000 & -0.5329 & 0.0136
\end{bmatrix}
\]
Example

\[ [A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y); \]

\[ r = \]

\[ 1.0000 \quad 0.5194 \quad 0.0910 \]

The first pair of canonical variates
Example

$$[A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y);$$

$$r = \begin{bmatrix} 1.0000 & 0.5194 & 0.0910 \end{bmatrix}$$

The second pair of canonical variates
Example

\[ [A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y); \]

\[
\begin{array}{ccc}
1.0000 & 0.5194 & 0.0910
\end{array}
\]

The third pair of canonical variates
Example

\[ [A, B, r, U, V, \text{stat}] = \text{canoncorr}(X, Y); \]

\[ r = \begin{pmatrix} 1.0000 & 0.5194 & 0.0910 \end{pmatrix} \]

The third pair of canonical variates
Example

\[ A, B, r, U, V, \text{stat} \] = canoncorr(X, Y);

**struct** with fields:

- **Wilks**: \[2.0261e-13 0.7242 0.9917]\n- **df1**: [9 4 1]
- **df2**: [7.4518 8 5]
- **F**: [1.3602e+05 0.3502 0.0418]
- **pF**: [1.5799e-18 0.8370 0.8461]
- **chisq**: [131.5237 1.4522 0.0600]
- **pChisq**: [5.7628e-24 0.8351 0.8065]
- **dfe**: [9 4 1]
- **p**: [5.7628e-24 0.8351 0.8065]
About Final Project

• You will be asked to present a paper that uses one of methods talked in the class

• 20% Grade!

• Start from April 11th

• The presentation will be 10 minutes, followed by a 2-minute question session. You’re expected to prepare some PPT slides for the presentation!

• Be clear about
  o What is the major goal of the paper?
  o How did it use the method we talked about to achieve its goal?
About Final Project

• Let me know before next Tuesday (3/21) if you want to find a paper by yourself that is more relevant to your area of research

• Otherwise, I will randomly assign a paper to you next Tuesday, as well as the time of your presentation.
CCA: Examples

Camera A

Camera B

\( \phi(\cdot) \)

Kernel Trick

KCCA

\( \alpha, \beta \)

Reprojection

KCCA Common Subspace