Canonical Correlation Analysis

Seungjin Choi

Department of Computer Science
POSTECH, Korea
seungjin@postech.ac.kr

Outline

- Canonical correlation analysis (CCA)
- Kernel CCA

Canonical Correlation Analysis (CCA)

- CCA is a multivariate analysis method of correlating linear relationships between two sets of variables, while PCA considers a single set of variables.
- Proposed by Hotelling (1936), CCA is viewed as the problem of finding basis vectors of two sets of variables such that correlations between the projections of the variables onto these basis vectors are mutually maximized.
- CCA first seeks a pair of linear combinations (a linear combination of the variables in one set and a linear combination of the variables in the other set) that has the largest correlation. Next, it determines a pair of linear combinations having the largest correlation among all pairs uncorrelated with the initially-selected pair, and so on.

Formulation for CCA

Consider two sets of multivariate data, $\mathcal{D}_x = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^{m_x}$ and $\mathcal{D}_y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\} \in \mathbb{R}^{m_y}$, where pairs $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ have correspondence.

CCA seeks a pair of linear transformations, $\boldsymbol{w}_x \in \mathbb{R}^{m_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{m_y}$ such that correlations between $\boldsymbol{w}_x^\top \boldsymbol{x}$ and $\boldsymbol{w}_y^\top \boldsymbol{y}$ are maximized.

The objective function to be maximized is:

$$
\begin{aligned}
\rho = \quad &= \frac{\mathrm{cov}\left(\boldsymbol{w}_x^\top \boldsymbol{x}, \boldsymbol{w}_y^\top \boldsymbol{y}\right)}{\sqrt{\mathrm{var}\left(\boldsymbol{w}_x^\top \boldsymbol{x}\right)\mathrm{var}\left(\boldsymbol{w}_y^\top \boldsymbol{y}\right)}} \\
&= \frac{\boldsymbol{w}_x^\top \boldsymbol{C}_{xy} \boldsymbol{w}_y}{\sqrt{\left(\boldsymbol{w}_x^\top \boldsymbol{C}_{xx} \boldsymbol{w}_x\right)\left(\boldsymbol{w}_y^\top \boldsymbol{C}_{yy} \boldsymbol{w}_y\right)}},
\end{aligned}
$$

where $\boldsymbol{C}_{xy} = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t - \boldsymbol{\mu}_x)(\boldsymbol{y}_t - \boldsymbol{\mu}_y)^\top$.

CCA: Optimization Formulation

CCA is formulated as

$$\boxed{\arg\max_{\boldsymbol{w}_x,\boldsymbol{w}_y} \ \boldsymbol{w}_x^\top \boldsymbol{C}_{xy}\boldsymbol{w}_y,}$$

subject to

$$\begin{aligned}\boldsymbol{w}_x^\top \boldsymbol{C}_{xx}\boldsymbol{w}_x &= 1,\\ \boldsymbol{w}_y^\top \boldsymbol{C}_{yy}\boldsymbol{w}_y &= 1.\end{aligned}$$

CCA: Generalized Eigenvalue Problem

Incorporating these two constraints, the Lagrangian $\mathcal{J}$ is given by

$$\boxed{\mathcal{J} = \boldsymbol{w}_x^\top \boldsymbol{C}_{xy}\boldsymbol{w}_y + \lambda_x\left(1 - \boldsymbol{w}_x^\top \boldsymbol{C}_{xx}\boldsymbol{w}_x\right) + \lambda_y\left(1 - \boldsymbol{w}_y^\top \boldsymbol{C}_{yy}\boldsymbol{w}_y\right).}$$

It follows from $\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}_x} = 0$ and $\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}_y} = 0$ that CCA is solved by the following generalized eigenvalue problem

$$\begin{bmatrix} 0 & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yx} & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{C}_{xx} & 0 \\ 0 & \boldsymbol{C}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix},$$

where $\lambda = 2\lambda_x = 2\lambda_y$.

CCA: Generalized Eigenvalue Problem (Cont'd)

Recall the generalized eigenvalue problem:

$$\begin{bmatrix} 0 & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yx} & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{C}_{xx} & 0 \\ 0 & \boldsymbol{C}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}.$$

This problem has m_x+m_y eigenvalues $\{\lambda_1, -\lambda_1, \ldots, \lambda_m, -\lambda_m, 0, \ldots, 0\}$, where $m = \min(m_x, m_y)$.

The generalized eigenvalue problem can be re-written as

$$\begin{bmatrix} \boldsymbol{C}_{xx} & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yx} & \boldsymbol{C}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = (1+\lambda) \begin{bmatrix} \boldsymbol{C}_{xx} & 0 \\ 0 & \boldsymbol{C}_{yy} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}.$$

This problem has m_x+m_y eigenvalues $\{1+\lambda_1, 1-\lambda_1, \ldots, 1+\lambda_m, 1-\lambda_m, 1, \ldots, 1\}$.

$\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}_x} = 0$ and $\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}_y} = 0$ lead to

$$\begin{aligned}\boldsymbol{C}_{xy}\boldsymbol{w}_y - 2\lambda_x \boldsymbol{C}_{xx}\boldsymbol{w}_x &= 0, && (1)\\ \boldsymbol{C}_{yx}\boldsymbol{w}_x - 2\lambda_y \boldsymbol{C}_{yy}\boldsymbol{w}_y &= 0. && (2)\end{aligned}$$

Pre-multiply (1) by $\boldsymbol{w}_x^\top$ and pre-multiply (2) $\boldsymbol{w}_y^\top$ to obtain

$$\begin{aligned}\boldsymbol{w}_x^\top \boldsymbol{C}_{xy}\boldsymbol{w}_y - 2\lambda_x \underbrace{\boldsymbol{w}_x^\top \boldsymbol{C}_{xx}\boldsymbol{w}_x}_{1} &= 0,\\ \boldsymbol{w}_y^\top \boldsymbol{C}_{yx}\boldsymbol{w}_x - 2\lambda_y \underbrace{\boldsymbol{w}_y^\top \boldsymbol{C}_{yy}\boldsymbol{w}_y}_{1} &= 0,\end{aligned}$$

leading to

$$\begin{aligned}\boldsymbol{w}_x^\top \boldsymbol{C}_{xy}\boldsymbol{w}_y &= 2\lambda_x,\\ \boldsymbol{w}_y^\top \boldsymbol{C}_{yx}\boldsymbol{w}_x &= 2\lambda_y.\end{aligned}$$

Since $\boldsymbol{w}_x^\top \boldsymbol{C}_{xy}\boldsymbol{w}_y = \boldsymbol{w}_y^\top \boldsymbol{C}_{yx}\boldsymbol{w}_x$, we have $\boxed{\lambda = 2\lambda_x = 2\lambda_y}$.

CCA: Extension to Multiple Sets of Variables

Consider n multiple sets of variables, $\boldsymbol{x}_1 \in \mathbb{R}^{m_1}, \boldsymbol{x}_2 \in \mathbb{R}^{m_2}, \ldots, \boldsymbol{x}_n \in \mathbb{R}^{m_n}$.

Then, CCA is formulated as

$$\boxed{\underset{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n}{\arg\max} \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{w}_i^\top \boldsymbol{C}_{ij} \boldsymbol{w}_j,}$$

subject to

$$\sum_{i=1}^{n} \boldsymbol{w}_i^\top \boldsymbol{C}_{ii} \boldsymbol{w}_j = 1.$$

CCA: Generalized Eigenvalue Problem

Incorporating these two constraints, the Lagrangian $\mathcal{J}$ is given by

$$\boxed{\mathcal{J} = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{w}_i^\top \boldsymbol{C}_{ij} \boldsymbol{w}_j + \lambda \left(1 - \sum_{i=1}^{n} \boldsymbol{w}_i^\top \boldsymbol{C}_{ii} \boldsymbol{w}_j\right).}$$

It follows from $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_i} = 0$ for $i = 1, \ldots, n$ that we have

$$\begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1n} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \cdots & \boldsymbol{C}_{2n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{C}_{n1} & \boldsymbol{C}_{n2} & \cdots & \boldsymbol{C}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_n \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{C}_{11} & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_{22} & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \boldsymbol{C}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_n \end{bmatrix}.$$

The minimal generalized eigenvalue has the fixed range $[0, 1]$, whereas the maximal generalized eigenvalue has a range dependent on the dimensions of the variables. The minimal generalized eigenvalue is more convenient.

Kernel CCA

Note that covariance matrices are

$$\boldsymbol{C}_{xy} = \boldsymbol{X}\boldsymbol{Y}^\top, \quad \boldsymbol{C}_{xx} = \boldsymbol{X}\boldsymbol{X}^\top, \quad \boldsymbol{C}_{yy} = \boldsymbol{Y}\boldsymbol{Y}^\top.$$

We use kernel trick, assuming $\boldsymbol{w}_x = \boldsymbol{X}\boldsymbol{\alpha}_x$ and $\boldsymbol{w}_y = \boldsymbol{Y}\boldsymbol{\alpha}_y$. Define kernel matrices as $\boldsymbol{K}_x = \boldsymbol{X}^\top\boldsymbol{X}$ and $\boldsymbol{K}_y = \boldsymbol{Y}^\top\boldsymbol{Y}$. Then KCCA is formulated as

$$\boxed{\underset{\boldsymbol{\alpha}_x,\boldsymbol{\alpha}_y}{\arg\max} \ \boldsymbol{\alpha}_x^\top \boldsymbol{K}_x \boldsymbol{K}_y \boldsymbol{\alpha}_y,}$$

subject to

$$\begin{aligned} \boldsymbol{\alpha}_x^\top \boldsymbol{K}_x^2 \boldsymbol{\alpha}_x &= 1, \\ \boldsymbol{\alpha}_y^\top \boldsymbol{K}_y^2 \boldsymbol{\alpha}_y &= 1. \end{aligned}$$

Kernel CCA: Generalized Eigenvalue Problem

As in CCA, KCCA reduces to a generalized eigenvalue problem:

$$\begin{bmatrix} 0 & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{K}_x^2 & 0 \\ 0 & \boldsymbol{K}_y^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix},$$

which can be re-written as

$$\begin{bmatrix} \boldsymbol{K}_x^2 & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & \boldsymbol{K}_y^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} = (1+\lambda) \begin{bmatrix} \boldsymbol{K}_x^2 & 0 \\ 0 & \boldsymbol{K}_y^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix},$$

which is further simplified as

$$\begin{bmatrix} \boldsymbol{K}_x & \boldsymbol{K}_y \\ \boldsymbol{K}_x & \boldsymbol{K}_y \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} = (1+\lambda) \begin{bmatrix} \boldsymbol{K}_x & 0 \\ 0 & \boldsymbol{K}_y \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix}.$$