Canonical Correlation Analysis (CCA)

Nanbo Sun

This document is about the derivation of CCA

1 Problem Statement

Given two random vectors $\vec{x}, \vec{y}$, which has been demeaned, we want to find two transformation $\vec{w}, \vec{c}$

$$t = \vec{w}^T \vec{x} \tag{1}$$

$$u = \vec{c}^T \vec{y} \tag{2}$$

to maximize the correlation between random variable t and u

$$\max\left(\frac{E[tu]}{\sqrt{E[t^2]E[u^2]}}\right)^2 \tag{3}$$

under the condition that

$$\vec{w}^T C_{xx} \vec{w} = 1 \tag{4}$$

$$\vec{c}^T C_{yy} \vec{c} = 1 \tag{5}$$

2 Solution

$$\max\left(\frac{E[tu]}{\sqrt{E[t^2]E[u^2]}}\right)^2 = \max \frac{(E[\vec{w}^T \vec{x} \vec{y}^T \vec{c}])^2}{\vec{w}^T C_{xx} \vec{w} \vec{c}^T C_{yy} \vec{c}} \tag{6}$$

$$= \max(\vec{w}^T C_{xy} \vec{c})^2 \tag{7}$$

By using Lagrange multiplier, we get following equation

$$L(\vec{w}, \vec{c}) = (\vec{w}^T C_{xy} \vec{c})^2 - \lambda_w(\vec{w}^T C_{xx} \vec{w} - 1) - \lambda_c(\vec{c}^T C_{yy} \vec{c} - 1) \tag{8}$$

$$\frac{\partial L}{\partial w} = 2(\vec{w}^T C_{xy} \vec{c}) C_{xy} \vec{c} - 2\lambda_w C_{xx} \vec{w} = 0 \tag{9}$$

$$\frac{\partial L}{\partial c} = 2(\vec{w}^T C_{xy} \vec{c})(\vec{w}^T C_{xy})^T - 2\lambda_c C_{yy} \vec{c} = 0 \tag{10}$$

Times $\vec{w}^T, \vec{c}^T$ in both sides

$$\vec{w}^T(\vec{w}^T C_{xy} \vec{c}) C_{xy} \vec{c} = \lambda_w \vec{w}^T C_{xx} \vec{w} \tag{11}$$

$$\vec{c}^T(\vec{w}^T C_{xy}\vec{c}) C_{yx}\vec{w} = \lambda_c \vec{c}^T C_{yy}\vec{c} \tag{12}$$

Since $\vec{w}^T C_{xx}\vec{w} = 1, \vec{c}^T C_{yy}\vec{c} = 1$

$$(\vec{w}^T C_{xy}\vec{c})^2 = \lambda_w \tag{13}$$

$$(\vec{w}^T C_{xy}\vec{c})^2 = \lambda_c \tag{14}$$

It's easy to see that $\lambda = \lambda_w = \lambda_c$.

We can also write the above function in this way

$$\vec{w}^T C_{xy}\vec{c} = \sqrt{\lambda} \tag{15}$$

Substitute above Equation into Eq. (9) and Eq. (10), we can get following

$$C_{xy}\vec{c} = \sqrt{\lambda} C_{xx}\vec{w} \tag{16}$$

$$C_{yx}\vec{w} = \sqrt{\lambda} C_{yy}\vec{c} \tag{17}$$

$$\begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & C_{yy}^{-1} \end{bmatrix} \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} = \sqrt{\lambda} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} \tag{18}$$

2.1 Wiki Solution

According to Eq. (17)

$$\vec{c} = \frac{1}{\sqrt{\lambda}} C_{yy}^{-1} C_{yx}\vec{w} \tag{19}$$

By substituting Eq. (19) into Eq. (16)

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}\vec{w} = \lambda \vec{w} \tag{20}$$

Similarly,

$$C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}\vec{c} = \lambda \vec{c} \tag{21}$$

Therefore, $\vec{w}, \vec{c}$ are eigenvectors of $C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}$ and $C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}$. This matches the solution in wiki `https://www.wikiwand.com/en/Canonical_correlation`

2.2 Matlab cononcorr Solution

According to Eq. (16) and (17) we can get

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} = \sqrt{\lambda} \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} \tag{22}$$

Theorem: Cholesky decomposition

The Cholesky decomposition of a Hermitian positive-definite matrix A is a decomposition of the form

$$A = LL^* \tag{23}$$

where L is a lower triangular matrix with real and positive diagonal entries, and L^* denotes the conjugate transpose of L

Because C_{xx}, C_{yy} are positive-definite matrix, we can decompose it

$$C_{xx} = T_x^T T_x \tag{24}$$

$$C_{yy} = T_y^T T_y \tag{25}$$

Then, Eq. (22) becomes

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} = \sqrt{\lambda} \begin{bmatrix} T_x^T T_x & 0 \\ 0 & T_y^T T_y \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{c} \end{bmatrix} \tag{26}$$

$$\begin{bmatrix} 0 & T_x^{-T} C_{xy} T_y^{-1} \\ T_y^{-T} C_{yx} T_x^{-1} & 0 \end{bmatrix} \begin{bmatrix} T_x\vec{w} \\ T_y\vec{c} \end{bmatrix} = \sqrt{\lambda} \begin{bmatrix} T_x\vec{w} \\ T_y\vec{c} \end{bmatrix} \tag{27}$$

If we set $\Omega = T_x^{-T} C_{xy} T_y^{-1}$ and $\vec{a} = T_x\vec{w}$ and $\vec{b} = T_y\vec{c}$, we get

$$\begin{bmatrix} 0 & \Omega \\ \Omega^T & 0 \end{bmatrix} \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} = \sqrt{\lambda} \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} \tag{28}$$

Based on

Theorem: Singular Value Decomposition (SVD)

If $A = USV^T$, then U and V are eigenvectors of AA^T and A^TA

$$A = USV^T \tag{29}$$

$$U^TU = V^TV = I \tag{30}$$

$$AA^TU = USV^TVSU^TU = US^2 \tag{31}$$

$$A^TAV = VSU^TUSV^TV = VS^2 \tag{32}$$

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} S \tag{33}$$

We know that $\vec{a}, \vec{b}$ are left and right singular vectors of Ω.

Theorem: QR decomposition

we can factor a complex $m \times n$ matrix A, with $m \geq n$, as the product of an $m \times m$ unitary matrix Q and an $m \times n$ upper triangular matrix R. As the bottom $(m-n)$ rows of an $m \times n$ upper triangular matrix consist entirely of zeroes, it is often useful to partition R, or both R and Q:

$$A = QR = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \tag{34}$$

where R_1 is an $n \times n$ upper triangular matrix, 0 is an $(m-n) \times n$ zero matrix, Q_1 is $m \times n$, Q_2 is $m \times (m-n)$, and Q_1 and Q_2 both have orthogonal columns.

If we do a QR decomposition for X, Y.

$$X = QR \tag{35}$$
$$Q^T Q = I \tag{36}$$
$$Y = PS \tag{37}$$
$$P^T P = I \tag{38}$$

Then,

$$\Omega = R^{-T} C_{xy} S^{-1} = R^{-T} X^T Y S^{-1} = R^{-T} R^T Q^T P S S^{-1} = Q^T P \tag{39}$$

$\vec{a}, \vec{b}$ are just singular values of $Q^T P$, after that we can compute $\vec{w}, \vec{c}$

$$\vec{w} = T_x^{-1} \vec{a} \tag{40}$$
$$\vec{v} = T_y^{-1} \vec{b} \tag{41}$$

The Matlab function cononcorr.m implement the CCA in this way.