Journal of Mathematics and Statistics

Original Research Paper

Optimization of Penalty Parameter in Penalized Nonlinear Canonical Correlation Analysis by using Cross-Validation

Isamu Nagai

Department of International Liberal Studies, School of International Liberal Studies, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota Aichi 470-0393, Japan

Article history
Received: 05-06-2015
Revised: 31-10-2015
Accepted: 09-11-2015

Corresponding Author:
Isamu Nagai
Department of International Liberal Studies, School of International Liberal Studies, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota Aichi 470-0393, Japan
E-mail: inagai@lets.chukyo-u.ac.jp

Abstract: There is Canonical Correlation Analysis (CCA) as a way to find a linear relationship between a pair of random vectors. However, CCA cannot find a nonlinear relationship between them since the method maximizes the correlation between linear combinations of the vectors. In order to find the nonlinear relationship, we convert the vectors through some known conversion functions like a kernel function. Then we find the nonlinear relationship in the original vectors through the conversion function. However, this method has a critical issue in that the maximized correlation sometimes becomes 1 even if there is no relationship between the random vectors. Some author proposed a penalized method with a penalty parameter that avoids this issue when the kernel functions are used for conversion. In this method, however, methods have not been proposed for optimizing the penalty and other hyper parameters in the conversion function, even though the results heavily depend on these parameters. In this study, we propose an optimization method for the penalty and other parameters, based on the simple cross-validation method.

Keywords: Canonical Correlation Analysis, Cross-Validation, Nonlinear Relationship, Penalized Method

Introduction

Let y and x be q_0- and p_0-dimensional random vectors. Without of generality, we assume $E[y] = 0_{q_0}$ and $E[x] = 0_{p_0}$ where 0_ℓ is an ℓ-dimensional vector of zeros. Moreover, let $\Sigma = E[(y', x')'(y', x')]$ be a $(q_0 + p_0) \times (q_0 + p_0)$ unknown matrix and $\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma'_{yx} & \Sigma_{xx} \end{pmatrix}$ where Σ_{yy} is a $q_0 \times q_0$ matrix, Σ_{yx} is a $q_0 \times p_0$ matrix and Σ_{xx} is a $p_0 \times p_0$ matrix and we assume $\det(\Sigma_{yy}) \neq 0$ and $\det(\Sigma_{xx}) \neq 0$. Note that $\Sigma_{yy} = \mathrm{Var}(y)$, $\Sigma_{yx} = \mathrm{Cov}(y, x)$ and $\Sigma_{xx} = \mathrm{Var}(x)$, since $E[y] = 0_{q_0}$ and $E[x] = 0_{p_0}$.

As a method for finding the linear relationship between y and x, Hotelling (1936) proposed Canonical Correlation Analysis (CCA). This method is formulated as follows:

$$\max_{\alpha \in \mathbb{R}^{q_0}, b \in \mathbb{R}^{p_0}} a'\Sigma_{yx} b \; s.t. \; a'\Sigma_{yy} a = 1 \; and \; b'\Sigma_{xx} b = 1 \qquad (1.1)$$

Usually, using the Lagrange method of undetermined multipliers, we can derive the solutions of a and b. More details of CCA can be found in Muirhead (1982), Gittins (1985), Srivastava (2002) and Weenink (2003). This method is currently being used for data analysis (see, e.g., Doeswijk *et al.*, 2011). CCA, however, can not find a nonlinear relationships between y and x, since the maximization term in (1.1) is equivalent to $\mathrm{Cov}(a'y, b'x)$, which evaluates the linear relationship between linear combinations $a'y$ and $b'x$.

In order to find a nonlinear relationship between y and x, we consider converting them by using some known functions like a kernel function. Then, CCA can then find a nonlinear relationship between y and x through the conversion functions. This method is referred to as a Nonlinear Canonical Correlation Analysis (NCCA) and it is shown in section 2. Hardoon *et al.* (2004) pointed out that NCCA has a critical issue which is also shown in section 2.

Using the same idea as is used in the penalized nonlinear regression model, Akaho (2000) proposed a penalized NCCA when the kernel functions are used for the conversion functions. We will refer to the penalized NCCA as PNCCA even when it uses any conversion functions instead of the kernel function.

In PNCCA, no criteria have yet been developed for optimizing the penalty and other hyper parameters in the conversion function. The reason of this problem, it is difficult to know how to evaluate the result of PNCCA. In particular, determining how to optimize the penalty and

other hyper parameters in conversion function is important, since the result of PNCCA heavily depends on these parameters. Hence, in this study, we create a evaluating function for evaluating the estimated value. Based on this function and the ordinary Cross-Validation (CV) method, we propose the simple form of CV method for optimizing these parameters in PNCCA. Details of the proposed function and CV method are presented in section 3.

The remainder of the present paper is organized as follows: In section 2, we present more details of CCA, NCCA and PNCCA. In section 3, we propose the simple CV method for optimizing several parameters in PNCCA. In section 4, we use numerical studies to compare CCA, NCCA and PNCCA based on the optimized parameters. In section 5, we present our conclusions. Using the proposed CV method, we can select the variables in y and x; we illustrate this method in the Appendix.

CCA, NCCA and PNCCA

In this section, we illustrate CCA, NCCA and PNCCA. We first illustrate CCA, which is expressed as (1.1). Using the Lagrange method of undetermined multipliers, since $\det(\Sigma_{xx}) \neq 0$ and $\det(\Sigma_{yy}) \neq 0$, CCA is the same as solving the following eigenvalue problem:

$$\Sigma_{xx}^{-1}\Sigma_{yx}'\Sigma_{yy}^{-1}\Sigma_{yx}\tilde{b} = \tilde{\theta}^2\tilde{b} \tag{2.1}$$

and $\tilde{a}=\Sigma_{yy}^{-1}\Sigma_{yx}\tilde{b}/\tilde{\theta}$ where $\tilde{\theta} = \tilde{a}'\Sigma_{yx}\tilde{b} > 0$. Hence, solving the eigenvalue problem in (2.1) and using the largest eigenvalue and the corresponding eigenvector, we can solve the maximization problem under several conditions in (1.1). More details of CCA can be found in e.g., Muirhead (1982).

However, CCA can not find a nonlinear relationship between y and x. In order to find a nonlinear relationship between them, we convert x as $w = \varphi(x)$ where $\varphi(.)$: $\mathbb{R}^{P_0} \to \mathbb{R}^{P_1}$ is a known conversion function. Without of generality, we also assume $E[w] = 0_{p_1}$ and we also assume $\det(\Sigma_{ww}) \neq 0$ where Σ_{ww} = Var (w). When we use CCA for y and w, we can find the nonlinear relationship between y and x through $\varphi(\cdot)$. This is the NCCA. However, Hardoon *et al*. (2004) pointed out that, even if there is no relationship between y and x, the result of NCCA shows there are heavily relationship between them.

In order to avoid this problem, Akaho (2000) proposed PNCCA only when we use the kernel functions as conversion functions. This is the primary method we consider in this study. Since, in our setting, only x is converted, PNCCA is expressed as follows:

$$\max_{a\in\mathbb{R}^{q_0}, d\in\mathbb{R}^{P_1}} a'\Sigma_{yw}d \; s.t.\; a'\Sigma_{yy}a = 1 \; and \; d'\left(\Sigma_{ww}+\lambda P\right)d = 1 \tag{2.2}$$

where, Σ_{yw} = Cov (y, w), λ is a nonnegative penalty parameter and P is a known $p_1 \times p_1$ nonnegative definite penalty matrix. Note that $\lambda d'Pd$ is the penalty term in (2.2) since $\lambda d'Pd \geq 0$ for any $d \in \mathbb{R}^{P_1}$. Furthermore, we note $\det(\Sigma_{ww} + \lambda P) \geq \det(\Sigma_{ww}) > 0$ since $\lambda \geq 0$ and P is the nonnegative definite matrix (see, e.g., Lütkepohl (1996) section 4.2.6, (11)). The same as for CCA in (1.1), in order to solve the maximization problem under various conditions in (2.2), we use the Lagrange method of undetermined multipliers as follows:

$$L_P(\eta_a,\eta_d,a,d,\lambda|P) = a'\Sigma_{yw}d - \frac{\eta_a}{2}(a'\Sigma_{yy}a-1) - \frac{\eta_d}{2}\left\{d'(\Sigma_{ww}+\lambda P)d-1\right\},$$

where, η_a and η_d are undetermined nonnegative constants. Akaho (2000) only showed the above expression without (2.2) when the conversion function is the kernel function. For the fixed λ, solving the simultaneous equations $\partial L_P(\eta_a,\eta_d,a,d,\lambda|P)/(\partial a)\big|_{a=\tilde{a}_\lambda} = 0_{q_0}$, $\partial L_P(\eta_a,\eta_d,a,d,\lambda|P)/(\partial d)\big|_{d=\tilde{d}_\lambda} = 0_{p_1}$, $\partial L_P(\eta_a,\eta_d,a,d,\lambda|P)/\left(\partial\eta_a\right)\big|_{\eta_a=\tilde{\eta}_{a,\lambda}} = 0$ and $\partial L_P(\eta_a,\eta_d,a,d,\lambda|P)/(\partial\eta_d)\big|_{\eta_d=\tilde{\eta}_{d,\lambda}} = 0$ coincides with solving the following eigenvalue problem:

$$\left(\Sigma_{ww}+\lambda P\right)^{-1}\Sigma_{yw}'\Sigma_{yy}^{-1}\Sigma_{yw}\tilde{d}_\lambda = \tilde{\eta}_\lambda^2\tilde{d}_\lambda \tag{2.3}$$

and $\tilde{a}_\lambda = \Sigma_{yy}^{-1}\Sigma_{yw}\tilde{d}_\lambda/\tilde{\eta}_\lambda$, where $\tilde{\eta}_\lambda = \tilde{a}'_\lambda\Sigma_{yw}\tilde{d}_\lambda > 0$ and $\tilde{\eta}_\lambda = \tilde{\eta}_{a,\lambda} = \tilde{\eta}_{d,\lambda}$. Hence, when the penalty parameter λ is given, we can solve (2.2) by using the largest eigenvalue and the corresponding eigenvector of the above eigenvalue problem.

However, although it is important, there are no optimization methods for λ and other parameters in the conversion function $\varphi(\cdot)$. In the next section, we propose a simple CV method for optimizing λ and some of the parameters in the known conversion function $\varphi(\cdot)$.

Proposed Method

In this section, we propose a simple CV method for optimizing the penalty and other hyper parameters in the conversion function $\varphi(\cdot)$ which are used in PNCCA. In order to propose CV method, we consider evaluating function for the results of PNCCA.

Firstly, since Σ_{ww}, Σ_{yw} and Σ_{yy} are unknown matrices, we use their unbiased estimators to estimate $\tilde{\eta}_\lambda$, $\tilde{a}_\lambda$, and $\tilde{d}_\lambda$. Let S be the ordinary unbiased estimators for Σ based on the sample $\{y_i, x_i\}_{i=1,\ldots,n}$ and $w_i = \varphi(x_i)$. Then we divide S as $\begin{pmatrix} S_{yy} & S_{yw} \\ S_{yw}' & S_{ww} \end{pmatrix}$ where S_{yy} is a $q_0 \times q_0$ matrix, S_{yw} is a $q_0 \times p_1$ matrix and S_{ww} is a $p_1 \times p_1$ matrix. In

order to estimate $\tilde{\eta}_\lambda, \tilde{a}_\lambda$ and $\tilde{d}_\lambda$, we use S_{yy}, S_{yw} and S_{ww} instead of using Σ_{yy}, Σ_{yw} and Σ_{ww} in (2.3), respectively. Let $\hat{\eta}_\lambda(>0), \hat{a}_\lambda$ and $\hat{d}_\lambda$ be the estimators for $\tilde{\eta}_\lambda$, $\tilde{a}_\lambda$ and $\tilde{d}_\lambda$, respectively. Then, $\hat{\eta}_\lambda^2$ and $\hat{d}_\lambda$ are derived as the largest eigenvalue and the corresponding eigenvector of $(S_{ww}+\lambda P)^{-1}S'_{yw}S_{yy}^{-1}S_{yw}$ and $\hat{a}_\lambda = S_{yy}^{-1}S_{yw}\hat{d}_\lambda / \hat{\eta}_\lambda$ from (2.3).

We consider creating an objective function in order to evaluate $\tilde{a}_\lambda$ and $\tilde{d}_\lambda$ for optimizing several parameters in PNCCA that are the penalty parameter and the other parameter in the conversion function. Since the purpose of PNCCA is maximizing $a'\Sigma_{yw}d$ under several conditions, we consider the following evaluation function:

$$R^* = E\left[\hat{a}'_\lambda\ \Sigma_{yw}\hat{d}_\lambda\right] \tag{3.1}$$

Maximizing the above function, we can optimize the parameters in PNCCA. Here, we note that $\hat{a}_\lambda$ and $\hat{d}_\lambda$ are derived from $\{y_i, x_i\}_{i=1,...,n}$. However, Σ_{yw} is an unknown covariance matrix. We therefore consider using an estimator for Σ_{yw} that does not depend on $\{y_i, x_i\}_{i=1,...,n}$ in order to estimate R^* in (3.1) since we use $\{y_i, x_i\}_{i=1,...,n}$ for deriving $\hat{a}_\lambda$ and $\hat{d}_\lambda$.

Then, let y^* and x^* be new variables that are obtained independently from $\{y_i, x_i\}_{i=1,...,n}$ and let S^* be the variance and covariance matrix between y^* and $w^* = \varphi(x^*)$. Then, letting $S_{y^*w^*}$ be the first $q_0 \times p_1$ matrix in S^*, we can regard $S_{y^*w^*}$ as an estimator for Σ_{yw}. Based on $S_{y^*w^*}$, the evaluation function R^* in (3.1) is estimated by using the average of the following value:

$$\hat{R}^* = \hat{a}'_\lambda\ S_{y^*w^*}\hat{d}_\lambda \tag{3.2}$$

Nevertheless, this evaluation function $\hat{R}^*$ in (3.2) also can not be used directly for optimizing the parameters in PNCCA since y^* and w^* are not obtained. We thus use the simple CV method to optimize the penalty parameter and other hyper parameter in the conversion function that are in PNCCA. As similar as the ordinary CV method for some regression model, we divide $\{y_i, x_i\}_{i=1,...,n}$ into two subsets. One of them is used for estimation, and other one is used for evaluating the estimated value.

Let $V = (v_1 \ldots v_n)'$ be $n \times (q_0+p_1)$ matrix, where $v_i = (y_i', w_i')'$, $(i = 1,...,n)$. The essence of the propose method is to obtain a matrix that is an alternative to $S_{y^*w^*}$. The alternative matrix to it can not be derived from using only one sample.

We now use v_i and v_j, $(i \neq j)$ to derive an alternative matrix to $S_{y^*w^*}$, which can be defined as:

$$\hat{S}_{[i,j]} = \frac{\left(y_i - y_j\right)\left(w_j - w_i\right)'}{4}, \left(i=1,...,n;\ j=1,...,i\neq j\right)$$

since $(y_i + y_j)/2$ and $(w_i + w_j)/2$ are the sample means based on v_i and v_j, $(i \neq j)$ and the sample covariance matrix between y_i and w_j is derived as $(y_i - (y_i + y_j)/2)(w_j - (w_i + w_j)/2)'$. Note that $\hat{S}_{[i,j]} = \hat{S}_{[j,i]}$ for any i and j, $i \neq j$. Let $V^{[-i,-j]}$, $(i = 1, ..., n; j = 1, ..., n; i \neq j)$ be obtained by deleting v'_i and v'_j, $(i \neq j)$ from V. Furthermore, let $S_{ww}^{[-i-j]}, S_{yw}^{[-i-j]}$ and $S_{yy}^{[-i-j]}$ be derived by using $V^{[-i,-j]}$ and be based on the ordinary estimation method for covariance matrices. Then, if λ is fixed, $\hat{d}_\lambda^{[-i,-j]}$ is derived as the eigenvector that corresponds to the largest eigenvalue of $\left(S_{ww}^{[-i-j]}+\lambda P\right)^{-1}S_{yw}^{[-i-j]\prime}\left(S_{yy}^{[-i-j]}\right)^{-1}S_{yw}^{[-i-j]}$. Using $\hat{d}_\lambda^{[-i,-j]}$ and the largest eigenvalue $\left(\hat{\theta}_\lambda^{[-i,-j]}\right)^2$, $\hat{a}_\lambda^{[-i,-j]}$ is obtained as $\hat{a}_\lambda^{[-i,-j]} = \left(S_{yy}^{[-i-j]}\right)^{-1}S_{yw}^{[-i-j]}\hat{d}_\lambda^{[-i,-j]} / \hat{\theta}_\lambda^{[-i,-j]}$, where $\hat{\theta}_\lambda^{[-i,-j]} > 0$. Note that $\hat{a}_\lambda^{[-i,-j]}$ and $\hat{d}_\lambda^{[-i,-j]}$ are derived from $V^{[-i,-j]}$ and $\hat{S}_{[i,j]}$ is derived from v_i and v_j, $(i \neq j)$, which are not used for deriving $\hat{a}_\lambda^{[-i,-j]}$ and $\hat{d}_\lambda^{[-i,-j]}$. Thus, we can evaluate $\hat{a}_\lambda^{[-i,-j]}$ and $\hat{d}_\lambda^{[-i,-j]}$ based on $\hat{S}_{[i,j]}$. In order to optimize the penalty parameter λ and the other hyper parameters in the conversion function, we use $T = \Sigma_{i\neq j}\left|c_{ij}\right|$ where:

$$c_{ij} = \hat{a}_\lambda^{[-i,-j]\prime}\hat{S}_{[i,j]}\hat{d}_\lambda^{[-i,-j]}, \left(i=1,...,n; j=1,...,n;\ i\neq j\right) \tag{3.3}$$

Thus, for example, the penalty parameter λ and hyper parameter ζ in PNCCA and the conversion function can be optimized as $\hat{\lambda} = \arg\max_{\lambda\geq 0,\zeta} T$.

When we use more number of rows of V for making the alternative matrix for $S_{y^*w^*}$, we can extend this simple CV method to subset CV method. However, we only focus on the simple CV method in order to save the space of paper.

Numerical Study

In this section, we compare CCA, NCCA and PNCCA optimized with the proposed CV method through numerical study. Note that NCCA can be defined by the same form as PNCCA in (2.2) when we fix $\lambda = 0$. Let $\Delta_r(\rho)$ be an $r \times r$ matrix whose (i, j)th element is derived as $\rho^{|i-j|}$ The $n \times p_0$ matrix X is generated from $X = U\Delta_{p_0}\left(\rho_x\right)^{1/2}$, where U is an $n \times p_0$ matrix whose elements were generated independently from the standard normal distribution. Then, $Y = (y_1,..., y_n)'$ are derived as follows:

$$(A)\ \ y_i = \delta x_i'\ x_i 1_{qo} + \varepsilon_{i.q0},$$

$$(B)\ \ y_i = \delta \begin{pmatrix} x_i'\ x_i\ / \max\left(x_{ij}^2\right) \\ \sin\left(2x_i'\ x_i\right) \\ \cos\left(2x_i'\ x_i\right) \end{pmatrix} + \varepsilon_{i.3},$$

$$(C)\ \ y_i = \delta \begin{pmatrix} x_i'\ x_i\ / \max\left(x_{ij}^2\right) \\ \sin\left(2x_i'\ x_i\right) \\ \cos\left(2x_i'\ x_i\right) \\ \exp\left(-x_i'\ x_i\ / 4\right) \end{pmatrix} + \varepsilon_{i.4},$$

where, $x_i = \left(x_{i1},\ ...,\ x_{ip_0}\right)'$ is the ith row of the standardized X, 1_r is a r-dimensional vector all of whose elements are 1 and $\varepsilon_{i.r}$ is generated independently from $N_r(0_r, \Delta_r(0.5))$ which is a r-dimensional multivariate normal distribution with mean 0_r and covariance matrix $\Delta_r(0.5)$. Here, δ controls the scale of the nonlinear relationship part.

Since NCCA and PNCCA need the converted values that are expressed as $w_i = \left(w_{i1},\ ..., w_{ip_0}\right)' = \varphi\left(x_i\right), \left(i = 1,\ ...,\ n\right)$, we set $w_{ij} = \exp\left\{-x_{ij}^2 / (2h)\right\}$ and $W = (w_1, ..., w_n)'$ Then, W is standardized. We choose h by comparing the maximized correlation for each value {0.05, 0.1, 0.5, 1, 2, 5} in each repetition. In PNCCA, c P is set to $P = K'K$, where $K = \left(k_1,\ ..., k_{p_0-2}\right)'$ is a $(p_0-2) \times p_0$ matrix and $k_j = \left(0'_{j-1}, 1, -2, 1, 0'_{p_0-j-2}\right)', \left(j = 1, ..., p_0 - 2\right)$. (More details of K can be found in Green and Silverman (1994).) Since the 'arg max' operator is equivalent to the 'arg min' operator with the reversed sign, we select λ by using 'fminbnd' function in Matlab which is 'fminsearch' in Matlab with a specified region and we restrict the region to 1 to exp(20) in order to shorten the computation time. Furthermore, in order to reduce computational tasks, we calculate c_{ij} in (3.3) for $i = 1,...n-1$ and $j = i+1$.

In order to derive R^* in (3.1), since we need $\Sigma_{y^*y^*}$, $\Sigma_{y^*x^*}$, $\Sigma_{x^*x^*}$, $\Sigma_{y^*w^*}$ and $\Sigma_{w^*w^*}$, we set n = 10, 000 and generate X for each p_0 and ρ_x and standardize them. Then, from each transformation function (A), (B) and (C) and each parameter δ and q_0, we obtain Y, which we also standardized.

Note that $q_0 = 3$ when the transformation function is in (B) and $q_0 = 4$ when the transformation function is in (C). In CCA, $\Sigma_{y^*y^*}$, $\Sigma_{y^*x^*}$ and $\Sigma_{x^*x^*}$ can be derived as the sample variance matrix of the standardized Y, the sample covariance matrix of the standardized Y and X and the sample variance matrix of the standardized X, respectively. In NCCA and PNCCA, we convert X as above for each h and standardize the converted values. The results of conversion is derived as W. Then, $\Sigma_{y^*w^*}$ and $\Sigma_{w^*w^*}$ can be derived as the sample covariance matrix of standardized Y and W and the sample variance matrix of the standardized W, respectively. Using these matrices, we evaluated the results of each method.

In order to evaluate these methods, we fixed X and generated Y for 1, 000 repetitions. We used the standardized X, Y and W in each repetition. For each repetition with CCA, we obtain S_{yx}, S_{yy} and S_{xx}. On the other hand, for each repetition with NCCA and PNCCA, we obtain S_{yw}, S_{yy} and S_{ww}.

For each repetition with CCA in (1.1), we calculated the maximized correlation under certain conditions by using S_{yx}, S_{yy} and S_{xx} instead of Σ_{yx}, Σ_{yy} and Σ_{xx}, respectively. We denote the maximized correlation as $\hat{\theta}^2$, the eigenvector that corresponds to the largest eigenvalue of $S_{xx}^{-1} S'_{yx} S_{yy}^{-1} S_{yx}$, as $\hat{b}_C$ and then $\hat{a}_c = S_{yy}^{-1} S_{yx} \hat{b}_c / \hat{\theta}$ is derived where $\hat{\theta} > 0$. For each repetition with NCCA which can be defined as (2.2) with λ = 0, we calculated the maximized correlation under certain conditions and the optimized h, for which we used S_{yw}, S_{yy} and S_{ww} instead of Σ_{yw}, Σ_{yy} and Σ_{ww}, respectively. We denote the maximized correlation as $\hat{\eta}_0^2$, the eigenvector that corresponds to the largest eigenvalue of $S_{ww}^{-1} S'_{yw} S_{yy}^{-1} S_{yw}$ as $\hat{d}_N$ and then $\hat{a}_N = S_{yy}^{-1} S_{yw} \hat{d}_N / \hat{\eta}_0$ is derived where $\hat{\eta}_0 > 0$. For each repetition with PNCCA in (2.2), we calculated the maximized correlation under certain conditions by using the optimized λ and optimized h and we used S_{yw}, S_{yy} and S_{ww} instead of Σ_{yw}, Σ_{yy} and Σ_{ww}, respectively. We denote the maximized correlation as $\hat{\eta}_{\hat{\lambda}}^2$ and the eigenvector that corresponds to the largest eigenvalue of $(S_{ww} + \hat{\lambda} P)^{-1} S'_{yw} S_{yy}^{-1} S_{yw}$ as $\hat{d}_p$, where $\hat{\lambda}$ is the optimized penalty parameter based on the proposed CV method and then $\hat{a}_P = S_{yy}^{-1} S_{yw} \hat{d}_P / \hat{\eta}_{\hat{\lambda}}$ is derived where $\hat{\eta}_{\hat{\lambda}} > 0$

Note that we considered the evaluating function in (3.1) in order to optimize λ based on the predictive values. Thus we also compared these methods by using the average values of $\hat{a}'_C \Sigma_{y^*x^*} \hat{b}_C, \hat{a}'_N \Sigma_{y^*w^*} \hat{d}_N$, and $\hat{a}'_P \Sigma_{y^*w^*} \hat{d}_P$ and then we denoted the average value of each value as R_C^*, R_N^* and R_P^* in Table 1-5. The reason of using R_N^* and R_P^* is that the purposes of the corresponding method are finding the nonlinear relationship. In Table 1 to 5, the bold and italic faces mean the biggest and second biggest values, respectively, in each situation.

First, we consider the results when using pattern (A), which are presented in Table 1-3. When ρ_x becomes large, the result values of CCA become small, the results of PNCCA become large and the result values of NCCA also become large except when (n, p_0) = (30, 5) and (n, p_0) = (30, 8).

The result values of each method become large in almost all cases when δ becomes large except when (n, p_0) = (100, 3). In this pattern, we can change q_0. Thus, next, we consider the result values when q_0 changes. When q_0 changes from 3 to 8, the result values of NCCA become large in almost cases. When q_0 becomes large, the result values of PNCCA become large in almost situations

in $(n, p_0, \delta) = (30, 3, 1)$, $(n, p_0, \delta) = (50, 3, 1)$ and $(n, p_0, \delta) = (100, 3, 1)$ and that of PNCCA become small in almost situations in $(n, p_0, \delta) = (30, 3, 3)$, $(n, p_0) = (30, 5)$ and $(n, p_0) = (30, 8)$. Next, we consider the results when p_0 becomes large. In $n = 50$ and $n = 100$, the result values of NCCA and PNCCA become large when p_0 becomes large. The result values of PNCCA also become large when $n = 30$ and p_0 becomes large. In this connection, we focus on the results when n becomes large.

When n changes from 30 to 50, the result values of CCA almost all become small in $(p_0, q_0) = (3, 5)$ and $(p_0, q_0) = (3, 8)$, the result values of NCCA become large and the result values of PNCCA become large when $p_0 = 3$ and $p_0 = 5$. When $(p_0, q_0) = (8, 5)$ and $(p_0, q_0) = (8, 8)$ and n changes from 30 to 50, the result values of PNCCA almost all become small. The result values of NCCA also become small when n changes from 30 to 50 in almost all situations when $(p_0, \rho_x) = (8, 0.8)$.

Table 1. Average values of R_C^* (CCA), R_N^* (NCCA) and R_P^* (PNCCA) for $n = 30$ and (A)

			$p_0 = 3$			$p_0 = 5$			$p_0 = 8$		
q_0	δ	ρ_x	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*
3	1	0.5	0.0055	*1.0848*	1.0888	0.0028	*1.0723*	1.2443	0.0070	*1.4206*	1.6244
		0.8	0.0029	*1.6265*	1.6713	0.0017	*1.6642*	2.4493	0.0020	*1.5904*	3.5193
		0.95	0.0021	2.3404	*2.2292*	0.0007	*0.8485*	3.6166	0.0010	*0.5696*	5.8488
	3	0.5	0.0156	*1.1349*	1.1383	0.0033	*1.0898*	1.2380	0.0063	*1.4944*	1.6558
		0.8	0.0019	*1.6746*	1.6993	0.0027	*1.6857*	2.4588	0.0032	*1.6175*	3.5545
		0.95	0.0012	2.4598	*2.2841*	0.0014	*0.8380*	3.6303	0.0006	*0.5747*	5.9185
5	1	0.5	0.0101	1.0912	*1.0904*	0.0040	*1.0728*	1.2353	0.0048	*1.4135*	1.6298
		0.8	0.0027	*1.6389*	1.6753	0.0026	*1.6435*	2.4555	0.0024	*1.5689*	3.4765
		0.95	0.0013	2.3832	*2.2656*	0.0009	*0.8670*	3.5930	0.0006	*0.5830*	5.7756
	3	0.5	0.0075	*1.1186*	1.1206	0.0041	*1.0938*	1.2437	0.0049	*1.4588*	1.6316
		0.8	0.0052	*1.6788*	1.7040	0.0024	*1.6680*	2.4355	0.0022	*1.6021*	3.5511
		0.95	0.0010	2.4589	*2.3020*	0.0007	*0.8388*	3.5998	0.0009	*0.5777*	5.8660
8 1		0.5	0.0068	1.0992	*1.0367*	0.0051	*1.0519*	1.2061	0.0052	*1.3710*	1.5844
		0.8	0.0055	*1.6640*	1.6981	0.0032	*1.6006*	2.4153	0.0032	*1.5513*	3.4621
		0.95	0.0017	2.3961	*2.2814*	0.0008	*0.8437*	3.5642	0.0007	*0.5925*	5.6960
	3	0.5	0.0065	1.1145	*1.0829*	0.0045	*1.0623*	1.2186	0.0052	*1.3998*	1.5976
		0.8	0.0037	*1.6753*	1.6990	0.0024	*1.6347*	2.4313	0.0024	*1.5424*	3.4618
		0.95	0.0020	2.4499	*2.2716*	0.0008	*0.8474*	3.5708	0.0005	*0.5795*	5.7208

Table 2. Average values of R_C^* (CCA), R_N^* (NCCA) and R_P^* (PNCCA) for n = 50 and (A)

			$p_0 = 3$			$p_0 = 5$			$p_0 = 8$		
q_0	δ	ρ_x	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*
3	1	0.5	0.0046	1.1368	*1.1325*	0.0055	1.3697	*1.3605*	0.0042	*1.5363*	1.5706
		0.8	0.0028	1.8384	*1.8374*	0.0020	*2.6066*	2.6390	0.0025	*3.4349*	3.5174
		0.95	0.0025	*2.4768*	2.4821	0.0007	*3.5793*	4.0919	0.0006	*5.1110*	6.2173
	3	0.5	0.0070	1.1820	*1.1759*	0.0037	1.3715	*1.3574*	0.0061	*1.5809*	1.6030
		0.8	0.0042	1.8803	*1.8716*	0.0038	2.6648	*2.6590*	0.0019	*3.5191*	3.5492
		0.95	0.0011	2.5765	*2.5221*	0.0007	*3.7883*	4.1181	0.0004	*5.4452*	6.2972
5	1	0.5	0.0069	1.1455	*1.1404*	0.0055	1.3621	*1.3533*	0.0045	*1.5577*	1.5855
		0.8	0.0032	1.8480	*1.8430*	0.0025	*2.6401*	2.6667	0.0021	*3.4211*	3.4885
		0.95	0.0013	*2.5102*	2.5123	0.0009	*3.5909*	4.1073	0.0005	*5.1608*	6.2080
	3	0.5	0.0074	1.1671	*1.1607*	0.0047	1.3896	*1.3751*	0.0053	*1.5632*	1.5857
		0.8	0.0041	1.8859	*1.8769*	0.0032	2.6548	*2.6521*	0.0019	*3.5276*	3.5593
		0.95	0.0013	2.5805	*2.5345*	0.0007	*3.7451*	4.1177	0.0006	*5.4121*	6.2852
8	1	0.5	0.0060	1.1588	*1.1532*	0.0076	1.3608	*1.3512*	0.0051	*1.5440*	1.5711
		0.8	0.0043	1.8764	*1.8695*	0.0028	*2.6122*	2.6418	0.0024	*3.4513*	3.5162
		0.95	0.0016	2.5381	*2.5289*	0.0008	*3.563*	4.0960	0.0006	*5.1504*	6.2035
	3	0.5	0.0063	1.1669	*1.1596*	0.0057	1.3798	*1.3646*	0.0052	*1.5522*	1.5769
		0.8	0.0039	1.8831	*1.8726*	0.0028	*2.6527*	2.6566	0.0023	*3.4788*	3.5145
		0.95	0.0018	2.5776	*2.5306*	0.0008	*3.7039*	4.1246	0.0005	*5.3411*	6.2406

Table 3. Average values of R_C^* (CCA), R_N^* (NCCA) and R_P^* (PNCCA) for $n = 100$ and (A)

			$p_0 = 3$			$p_0 = 5$			$p_0 = 8$		
q_0	δ	ρ_x	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*
3	1	0.5	0.0055	*1.2018*	1.2057	0.0032	*1.5306*	1.5352	0.0041	*1.5487*	1.5695
		0.8	0.0029	*1.9092*	1.9138	0.0021	*2.8659*	2.8950	0.0023	*3.2276*	3.2883
		0.95	0.0021	2.5552	*2.539*	60.0007	*3.9258*	4.2784	0.0007	*5.5941*	6.0614
	3	0.5	0.0067	*1.2531*	1.2561	0.0043	*1.5357*	1.5381	0.0065	*1.5865*	1.6023
		0.8	0.0045	1.9513	*1.9497*	0.0029	*2.9057*	2.9187	0.0023	*3.2731*	3.3141
		0.95	0.0014	2.6314	*2.5962*	0.0006	*4.0318*	4.3034	0.0005	*5.7878*	6.1337
5	1	0.5	0.0062	*1.2067*	1.2105	0.0047	*1.5263*	1.5302	0.0051	*1.5673*	1.5850
		0.8	0.0033	*1.9120*	1.9155	0.0024	*2.9000*	2.9256	0.0021	*3.2101*	3.2612
		0.95	0.0015	2.5791	*2.5716*	0.0008	*3.9569*	4.2926	0.0005	*5.6221*	6.0548
	3	0.5	0.0068	*1.2363*	1.2393	0.0042	*1.5575*	1.5604	0.0055	*1.5703*	1.5851
		0.8	0.0034	1.9561	*1.9548*	0.0028	*2.9022*	2.9150	0.0019	*3.2823*	3.3243
		0.95	0.0017	2.6327	*2.6061*	0.0007	*4.0373*	4.3078	0.0007	*5.7813*	6.1237
8	1	0.5	0.0062	*1.2187*	1.2221	0.0061	*1.5266*	1.5304	0.0053	*1.5545*	1.5720
		0.8	0.0041	*1.9385*	1.9404	0.0027	*2.8822*	2.9037	0.0024	*3.2381*	3.2901
		0.95	0.0020	2.5963	*2.5784*	0.0008	*3.9735*	4.2905	0.0006	*5.6429*	6.0518
	3	0.5	0.0060	*1.2348*	1.2378	0.0056	*1.5482*	1.5508	0.0048	*1.5607*	1.5771
		0.8	0.0043	1.9512	*1.9502*	0.0029	*2.9138*	2.9264	0.0024	*3.2405*	3.2813
		0.95	0.0016	2.6274	*2.5991*	0.0009	*4.0515*	4.3232	0.0005	*5.7381*	6.0822

Table 4. Average values of R_C^* (CCA), R_N^* (NCCA) and R_P^* (PNCCA) for (B)

			$p_0 = 3$			$p_0 = 5$			$p_0 = 8$		
n	δ	ρ_x	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*
30	1	0.5	0.0050	*0.1803*	0.1924	0.0060	*0.1426*	0.2362	0.0042	*0.1415*	0.3359
		0.8	0.0029	*0.3026*	0.3593	0.0017	*0.1841*	0.5203	0.0020	*0.2017*	1.0257
		0.95	0.0027	*0.570*	0.7467	0.0014	*0.2478*	1.3996	0.0005	*0.1366*	2.7816
	3	0.5	0.0088	0.5705	*0.5645*	0.0025	*0.5279*	0.6730	0.0049	*0.8081*	1.0558
		0.8	0.0015	*0.9216*	0.9264	0.0036	*0.8489*	1.5573	0.0026	*0.9735*	2.3869
		0.95	0.0038	1.5283	*1.5091*	0.0025	*0.6209*	2.5992	0.0007	*0.3815*	4.4913
50	1	0.5	0.0051	*0.1904*	0.1936	0.0048	*0.1715*	0.2568	0.0051	*0.2230*	0.3839
		0.8	0.0032	*0.3164*	0.4315	0.0020	*0.3590*	0.6171	0.0021	*0.5467*	1.1514
		0.95	0.0023	*0.4452*	0.8387	0.0011	*0.6805*	1.7296	0.0005	*0.6244*	3.3308
	3	0.5	0.0053	*0.6274*	0.6285	0.0027	*0.6998*	0.7309	0.0033	*0.9603*	1.0420
		0.8	0.0025	*1.0326*	1.1067	0.0060	*1.5704*	1.7519	0.0012	*2.1235*	2.4541
		0.95	0.0011	*1.4164*	1.6434	0.0010	*2.2205*	2.9867	0.0005	*2.8972*	5.0905
100	1	0.5	0.0051	*0.2578*	0.2644	0.0075	*0.2691*	0.3349	0.0037	*0.2601*	0.3816
		0.8	0.0030	*0.4277*	0.5193	0.0014	*0.4836*	0.7042	0.0019	*0.6373*	1.0948
		0.95	0.0014	*0.5038*	0.9007	0.0007	*0.8114*	1.7327	0.0005	*0.9769*	3.3225
	3	0.5	0.0044	*0.6661*	0.6800	0.0030	*0.7958*	0.8217	0.0029	*0.9588*	1.0304
		0.8	0.0017	*1.1146*	1.1567	0.0022	*1.8378*	1.9598	0.0017	*2.0936*	2.3122
		0.95	0.0010	*1.3845*	1.6391	0.0006	*2.4001*	3.1127	0.0005	*3.3980*	4.9799

Table 5. Average values of R_C^* (CCA), R_N^* (NCCA) and R_P^* (PNCCA) for (C)

			$p_0 = 3$			$p_0 = 5$			$p_0 = 8$		
n	δ	ρ_x	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*	R_C^*	R_N^*	R_P^*
30	1	0.5	0.0074	0.2507	*0.2479*	0.0047	*0.1690*	0.2528	0.0050	*0.1576*	0.4187
		0.8	0.0063	*0.3503*	0.4134	0.0016	*0.1914*	0.5353	0.0032	*0.2065*	1.1178
		0.95	0.0016	*0.5057*	0.6864	0.0012	*0.2715*	1.4164	0.0005	*0.1110*	2.3199
	3	0.5	0.0082	0.6628	*0.6571*	0.0069	*0.6211*	0.7752	0.0046	*0.8311*	1.0931
		0.8	0.0023	*1.1161*	1.1373	0.0028	*0.8099*	1.4641	0.0019	*0.9546*	2.3910
		0.95	0.0012	*1.4540*	1.4771	0.0011	*0.6674*	2.5052	0.0006	*0.3961*	4.7958
50	1	0.5	0.0064	0.3006	*0.2786*	0.0051	*0.2095*	0.3011	0.0045	*0.2809*	0.4944
		0.8	0.0049	*0.3943*	0.5459	0.0028	*0.4195*	0.7251	0.0018	*0.6204*	1.2818
		0.95	0.0012	*0.4479*	0.8000	0.0020	*0.6926*	1.7492	0.0005	*0.5374*	2.7939
	3	0.5	0.0067	0.7499	*0.7488*	0.0053	*0.8343*	0.8708	0.0027	*0.9939*	1.0766
		0.8	0.0029	*1.3193*	1.3724	0.0029	*1.5655*	1.7197	0.0019	*2.1370*	2.4582
		0.95	0.0009	*1.5079*	1.7138	0.0008	*2.2223*	2.9143	0.0004	*3.1285*	5.4998
100	1	0.5	0.0057	*0.3862*	0.3951	0.0053	*0.3193*	0.3888	0.0051	*0.3464*	0.5038

Table 5. Continue

	0.8	0.0044	*0.5380*	0.6380	0.0015	*0.6137*	0.8462	0.0023	*0.7448*	1.2510
	0.95	0.0013	*0.5335*	0.8878	0.0007	*0.8215*	1.8003	0.0005	*0.8568*	2.8302
3	0.5	0.0083	*0.8072*	0.8198	0.0062	*0.9455*	0.9823	0.0036	*1.0074*	1.0798
	0.8	0.0025	*1.3925*	1.4396	0.0029	*1.7927*	1.8991	0.0012	*2.1214*	2.3359
	0.95	0.0008	*1.6185*	1.7614	0.0006	*2.4415*	3.0633	0.0004	*3.8614*	5.3736

When n changes from 50 to 100 and p_0 = 8, the result values of PNCCA almost always become small. The result values of PNCCA also become small when n changes from 30 to 100 except when (p_0, ρ_x) = (8, 0.95). The result values of NCCA become large when n changes from 30 to 100. Next, we consider the results when using pattern (B), which are presented in Table 4. When ρ_x becomes large, the result values of CCA become small when (p_0, δ) = (5, 1), p_0 = 8 except when (n, δ, p_0) = (30, 3, 3). When ρ_x becomes large, the result values of NCCA become large when p_0 = 3, (n, p_0) = (100, 8) and (n, δ, p_0) = (50, 3, 8) but not when (n, δ, p_0) = (30, 1, 5). The result values of PNCCA become large when ρ_x becomes large. The result values of NCCA and PNCCA become large when δ becomes large. When δ becomes large, the result values of CCA also become large When (n, p_0) = (30, 8) and that of CCA become small when (n, p_0) = (50, 8) and (n, p_0) = (100, 8).

When p_0 = 5 and δ becomes large, the result values of CCA also almost all become small, except when ρ_x = 0.8. Next, we compare the result values when p_0 and n both become large. When p_0 becomes large, the result values of PNCCA become large. The result values of CCA become small when p_0 changes from 3 to 5 except when (δ, ρ_x) = (3, 0.8). When p_0 changes from 3 to 8, the result values of CCA almost all become small. The result values of NCCA become large when p_0 changes from 3 to 5 in (n, δ) = (50, 3) and (n, δ) = (100, 3) and when p_0 changes from 3 to 8 in n = 50 and n = 100. When p_0 changes from 5 to 8 and δ = 3, the result values of NCCA almost all become large. In contrast to this, the result values of CCA become small when p_0 changes from 3 to 5 in n = 30. Moreover, when n changes from 50 to 100 and p_0 = 3 and it changes from 30 to 100 and p_0 = 8, the result values of CCA become small. The result values of NCCA become large except when (p_0, ρ_x) = (3, 0.95) and (p_0, δ) = (8, 3), when n changes from 50 to 100. The result values of PNCCA almost always become large when n becomes large and p_0 = 3 and p_0 = 5. When (p_0, δ) = (8, 1) and n changes from 30 to 50 and 30 to 100, the result values of PNCCA also become large. When n changes from 50 to 100 and p_0 = 8, the result values of PNCCA become small.

Finally, we consider the results with pattern (C), which are in Table 5. When ρ_x becomes large, the result values of CCA and PNCCA become small and large, respectively. When ρ_x becomes large, the result values of NCCA almost always become large in p_0 = 3, p_0 = 5, (n, p_0, δ) = (50, 8, 3) and (n, p_0) = (100, 8). When δ becomes large, the result values of NCCA and PNCCA become large. When δ becomes large, the result values of CCA become small when (n, p_0) = (100, 8) and p_0 = 3, but not when (p_0, ρ_x) = (3, 0.5). Next, we compare the results when p_0 becomes large and n becomes large. When ρ_x becomes large, the result values of PNCCA almost always become large. The result values of NCCA become small when p_0 changes from 3 to 5 and n = 30. Moreover, when n becomes large, the result values of NCCA and PNCCA almost always become large.

Based on these results, we recommend using PNCCA with the proposed optimization method in order to find a nonlinear relationship.

Conclusion

In the present paper, we considered finding a nonlinear relationship between random vectors. CCA (Hotelling, 1936) can find only a linear relationship between random vectors, based on the correlation between linear combinations of them. The use of conversion functions allows a nonlinear relationship to be found by using CCA on the converted variables. Hardoon *et al.* (2004) pointed out that this method has a critical issue and, to avoid this, Akaho (2000) proposed PNCCA when the conversion functions are the kernel functions. Although the result of PNCCA heavily depends on the penalty and other hyper parameters in the conversion function, there has been no optimization methods proposed for them until the present paper. The reason for this is that the evaluation method for the covariance matrix is not defined.

In order to optimize the penalty and other parameters in PNCCA, we considered the evaluating function in (3.1) and proposed using the simple CV method in Section 3. Using the two samples $\{y_i, w_i\}$ and $\{y_j, w_j\}$, where $i \neq j$, we define $\hat{S}_{[i,j]}$ for all i and j, ($i \neq j$). On the other hands, for the fixed parameters, the results of PNCCA are derived based on the remains datum. Using $\hat{S}_{[i,j]}$ and the results of PNCCA for each i and j, ($i \neq j$), we then proposed the optimization method for the penalty and other hyper parameters in the conversion function based on the sum of evaluation (3.3). We can easily extend this method for subset CV method. Our numerical studies showed that PNCCA is almost always the best of the three we tested. Thus, we recommend using PNCCA, optimized by using the proposed simple CV method.

Acknowledgment

I would like to express my deepest gratitude to Prof. Hirokazu Yanagihara of Hiroshima University for his valuable comments.

Ethics

We consider there are no problem about any ethical issues.

Appendix: Using Proposed CV Method to Select Variables in *y* and *x*

Using the optimized penalty parameter $\hat{\lambda}$, the maximized value of $a'\Sigma_{yw}d$ in (2.2) is estimated by using $\hat{\eta}_{\hat{\lambda}}$, which coincides with the square root of the largest eigenvalue of $\left(S_{ww}+\hat{\lambda}P\right)^{-1}S'_{yw}S_{yy}^{-1}S_{yw}$. In this section, we illustrate the variable selection method.

Let $y^{[1]}$ and $x^{[1]}$ be subsets of y and x, respectively and $w^{[1]} = \psi(x^{[1]})$, where $\psi(\cdot)$ is any known conversion function that does not need to correspond with $\varphi(\cdot)$. Let $S_{w[1]w[1]}$, $S_{y[1]w[1]}$ and $S_{y[1]y[1]}$ be the sample variance and covariance matrices of $w^{[1]}$, $y^{[1]}$ and $w^{[1]}$ and $y^{[1]}$, respectively and let $P^{[1]}$ be some known nonnegative penalty matrix. Based on the proposed simple CV method, the optimized penalty parameter $\hat{\lambda}^{[1]}$ is derived. We can then obtain $\left(\hat{\eta}^{[1]}_{\hat{\lambda}^{[1]}}\right)^2$, which is the estimator of the maximized correlation between the linear combinations of $y^{[1]}$ and $w^{[1]}$.

Next, $\left(\hat{\eta}^{[2]}_{\hat{\lambda}^{[2]}}\right)^2$ is derived using the same procedure as in PNCCA and the above procedure based on $y^{[2]}$ and $x^{[2]}$, where $y^{[2]}$ and $x^{[2]}$ are also subsets of y and x but are not the same as $y^{[1]}$ and $x^{[1]}$. If it holds that $\hat{\eta}^{[1]}_{\hat{\lambda}^{[1]}} > \hat{\eta}^{[2]}_{\hat{\lambda}^{[2]}}$, where $\hat{\eta}^{[1]}_{\hat{\lambda}^{[1]}} > 0$ and $\hat{\eta}^{[2]}_{\hat{\lambda}^{[2]}} > 0$, we select $y^{[1]}$ and $x^{[1]}$; $y^{[2]}$ and $x^{[2]}$ are selected if it does not hold.

Since we evaluate the covariance matrix by using the subset CV method, we conjecture that we may select another statistical estimation method based on the covariance matrix.

References

Akaho, S., 2000. A kernel method for canonical correlation analysis.

Doeswijk, T.G., J.A. Hageman, J.A. Westerhuis, Y. Tikunov and Y.M. Bovy *et al*., 2011. Canonical correlation analysis of multiple sensory directed metabolomics data blocks reveals corresponding parts between data blocks. Chemometr. Intell. Lab. Syst., 107: 371-176. DOI: 10.1016/j.chemolab.2011.05.010

Gittins, R., 1985. Canonical Analysis.

Green, P.J. and B.W. Silverman, 1994. Nonparametric regression and generalized linear models. Chapman and Hall/CRC.

Hardoon, D.R., S. Szedmak and J. Shawe-Taylor, 2004. Canonical correlation analysis: An overview with application to learning methods. Neural Comput., 16: 2639-2664. DOI: 10.1162/0899766042321814

Hotelling, H., 1936. Relations between two sets of variates. Biometrika, 28: 321-377. DOI: 10.1093/biomet/28.3-4.321

Lütkepohl, K., 1996. Handbook of Matrices. 1st Edn., John Wiley Sons, New York, ISBN-10: 0471970158, pp: 304.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory, 1st Edn., John Wiley Sons, New York ISBN-10: 0471094420, pp: 704.

Srivastava, M.S., 2002. Methods of Multivariate Statistics. 1st Edn., John Wiley and Sons, New York, ISBN-10: 0471223816, pp: 728.

Weenink, D., 2003. Canonical Correlation Analysis. Institute of Phonetic Sciences, University of Amsterdam.