Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5545061, 9 pages
https://doi.org/10.1155/2021/5545061

Research Article

The Simplified Expression of Machine Learning and Multivariate Statistical Analysis Based on the Centering Matrix

Zhen Xue and Liangliang Zhang

School of Science, North University of China, Taiyuan, Shanxi 030051, China

Correspondence should be addressed to Zhen Xue; xuezhen@nuc.edu.cn

Received 31 January 2021; Revised 24 March 2021; Accepted 26 March 2021; Published 28 April 2021

Academic Editor: Maria Patrizia Pera

Copyright © 2021 Zhen Xue and Liangliang Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In machine learning (ML) algorithms and multivariate statistical analysis (MVA) problems, it is usually necessary to center (zero centered or mean subtraction) the original data. The centering matrix plays an important role in this process. The full consideration and use of its properties may contribute to the speed or stability improvement of some related ML algorithms. Therefore, in this paper, we discussed in detail the properties of the centering matrix, proved some previously known properties, and deduced some new properties. The involved properties mainly consisted of the centering, quadratic form, spectral decomposition, null space, projection, exchangeability, Kronecker square, and so on. Based on this, we explored the simplified expression role of the centering matrix in the principal component analysis (PCA) and regression analysis theory. The results show that the sum of deviation squares which is widely used in regression theory and variance analysis can be expressed by the quadratic form with the centering matrix as the kernel matrix. The ML algorithms introducing the centering matrix can greatly simplify the learning process and improve predictive ability.

1. Introduction

Among the regression problems of machine learning (ML) and during the course of training the neural networks, we usually need to center (zero centered or mean subtraction) the original data. The centering matrix can achieve this goal. It is a symmetric idempotent matrix and has been successfully applied in various fields, such as transfer learning [1, 2], feature learning [3], extreme learning [4], ensemble learning [5–9], dictionary learning [10, 11], and multivariate statistical analysis (MVA) [12–16]. The practical applications involve a wide range of aspects (e.g., data-driven fault diagnosis and prognosis [3, 17, 18], sentiment analysis [7, 8], web page classification [9], information retrieval [4, 19, 20], image denoising [10], and signal processing of electronics [11], theoretical chemistry and graph theory [21], and rank data analysis [22]). For instance, Chen S. Z. et al.[4] proposed a flexible ranking extreme learning machine (ELM) method based on matrix-centering transformation to replace the traditional graph Laplacian matrix-based methods. In [19, 20], T. Pahikkala et al. presented two kinds of ranking algorithm for information retrieval through introducing the centering matrix to construct block diagonal matrix as the weighted matrix and finding the linear solution of Rank Reweighted Least Squares (RankRLS). The application of the centering matrix in the optimization problem of transfer learning is discussed in [1, 2]. Long M. S. et al.[2] established an optimization problem and then proposed a feature transformation by adopting the centering matrix. In a recent study, the covariance matrices of the source and target features were expressed by the centering matrix [3]. Wang Q. and Zhang L. investigated the online updating of the generalized inverse of the centered matrix of the data matrix in [5, 6]. In [23], Liu L. P. et al. pointed out the least square solution to linear discriminant analysis (LDA) [24] is the multiplication of pseudoinverse of the centered data matrix and the indicator matrix, which makes the solution being updated without eigenanalysis. In another study, Sun L. et al.[25] investigated the relationship between the generalized eigenvalue problem and least squares problem by

using the centering matrix. In [26], the application of the centering matrix in data dimension reduction algorithm of principal component analysis (PCA) [27, 28] is discussed. The application in statistical analysis of the centering matrix is addressed in [12–16]. In [21], Gutman I. and Xiao W. examined the generalized inverse of the Laplacian matrix of a connected graph. One of the conclusions obtained by them was the Laplacian matrix which commutes with its generalized inverse, and their product is the centering matrix. Recently, a MATLAB toolbox data-based key-performance-indicator oriented fault detection toolbox (DB-KIT) was developed for prognosis and fault diagnosis of the complex systems in [17] and the canonical variate dissimilarity analysis was proposed for process incipient fault detection in [18]. Both studies in which the centering matrix plays an important role greatly promoted the development of the data-driven fault diagnosis and prognosis in the modern industry.

With regard to the centering matrix, more consideration and use of its properties and its matrix notation simplified role may contribute to the speed or stability improvement of some related ML algorithms. For this purpose, in this paper, we firstly discussed in detail and established some properties of the centering matrix and drew a mind map to clearly illustrate the relationship between these properties and their potential role in ML and MVA. Then, we explored the theoretical simplification applications of the centering matrix in ML and MVA theory by two examples. The research results can provide help for the proposal or speed improvement of some related ML algorithms on the regression problem or neural network algorithms.

The centering matrix is defined as follows [13].

Definition 1. $C_n = I_n - (1/n)J_n = I_n - (1/n)1_n 1_n^T$ is called the centering matrix of order n, where I_n is the identity matrix of order n, $1_n = (1, 1, \ldots, 1)^T$ is the column vector of n ones, J_n is a square matrix of order n with all elements unity, and $1_n {1_n}^T = J_n$ holds.

For example, $C_3 = \begin{pmatrix} 2/3 & -(1/3) & -(1/3) \\ -(1/3) & 2/3 & -(1/3) \\ -(1/3) & -(1/3) & 2/3 \end{pmatrix}$ is a centering matrix of order 3.

The rest of this paper is organized as follows. Section 2 discusses the properties of the centering matrix. Section 3 presents the main application of the centering matrix in ML and MVA. Finally, we conclude this paper in Section 4.

2. Properties of the Centering Matrix

It can be proved that the centering matrix has the following important properties. In what follows, C_n will always represent the centering matrix, unless otherwise stated.

Lemma 1 (centering property, see [13], and we prove it). *For any data vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ of order n, when performing a left multiplication C_n on $\mathbf{x}$, denoted by $C_n x$ (or x_C), it has the same effect as subtracting the mean of the elements of $\mathbf{x}$ from each element, i.e., $\mathbf{C}_n\mathbf{x} = (x_1 - \overline{x}, x_2 - \overline{x}, \ldots, x_n - \overline{x})^T$, where $\overline{x} = (1/n)\sum_{k=1}^{n} x_k$.*

Proof.

$$\begin{aligned}
\mathbf{C}_n\mathbf{x} = \left(I_n - \frac{1}{n}J_n\right)x &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \frac{1}{n}\begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \left(\frac{1}{n}\sum_{k=1}^{n} x_k \frac{1}{n}\sum_{k=1}^{n} x_k \vdots \frac{1}{n}\sum_{k=1}^{n} x_k\right) = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \mathbf{1}_n. \\
&= x - \bar{x}
\end{aligned} \tag{1}$$

This property illustrates that the data vector can be centered through left multiplication by C_n. This also reveals the origin of the name centering matrix for C_n. Meanwhile, $C_n x$ is called the centered vector of $\mathbf{x}$. It is easy to know that the sum of all entries of the centered vector is 0. □

Corollary 1. *Let X be a data matrix of order $n \times m$, each column represents the n observations on a particular variable and each row represents a single observation on m variables. Then, the centered matrix of X is $C_n X$.*

Proof. X can be partitioned as $X = (x_1, x_2, \ldots, x_m)$, and each column is the data vector $\mathbf{x}_j$, $j = 1, 2, \ldots, m$. We can obtain that its centered vector is $C_n x_j$ by Lemma 1. So, the centered matrix of X is $C_n X$.

Multiplication by the centering matrix can be used to remove not only the mean of a single vector but also multiple vectors stored in the rows or columns of a matrix. For any matrix of $n \times m$, the multiplication $C_n X$ removes the means from each of the m columns, while the multiplication XC_m removes the means from each of the n rows [14].

It can be easily verified that the centering matrix C_n is symmetric. Since

$$C_n^2 = \left(I_n - \frac{1}{n}J_n\right)\left(I_n - \frac{1}{n}J_n\right) = I_n^2 - \frac{1}{n}I_n J_n - \frac{1}{n}J_n I_n + \frac{1}{n^2}J_n^2,$$
$$= I_n - \frac{2}{n}J_n + \frac{1}{n^2}nJ_n = I_n - \frac{1}{n}J_n = C_n, \tag{2}$$

the centering matrix C_n is idempotent, i.e.,$C_n^2 = C_n$. The extension form is $C_n^k = C_n, k = 1, 2, 3 \ldots$.

It can be seen from the above two paragraphs that C_n is symmetric and idempotent. Therefore, C_n is a projection matrix. It can be easily proved that $I_n - C_n$ is also a projection matrix.

Let the adjoint matrix of C_n be denoted by C_n^*. Since $(C_n^*)^T = (C_n^T)^* = C_n^*$ and $(C_n^*)^2 = C_n^* C_n^* = (C_n C_n)^* = (C_n^2)^* = C_n^*$, C_n^* is also a projection matrix. □

Theorem 1 (quadratic form property, see [13]). *Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ be a vector with a set of data $x_1, x_2, \ldots, x_n$ as entries. Then, the covariance or total sum of squares (SST) of the data can be expressed by a quadratic form whose kernel matrix is the centering matrix C_n, i.e.,* $\mathrm{SST} = \sum_{i=1}^n (x_i - \bar{x})^2 = x^T C_n x$.

Proof. By Lemma 1, $C_n x = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})^T$ can be obtained. Then, we calculate the inner product of the vector with itself:

$$(C_n x)^T C_n x = \left(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}\right)\left(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}\right)^T = \sum_{i=1}^n \left(x_i - \bar{x}\right)^2. \tag{3}$$

Using the symmetry and idempotency of C_n, we can obtain the equation $x^T C_n x = \sum_{i=1}^n (x_i - \bar{x})^2 = \mathrm{SST}$.

The centering matrix has close relationship with the Helmert matrix which is an orthogonal matrix and has important application in the defining contrasts in general linear models of statistics. The corresponding conclusion is as follows. □

Proposition 1 (see [12, 29], and we prove it). *Let the Helmert matrix of order n be partitioned as $H = \begin{pmatrix} h^T \\ K \end{pmatrix}$, where $h = n^{-(1/2)} 1_n^T$, K is a matrix consisting of the last $n-1$ rows of H, and each row of K is given by the following formula:*

$$\frac{1}{\sqrt{\lambda_i}}\left(1_i^T, -i, 0_{n-i-1}^T\right), \quad \lambda_i = i(i+1), i = 1, 2, \ldots, n-1, \tag{4}$$

where 1_i^T and 0_i^T are row vectors of order i with all elements unity and zero, respectively. Then, $\mathrm{SST} = \sum_{i=1}^n (x_i - \bar{x})^2 = x^T K^T K x$.

Proof. $\mathrm{SST} = \sum_{i=1}^n (x_i - \bar{x})^2 = x^T C x$ can be obtained by Theorem 1. For H is orthogonal, i.e., $HH^T = H^T H = I_n$, whereas

$$H^T H = \begin{pmatrix} h & K^T \end{pmatrix}\begin{pmatrix} h^T \\ K \end{pmatrix} = hh^T + K^T K,$$
$$= n^{-(1/2)} 1_n \cdot n^{-(1/2)} 1_n^T + K^T K = \frac{1}{n} 1_n 1_n^T + K^T K, \tag{5}$$

so $(1/n) 1_n 1_n^T + K^T K = I_n$; thus, $K^T K = I_n - (1/n) 1_n 1_n^T = C_n$. Hence, the conclusion holds.

The centering matrix also has a close association with the Laplacian matrix of a connected graph. The related property is as follows. □

Proposition 2. *(see [21]). The Laplacian matrix L_n of a connected graph with n vertices and its generalized inverse L_n^+ satisfy the following equality: $L_n L_n^+ = L_n^+ L_n = C_n$, $n \geq 2$.*

Theorem 2. *The eigenvalues of C_n are 0 of multiplicity 1 and 1 of multiplicity $n-1$.*

Proof. Owing to $C_n = I_n - (1/n) J_n$, we can find the eigenvalue of J_n at first. Assume that its eigenvalue be λ and the corresponding eigenvector be x. Then, $J_n x = \lambda x$. Since $J_n^2 x = \lambda^2 x$ and $J_n{}^2 = nJ_n$, we can obtain $\lambda^2 x = n\lambda x$, whereas $x \neq 0$, so $\lambda^2 - n\lambda = 0$; therefore, $\lambda = 0$ or $\lambda = n$.

Let the multiplicity of eigenvalue 0 be k_1. Since J_n is real symmetric, it can be diagonalizable; thus, $R(0I - J_n) = n - k_1$. It is easy to know that $R(J_n) = 1$, so $k_1 = n - 1$. Because there must be n eigenvalues for square matrix of order n, we can obtain that the multiplicity of eigenvalue n is 1.

The eigenvalue of C_n is equivalent to $-(1/n)\lambda + 1$, whereas $\lambda = n$ or $\lambda = 0$, so the eigenvalues of C_n are 0 of multiplicity 1 and 1 of multiplicity $n - 1$. □

Corollary 2 (the spectral decomposition of the centering matrix). *For the centering matrix C_n, there must exist an orthogonal matrix of order n $Q = (y_1, y_2, \ldots, y_{n-1}, y_n)$ such that $C_n = QD_nQ^T = y_1y_1^T + y_2y_2^T + \cdots + y_{n-1}y_{n-1}^T + 0y_ny_n^T$, where $y_1, y_2, \ldots, y_{n-1}$ are the unit orthogonal eigenvectors corresponding to eigenvalue 1, y_n is the unit orthogonal eigenvector corresponding to eigenvalue 0, and $D_n = \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix}$.*

Proof. Since the centering matrix C_n is real symmetric, it can be diagonalizable. We can know that the eigenvalues of C_n are 0 of multiplicity 1 and 1 of multiplicity $n - 1$ by Theorem 2. Suppose that the unit orthogonal eigenvectors corresponding to eigenvalue 1 are $y_1, y_2, \ldots, y_{n-1}$ and the unit orthogonal eigenvector corresponding to eigenvalue 0 is y_n. Let $Q = (y_1, y_2, \ldots, y_{n-1}, y_n)$. Then, Q is orthogonal and satisfies $Q^TC_nQ = D_n$, where $D_n = \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix}$ is a diagonal matrix. The following formula can be obtained by using the equation $Q^TC_nQ = D_n$ for left multiplication by Q and right multiplication by Q^T:

$$\begin{aligned} C_n &= QD_nQ^T = (y_1, y_2, \ldots, y_n)\operatorname{diag}(1, 1, \ldots, 1, 0)\left(y_1^T, y_2^T, \ldots, y_{n-1}^T, y_n^T\right)^T, \\ &= y_1y_1^T + y_2y_2^T + \cdots + y_{n-1}y_{n-1}^T + 0y_ny_n^T. \end{aligned} \tag{6}$$

□

Remark 1. In the previous formula, each term at the right end is the outer product of each of the unit orthogonal eigenvector of C_n with itself multiplying the corresponding eigenvalue. Each term is a square matrix with rank 1. The above formula has simple and special structure, so it has important applications in statistics.

Corollary 3. *The centering matrix C_n can be decomposed as $C_n = LL^T$, where L satisfies the relation $L^TL = I_{n-1}$.*

Proof. The equation $C_n = QD_nQ^T$ can be obtained by the proof process of Corollary 2, where the definitions of Q and D_n are given in Corollary 2. It is easy to know that $D_n = D_nD_n^T$. Therefore, $C_n = QD_nQ^T = QD_nD_n^TQ^T = (QD_n)(QD_n)^T = LL^T$, where $L = QD_n$ is a matrix of order $n \times (n - 1)$ such that $L^TL = (QD_n)^T(QD_n)$:

$$\begin{aligned} &= \begin{pmatrix} y_1 & y_2 & \cdots & y_{n-1} \end{pmatrix}^T \begin{pmatrix} y_1 & y_2 & \cdots & y_{n-1} \end{pmatrix}, \\ &= \begin{pmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_{n-1}^T \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \cdots & y_{n-1} \end{pmatrix}, \\ &= \begin{pmatrix} y_1^Ty_1 & y_1^Ty_2 & \cdots & y_1^Ty_{n-1} \\ y_2^Ty_1 & y_2^Ty_2 & \cdots & y_2^Ty_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n-1}^Ty_1 & y_{n-1}^Ty_2 & \cdots & y_{n-1}^Ty_{n-1} \end{pmatrix} = I_{n-1}. \end{aligned} \tag{7}$$

We can obtain that C_n has 0 eigenvalue by Theorem 2; therefore, $|C_n| = 0$, and thus, C_n is irreversible (singular).

The equality $C_n = Q\begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix}Q^T$ can be obtained by Corollary 2. Hence, we have [30]

$$C_n^+ = Q\begin{pmatrix} I_{n-1}^{-1} & \\ & 0 \end{pmatrix}Q^T = Q\begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix}Q^T = C_n.$$

It indicates that the Moore–Penrose generalized inverse of C_n is itself, i.e., $C_n^+ = C_n$.

We can know that the eigenvalues of C_n is nonnegative by Theorem 2, so it is positive semidefinite, denoted by $C_n \geq 0$; hence, it can be used as the covariance matrix of some random variable.

Because $C_n \geq 0$, so for arbitrary nonzero column vector x, $x^TC_nx \geq 0$. Let $x = e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, where 1 is the ith element. Then, $0 \leq e_i^TC_ne_i = c_{ii}$, for $i = 1, 2, \ldots, n$, i.e., the diagonal elements of C_n are nonnegative.

We can know that C_n is similar to diagonal matrix D_n by Corollary 2. The similar matrices have the same rank, so $R(C_n) = R(D_n) = n - 1$, i.e., the rank of C_n is $n - 1$.

We can verify that the rank of C_n^* is 1 by the theory of linear algebra. The proof process is as follows. On the one hand, $|C_n| = 0$; therefore, $C_n \cdot C_n^* = |C_n|I = O$, and thus, $R(C_n) + R(\operatorname{adj}(C_n)) \leq n$. Because $R(C_n) = n - 1$, so $R(C_n^*) \leq 1$. On the other hand, since $R(C_n) = n - 1$, there must exist a nonzero minor of order n-1 for C_n. Then, we can obtain $C_n^* \neq O$, and thus, $R(C_n^*) \geq 1$. In summary, we have $R(C_n^*) = 1$.

We can know that the eigenvalues of C_n are 0 with multiplicity 1 and 1 with multiplicity $n - 1$ by Theorem 2. Since the trace of a square matrix equals the sum of all eigenvalues of the matrix, $\operatorname{tr}(C_n) = n - 1$, i.e., the trace of the centering matrix C_n is $n - 1$.

Because $\operatorname{tr}(C_n^*) = [\operatorname{tr}(C_n)]^*$ and $\operatorname{tr}(C_n) = n-1$, so $\operatorname{tr}(C_n^*) = n-1$, i.e., the trace of C_n^* is $n-1$.

Although C_n is singular, and $C_n + aJ_n$ is nonsingular for $a \neq 0$. The related conclusion is as follows: □

Proposition 3 (see [12], and we prove it). *The centering matrix C_n and the matrix J_n satisfy the relations $J_n(C_n + aJ_n)^{-1} = (C_n + aJ_n)^{-1}J_n = (1/an)J_n$ and $(C_n + aJ_n)^{-1} - (C_n + bJ_n)^{-1} = ((b-a)/(abn^2))J_n$ for $a \neq 0, b \neq 0$.*

Proof. For $p \neq 0$ and $p + qn \neq 0$,

$$\begin{aligned}(pI + qJ_n)\frac{1}{p}\left(I - \frac{q}{p+qn}J_n\right) &= (pI + qJ_n)\left(\frac{1}{p}I - \frac{q}{p(p+qn)}J_n\right),\\ &= (pI)\left(\frac{1}{p}I\right) - p\frac{q}{p(p+qn)}J_n + \frac{q}{p}J_n - \frac{q^2}{p(p+qn)}J_n^2,\\ &= I - \frac{q}{(p+qn)}J_n + \frac{q}{p}J_n - \frac{q^2}{p(p+qn)}nJ_n,\\ &= I + \left[\frac{q}{p} - \frac{q}{(p+qn)} - \frac{nq^2}{p(p+qn)}\right]J_n = I,\end{aligned} \tag{8}$$

so $(pI + qJ_n)^{-1} = (1/p)(I - (q/(p+qn))J_n)$. Using this equation, we have $(C_n + aJ_n)^{-1} = (I - (1/n)J_n + aJ_n)^{-1}$

$$\begin{aligned}&= \left[I + \left(a - \frac{1}{n}\right)J_n\right]^{-1} = I - \frac{a - (1/n)}{1 + (a - (1/n))n}J_n\\ &= I - \frac{na - 1}{n + (na-1)n}J_n = I - \frac{1}{n}J_n + \frac{1}{n^2a}J_n\\ &= C_n + \frac{1}{n^2a}J_n.\end{aligned} \tag{9}$$

Therefore,

$$\begin{aligned}J_n(C_n + aJ_n)^{-1} &= J_n(I - (1/n)J_n + (1/n^2a)J_n)\\ &= J_n - \frac{1}{n}nJ_n + \frac{1}{n^2a}nJ_n = \frac{1}{na}J_n.\end{aligned} \tag{10}$$

In the same way, we can obtain $(C_n + aJ_n)^{-1}J_n = (1/na)J_n$.

Using the equation $(C_n + aJ_n)^{-1} = C_n + (1/n^2a)J_n$, we get the result

$$(C_n + aJ_n)^{-1} - (C_n + bJ_n)^{-1} = \left(C_n + \frac{1}{n^2a}J_n\right) - \left(C_n + \frac{1}{n^2b}J_n\right) = \frac{b-a}{abn^2}J_n. \tag{11}$$

□

Lemma 2 (see [12], and we prove it). *The centering matrix C_n commutes with matrix J_n, and their product is zero matrix, i.e., $C_nJ_n = J_nC_n = O$.*

Proof. $C_nJ_n = (I_n - (1/n)J_n)J_n = J_n - (1/n)J_n^2 = J_n - (1/n)\ nJ_n = \mathrm{O}$ and $J_nC_n = J_n(I_n - (1/n)J_n) = J_n - (1/n)J_n^2 = J_n - (1/n)nJ_n = \mathrm{O}$. □

Corollary 4. *The null space of C_n is $\mathbf{N}(C_n) = \{\mathbf{k1}_n, k \in R\}$ whose dimension is 1 and a set of bases is $\mathbf{1}_n$.*

Proof. Since $\dim(N(C_n)) + \dim(R(C_n)) = \mathrm{n}$ and $\dim(R(C_n)) = \mathrm{n}\text{-}1, \dim(N(C_n)) = 1$, that is, the dimension of the null space is 1. The equation $C_n1_n = 0_n$ can be obtained by Lemma 2. So, the null space of C_n is $\operatorname{Span}(1_n)$, i.e., $\mathbf{N}(C_n) = \{\mathbf{k1}_n, k \in R\}$. Obviously, a set of bases of $\mathbf{N}(C_n)$ is 1_n. □

Corollary 5. *For the centering matrix C_n, the set of all eigenvectors corresponding to the eigenvalue 0 is $\{k1_n, k \in (R/\{0\})\}$.*

Proof. Since the vector 1_n is the column of the matrix J_n and uses Lemma 2, the equation $C_n1_n = 0_n = 01_n$ holds. We get the result that the sum vector 1_n is an eigenvector of C_n corresponding to the eigenvalue 0 according to the definition of eigenvector. Then, we have the desired result. □

Proposition 4 (see [14], and we prove it). *The vector resulted from the vector $y = (y_1, y_2, \ldots, y_n)^T$ by the transformation C_n and denoted by C_ny, which is the projection of y to $n-1$ dimension subspace $V_1 = \{(x_1, x_2, \ldots, x_n) | x_1 + x_2 + \cdots + x_n = 0\}$, where the subspace V_1 is orthogonal to the subspace $V_0 = \{k1_n, k \in R\}$.*

Proof. We can obtain that the eigenvalues of C_n are 0 with multiplicity 1 and 1 with multiplicity $n-1$ by Theorem 2. Suppose that the eigenspace of eigenvalue 1 is $n-1$ dimension subspace V_1 and the eigenspace of eigenvalue 0 is 1 dimension subspace V_0; then, V_1 is orthogonal to $\mathbf{V}_0$[22]. It is worth mentioning that the eigenspace V_0 is also the null space of the centering matrix C_n.The result $V_0 = \{k1_n, k \in R\}$ can be obtained by Corollary 5. For C_n, assuming that the eigenvector corresponding to eigenvalue 1 is $x = (x_1, x_2, \ldots, x_n)^T$, then we can obtain that $C_n x = 1 \cdot x$ by the definition of eigenvalue and eigenvector, i.e., $(I_n - (1/n)J_n)x = x$. Replacing the corresponding value, we have $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - (1/n)\begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$; thus, $x_1 + x_2 + \cdots + x_n = 0$, i.e., $V_1 = \{(x_1, x_2, \ldots, x_n) | x_1 + x_2 + \cdots + x_n = 0\}$. Since C_n is a projection matrix, $C_n y$ is the projection of the vector y to $n-1$ dimension subspace $V_1 = \{(x_1, x_2, \ldots, x_n) | x_1 + x_2 + \cdots + x_n = 0\}$.

In the following, we will discuss the Kronecker product property of C_n.

From the previous text, we can know that C_n is a positive semidefinite matrix. Suppose B_m is a positive semidefinite matrix of order m. Since the Kronecker product of two positive semidefinite matrices is also positive semidefinite, $B_m \otimes C_n$ is also a positive semidefinite matrix [31]. In particular, $C_n \otimes C_n$ is also a positive semidefinite matrix. Next, we will further discuss the other property of $C_n \otimes C_n$. □

Proposition 5 (Kronecker square property). *The Kronecker square of C_n is defined as $C_n^{[2]} = C_n \otimes C_n$. It has the following characteristics:*

(i) It is a projection matrix

(ii) It is irreversible

(iii) Its Moore–Penrose Generalized inverse is itself, i.e., $(C_n^{[2]})^+ = C_n^{[2]}$

(iv) Its rank and trace are all equal to $(n-1)^2$, i.e., $R(C_n^{[2]}) = \mathrm{tr}(C_n^{[2]}) = (n-1)^2$

Proof

(i) On the one hand, $(C_n^{[2]})^T = (C_n \otimes C_n)^T = C_n^T \otimes C_n^T = C_n \otimes C_n = C_n^{[2]}$.
It shows that $C_n^{[2]}$ is symmetric. On the one hand,

$$\begin{aligned} \left(C_n^{[2]}\right)^2 &= (C_n \otimes C_n)^2 = (C_n \otimes C_n)(C_n \otimes C_n) \\ &= (C_n C_n) \otimes (C_n C_n), \end{aligned} \tag{12}$$

$$\left(C_n^2\right) \otimes \left(C_n^2\right) = C_n \otimes C_n = C_n^{[2]}.$$

Hence, $C_n^{[2]}$ is an idempotent matrix. To sum up, $C_n^{[2]}$ is a projection matrix.

(ii) $\det(C_n^{[2]}) = \det(C_n \otimes C_n) = (\det(C_n))^n \cdot (\det(C_n))^n = 0$. Hence, $C_n^{[2]}$ is irreversible.

(iii) $(C_n^{[2]})^+ = (C_n \otimes C_n)^+ = C_n^+ \otimes C_n^+ = C_n \otimes C_n = C_n^{[2]}$. So, the Moore–Penrose generalized inverse of $C_n^{[2]}$ is equal to itself.

(iv) $R(C_n^{[2]}) = \mathrm{rank}(C_n^{[2]}) = \mathrm{rank}(C_n \otimes C_n) = \mathrm{rank}(C_n) \cdot \mathrm{rank}(C_n) = (n-1)^2$, and $\mathrm{tr}(C_n^{[2]}) = \mathrm{tr}(C_n \otimes C_n) = \mathrm{tr}(C_n) \cdot \mathrm{tr}(C_n) = (n-1)^2$.

Therefore, $R(C_n^{[2]}) = \mathrm{tr}(C_n^{[2]}) = (n-1)^2$.

The relationship between some previous properties such as the centering property, quadratic form property, spectral property, and their roles in ML, MVA, theoretical chemistry, and graph theory can be illustrated in Figure 1. □

3. The Application of the Centering Matrix in ML and Statistical Analysis

3.1. The Application in the PCA. It is often necessary to calculate the eigenvector of the sample covariance matrix in dimensionality reduction by using the PCA, one of the ML algorithms. In this case, it is more convenient to represent the covariance matrix by the centering matrix.

Example 1. Suppose the sample data matrix is

$$\begin{aligned} X &= \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{pmatrix}_{n \times m} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}, \\ &= \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^T. \end{aligned} \tag{13}$$

Then, the mean of the sample is

$$\bar{x}_{m \times 1} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad = \frac{1}{n}X^T \mathbf{1}_n. \tag{14}$$

The sample covariance matrix is as follows [26]:

$$\begin{aligned} S_X &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T, \\ &= \frac{1}{n-1}(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})^T, \\ &= \frac{1}{n-1}\left(X^T - \bar{x}\mathbf{1}_n^T\right)\left(X^T - \bar{x}\mathbf{1}_n^T\right)^T. \end{aligned} \tag{15}$$

Replacing $\overline{x}$ by $(1/n)X^T \mathbf{1}_n$ and using the symmetry and idempotency of C_n, we can obtain

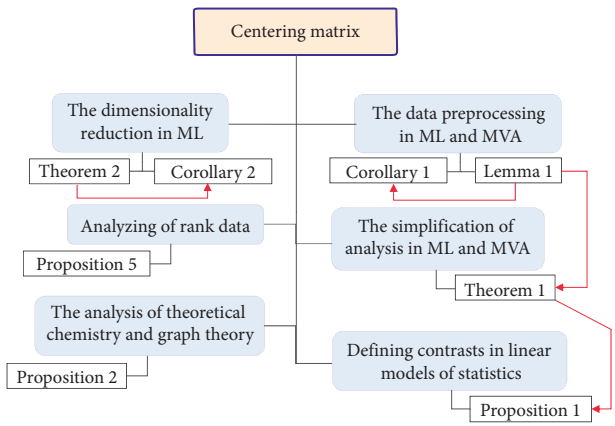

FIGURE 1: The relationship between some properties of the centering matrix and their roles in ML and MVA.

$$\begin{aligned} S_X &= \frac{1}{n-1}\left(X^T - \frac{1}{n}X^T\mathbf{1}_n\mathbf{1}_n^T\right)\left(X^T - \frac{1}{n}X^T\mathbf{1}_n\mathbf{1}_n^T\right)^T, \\ &= \frac{1}{n-1}X^T\left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)\left[X^T\left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)\right]^T, \\ &= \frac{1}{n-1}X^TC_n\left(X^TC_n\right)^T = \frac{1}{n-1}X^TC_nC_n^TX, \\ &= \frac{1}{n-1}X^TC_n^2X = \frac{1}{n-1}X^TC_nX, \end{aligned} \tag{16}$$

where $(n-1)S_X = X^TC_nX$ is the scatter matrix.

This example shows that the sample covariance matrix and scatter matrix can be expressed succinctly by the centering matrix. Meanwhile, we can know that the sample covariance matrix is a Gramian matrix and a positive semidefinite matrix. Similarly, the correlation matrix can also be expressed simply by the centering matrix.

3.2. The Application in the Regression Analysis

Example 2. The centering matrix has important simplified expression role in regression problem of supervised learning. Suppose there is a linear regression model as shown below between the random vector y and $p-1$ nonrandom factors $x_1, x_2, \ldots, x_{p-1}$. $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_{p-1}x_{p-1} + \varepsilon$, where ε is the random error, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ are unknown parameters. In order to determine their value, substitute the data $(x_{i1}, x_{i2}, \ldots, x_{i,p-1}; y_i), i = 1, 2, \ldots, n$, by $n (n \geq p)$ independent observations into the above formula; then, we can obtain the following equations:

$$y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \cdots + \beta_{p-1}x_{i,p-1} + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{17}$$

The corresponding matrix form is $y = X\beta + \varepsilon$, where

$$\begin{aligned} y &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \\ X &= \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \ldots & x_{2,p-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{n,p-1} \end{pmatrix}, \\ \beta &= \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \\ \varepsilon &= \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \end{aligned} \tag{18}$$

It can be found that the least square estimation of the parameter β is $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta_{p-1}})^T = (X^TX)^{-1}X^Ty$.

The fitted value of y can be obtained by substituting the above formula into $y = X\beta + \varepsilon$ and omitting error term ε:

$$\widehat{y} = (\widehat{y}_1, \widehat{y}_2, \ldots, \widehat{y}_n)^T = X\widehat{\beta} = X\left(X^TX\right)^{-1}X^Ty = My, \tag{19}$$

where $M = X(X^TX)^{-1}X^T$ is a projection matrix of order n. The expression of residual vector is shown as follows:

$$\begin{aligned} \widehat{\varepsilon} = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n)^T &= y - \widehat{y} = y - X\widehat{\beta}, \\ &= y - X\left(X^TX\right)^{-1}X^Ty, \\ &= \left[I - X\left(X^TX\right)^{-1}X^T\right]y = (I - M)y, \end{aligned} \tag{20}$$

where I is the unity matrix of order n, and if X^TX is singular, then find its Moore–Penrose generalized inverse.

Next, we are going to express the regression sum of squares (SSR), the error sum of squares (SSE), and SST by the centering matrix and do further derivation.

(i) The following formula about SSR can be obtained by Theorem 1 and the equation $\widehat{y} = My$:

$$\begin{aligned} \mathrm{SSR} &= \sum_{i=1}^{n}\left(\widehat{y}_i - \frac{1}{n}\sum_{i=1}^{n}\widehat{y}_i\right)^2 \\ &= \widehat{y}^TC_n\widehat{y} = \widehat{y}^T\left(I - \frac{1}{n}J\right)\widehat{y} \\ &= y^TM^T\left(I - \frac{1}{n}J\right)\mathrm{My} = y^TM\left(I - \frac{1}{n}1_n1_n^T\right)\mathrm{My}. \end{aligned} \tag{21}$$

Table 1: Quadratic form expressions for SSR, SSE, and SST.

Type	Quadratic form expression with the kernel matrix C_n	Quadratic form expression of y
SSR	$\widehat{y}^T C_n \widehat{y}$	$y^T (M - (1/n)J) y$
SSE	$\widehat{\varepsilon}^T C_n \widehat{\varepsilon}$	$y^T (I - M) y$
SST	$y^T C_n y$	$y^T C_n y$

On the one hand, the equality $\sum_{i=1}^n \widehat{\varepsilon}_i = 0$ holds [32]. On the other hand, $\sum_{i=1}^n \widehat{\varepsilon}_i = 1_n^T \widehat{\varepsilon} = 1_n^T (I - M) y$. Combining the above two equations, we have

$$1_n^T y = 1_n^T \mathrm{My}. \tag{22}$$

Therefore,

$$\begin{aligned}
\mathrm{SSR} &= y^T M y - \frac{1}{n} y^T M 1_n 1_n^T M y \\
&= y^T M y - \frac{1}{n} y^T M^T 1_n 1_n^T M y \\
&= y^T M y - \frac{1}{n} \left(1_n^T M y\right)^T 1_n^T M y \\
&= y^T M y - \frac{1}{n} \left(1_n^T y\right)^T 1_n^T y \\
&= y^T M y - y^T \frac{1}{n} 1_n 1_n^T y \\
&= y^T \left(M - \frac{1}{n} J\right) y.
\end{aligned} \tag{23}$$

(ii) SSE is also called residual sum of squares and defined as $\mathrm{SSE} = \sum_{i=1}^n (y_i - \widehat{y}_i)^2 = \sum_{i=1}^n \widehat{\varepsilon}_i^2$. Due to $\sum_{i=1}^n \widehat{\varepsilon}_i = 0$, the above formula is changed into

$$\mathrm{SSE} = \sum_{i=1}^n \left(\widehat{\varepsilon}_i - \sum_{i=1}^n \widehat{\varepsilon}_i\right)^2 = \sum_{i=1}^n \left(\widehat{\varepsilon}_i - \frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i\right)^2. \tag{24}$$

We can obtain the following expression of $\widehat{\varepsilon}$ by Theorem 1; hence,

$$\begin{aligned}
\mathrm{SSE} &= \widehat{\varepsilon}^T C_n \widehat{\varepsilon} = \widehat{\varepsilon}^T \left(I - \frac{1}{n} J\right) \widehat{\varepsilon} \\
&= [(I - M) y]^T \left(I - \frac{1}{n} J\right) (I - M) y \\
&= y^T (I - M)^T \left(I - \frac{1}{n} J\right) (I - M) y \\
&= y^T (I - M)^T (I - M) y - \frac{1}{n} y^T (I - M)^T J (I - M) y \\
&= y^T (I - M)^T (I - M) y - \frac{1}{n} y^T (I - M)^T 1_n 1_n^T (I - M) y.
\end{aligned} \tag{25}$$

We can obtain $1_n^T (I - M) y = 0$ by (i). Therefore, the preceding formula reduces to $\mathrm{SSE} = y^T (I - M) y$.

(iii) According to Theorem 1, SST can be expressed as

$$\mathrm{SST} = \sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i\right)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = y^T C_n y. \tag{26}$$

In the above example, for the sum of squares such as SSR and SSE, we expressed it by the centering matrix firstly and then obtained the quadratic form expression of y after a series of derivation. For the sum of squares SST, we directly obtained its quadratic form expression by Theorem 1. The Table 1 gives the quadratic form expression in which the kernel function is C_n and the quadratic form expression of y for SSR, SSE, and SST, respectively.

From Table 1, we can observe that SSR, SSE, and SST are all expressed as the quadratic form with the kernel matrix C_n. Generally speaking, sum of squares all can be considered to be expressed by the centering matrix. For instance, the sum of squares of normally distribution known as χ^2 distribution is such a case. From formulas (23), (25), and (26), we can easily deduce the famous decomposition equality SST = SSR + SSE which is of vital importance in regression theory and analysis of variance.

In addition, the centering matrix has important simplified role in the ranking algorithms in information retrieval [4, 19, 20]. For instance, in [19], T. Pahikkala et al. introduced the centering matrix to construct a block diagonal matrix as weighted matrix, express the loss function, obtain the matrix notation of a least squares problem, and conveniently find the linear solution of RankRLS algorithm.

4. Conclusion

As we know, the centering matrix plays an important role in ML and MVA. In this paper, we mainly carried out the theory research of the centering matrix. We discussed in detail some properties of the centering matrix, proved some previously known properties, deduced some new properties of the centering matrix, and finally drew a mind map to clearly illustrate the relationship between these properties and their potential role in ML and MVA. Then, we explore the application of the centering matrix such as PCA and regression analysis. The research results show that the centering matrix has excellent characteristics and simplified expression role by matrix notation in ML and MVA. The sample covariance matrix, scatter matrix, and sum of deviation squares all can be expressed succinctly by the centering matrix. Since the centering matrix has good properties such as symmetry and idempotency, for the algorithms based on it, the learning process can be greatly simplified and the performance such as training complexity, stability, and speed can be improved.

Data Availability

This paper is the theoretical research on the centering matrix. There are no corresponding programs and software.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61801437), Research Topics of Social and Economic Statistics in Shanxi Province (KY2020128), and Shanxi Natural Science Foundation (201801D221040).

References

[1] M. S. Long, J. M. Wang, G. G. Ding, J. G. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the ICCV*, pp. 2200–2207, Sydney, Australia, December 2013.

[2] T. Yang, The centering matrix [Online] https://orzyt.cn/posts/centering-matrix/, 2020.

[3] J. An, P. Ai, and D. K. Liu, "Deep domain adaptation model for bearing fault diagnosis with domain alignment and discriminative feature learning," *Shock and Vibration*, vol. 202014 pages, 2020.

[4] S. Z. Chen, K. Chen, C. F. Xu, and L. Lan, "Flexible ranking extreme learning machine based on matrix-centering transformation," in *Proceedings of the IJCNN*, Rio de Janeiro, Brazil, July 2018.

[5] Q. Wang and L. Zhang, "Online updating the generalized inverse of centered matrices," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, San Francisco, California USA, August 2011.

[6] Q. Wang, *Research on Several Key Problems of Ensemble Learning Algorithms*, Fudan University, Shanghai, China, 2011.

[7] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.

[8] A. Onan, S. Korukoğlu, and H. Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Systems with Applications*, vol. 62, pp. 1–16, 2016.

[9] A. Onan, "Classifier and feature set ensembles for web page classification," *Journal of Information Science*, vol. 42, no. 2, pp. 150–165, 2016.

[10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[11] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends in Computer Graphics and Vision*, vol. 8, pp. 46–56, 2014.

[12] S. R. Searle, *Matrix Algebra Useful for Statistics*, pp. 152-153, John Wiley & Sons, New York, NY, USA, 1982.

[13] X. D. Zhang, *Matrix Analysis and Application*, pp. 153–156, Higher education press, Beijing, China, 2004.

[14] Wikipedia, "Centering matrix [online]," 2020, https://encyclopedi a.thefreedictionary.com/Centering+matrix.

[15] K. T. Fang and M. Chen, *Matrix Algebra in Statistics*, pp. 177-178, Higher Education Press, Beijing, China, 2013.

[16] J. E. Gentle, *Matrix Algebra: Theory, Computations, and Application in Statistics*, pp. 293–296, Springer Science+-Business Media, LLC, New York, NY, USA, 2007.

[17] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: db-kit," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2849–2858, 2019.

[18] K. E. S. Pilario and Y. Cao, "Canonical variate dissimilarity analysis for process incipient fault detection," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5308–5315, 2018.

[19] T. Pahikkala, A. Airola, P. Naula, and T. Salakoski, "Greedy rankrls: a linear time algorithm for learning sparse ranking models," in *Proceedings of the Sigir 2010 Workshop on Feature Generation and Selection for Information Retrieval*, pp. 11–18, Geneva, Switzerland, July 2010.

[20] T. Pahikkala, W. Waegeman, A. Airola, T. Salakoski, and B. D. Baets, "Conditional ranking on relational data," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 1–16, Athens, Greece, September 2010.

[21] I. Gutman and W. Xiao, "Generalized inverse of the Laplacian matrix and some applications," *Bulletin: Classe des sciences mathematiques et natturalles*, vol. 129, no. 29, pp. 15–23, 2004.

[22] J. I. Marden, *Analyzing and Modeling Rank Data*, Vol. 59, Taylor & Francis Group, LLC, Boca Raton, FL, USA, 1995.

[23] L. P. Liu, Y. Jiang, and Z. H. Zhou, "Least square incremental linear discriminant analysis," in *Proceedings of ICDM*, pp. 298–306, Miami, Florida, December 2009.

[24] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[25] L. Sun, S. W. Ji, and J. P. Ye, "A least squares formulation for a class of generalized eigenvalue problems in machine learning," in *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pp. 1207–1216, Montreal, Canada, 2009.

[26] CSDN, "Dimension reduction," 2020, https://blog.csdn.net/cengjing12/article/details/106268447?utm_medium=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-5.nonecase&depth_1-utm_source=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-5.nonecase.

[27] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.

[28] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[29] J. Li, X. L. Li, Y. T. Su, and Z. Z. Ding, *Problem Solutions for Matrix Analysis and Applications*, pp. 102–104, Tsinghua University Press, Beijing, China, 2007.

[30] S. G. Wang, *The Theory and its Application of Linear Models*, pp. 17-18, The Anhui Education Press, Hefei, China, 1987.

[31] R. F. Yan and X. M. Fan, "The probability proof of several theorems in matrix theory," *Journal of Northwest Normal University (Natural Science)*, vol. 63, no. 2, pp. 71-72, 1987.

[32] C. L. Mei and J. C. Fan, *Data Analysis Method*, pp. 38–46, Higher Education Press, Beijing, China, 2006.