Multivariate Statistics

Lecture 12

Fudan University

Outline

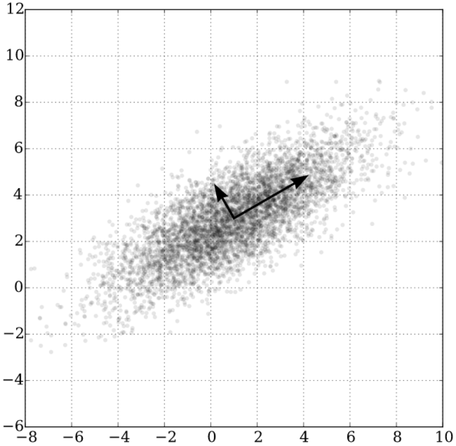
Outline

1. Principal Components

2. Canonical Correlation

Principal Components

In statistical practice, the method of principal components is used to find the linear combinations with large variance.



Principal Components

Let random vector $\mathbf{x}$ of p component has mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

Let $\boldsymbol{\beta}$ be a p-component column vector such that $\|\boldsymbol{\beta}\|_2 = 1$.

1. The variance of $\boldsymbol{\beta}^\top \mathbf{x}$ is
$$\mathbb{E}\left[(\boldsymbol{\beta}^\top \mathbf{x})^2\right] = \boldsymbol{\beta}^\top \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right]\boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}.$$

2. Maximizing $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}$ must satisfy
$$(\boldsymbol{\Sigma} - \lambda_1 \mathbf{I})\boldsymbol{\beta} = \mathbf{0},$$
where λ_1 is the largest root of
$$\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) = 0.$$

3. Let $\boldsymbol{\beta}^{(1)} = \arg\max_{\|\boldsymbol{\beta}\|_2 = 1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}$.

Principal Components

Find $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}^\top \mathbf{x}$ has maximum variance and is uncorrelated with $u_1 = {\boldsymbol{\beta}^{(1)}}^\top \mathbf{x}$.

1. Lack of correlation means

$$\boldsymbol{\beta}^\top \boldsymbol{\beta}^{(1)} = 0.$$

2. The vector $\boldsymbol{\beta}$ must satisfy

$$(\boldsymbol{\Sigma} - \lambda_2 \mathbf{I})\boldsymbol{\beta} = \mathbf{0},$$

where λ_2 is the second largest root of

$$\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) = 0.$$

Principal Components

At the $(r+1)$-th step, we want to find a vector such that $\boldsymbol{\beta}^\top \mathbf{x}$ has maximum variance and lacks correlation with $u_1 \dots, u_r$, that is

$$0 = \mathbb{E}\left[\boldsymbol{\beta}^\top \mathbf{x} u_i\right] = \mathbb{E}\left[\boldsymbol{\beta}^\top \mathbf{x}\mathbf{x}^\top \boldsymbol{\beta}^{(i)}\right] = \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}^{(i)} = \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}^{(i)}$$

for $i = 1, \dots, r$, where $u_i = \boldsymbol{\beta}^{(i)\top} \mathbf{x}$

Finally, we obtain $\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(p)}$ and $\lambda_1 \geq \dots \geq \lambda_p$ such that

$$\boldsymbol{\Sigma}\mathbf{B} = \mathbf{B}\boldsymbol{\Lambda}$$

where $\mathbf{B} = [\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(p)}]$ satisfying $\mathbf{B}^\top \mathbf{B} = \mathbf{I}$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}.$$

Principal Components

The transformation

$$\mathbf{u} = \mathbf{B}^\top \mathbf{x}$$

leads to the r-th component of $\mathbf{u}$ has maximum variance of all normalized linear combinations uncorrelated with $u_1, \ldots, u_{r-1}$.

The vector $\mathbf{u}$ is defined as the vector of principal components of $\mathbf{x}$.

Principal Components

Theorem 1

An orthogonal transformation $\mathbf{v} = \mathbf{C}\mathbf{x}$ of a random vector $\mathbf{x}$ with $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ leaves invariant the generalized variance and the sum of the variances of the components.

Corollary 1

The generalized variance of the vector of principal components is the generalized variance of the original vector, and the sum of the variances of the principal components is the sum of the variances of the original variates.

Principal Components

Another approach is based on the density of the normal distribution.

1. Let $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, whose density is

$$\frac{1}{(2\pi)^{\frac{p}{2}}(\det(\mathbf{\Sigma}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}\right).$$

2. Surfaces of constant density are ellipsoids

$$\left\{\mathbf{x} : \mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x} = C\right\}$$

3. A principal axis of this ellipsoid is defined as the line from $-\mathbf{y}$ to $\mathbf{y}$, where $\mathbf{y}$ is a point on the ellipsoid where its squared distance has a stationary point.

Maximum Likelihood Estimators of Principal Components

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be N observations from $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ has p different characteristic roots and $N > p$. Then a set of maximum likelihood estimators of $\lambda_1, \ldots, \lambda_p$ and $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(p)}$ consists of the roots $\lambda_1 > \cdots > \lambda_p$ of

$$\det(\hat{\boldsymbol{\Sigma}} - \lambda \mathbf{I}) = 0$$

and a set of vectors $\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(p)}$ satisfying $\|\hat{\boldsymbol{\beta}}^{(i)}\|_2 = 1$ and

$$(\hat{\boldsymbol{\Sigma}} - \lambda_i \mathbf{I})\hat{\boldsymbol{\beta}}^{(i)} = \mathbf{0}$$

for $i = 1, \ldots, p$, where $\hat{\boldsymbol{\Sigma}}$ is the the maximum likelihood estimate of $\boldsymbol{\Sigma}$.

Asymptotic Distributions

Let the characteristic roots of $\mathbf{\Sigma}$ be $\lambda_1 > \cdots > \lambda_p$ and the corresponding characteristic vectors be $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(p)}$ with $\|\boldsymbol{\beta}^{(i)}\|_2 = 1$ and $\boldsymbol{\beta}_{1i} \geq 0$.

Let the characteristic roots of $\mathbf{S}$ be $l_1 > \cdots > l_p$ and the corresponding characteristic vectors be $\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(p)}$ with $\|\mathbf{b}^{(i)}\|_2 = 1$ and $b_{1i} \geq 0$.

Let $d_i = \sqrt{n}\,(l_i - \lambda_i)$ and $\mathbf{g}^{(i)} = \mathbf{b}^{(i)} - \boldsymbol{\beta}^{(i)}$ for $i = 1, \ldots, p$.

1. The limiting normal distribution the sets $\{d_1, \ldots, d_p\}$ and $\{\mathbf{g}^{(1)}, \ldots, \mathbf{g}^{(p)}\}$ are independent and $d_1, \ldots, d_p$ are mutually independent.
2. The element d_i has the limiting distribution $\mathcal{N}(0, 2\lambda_i^2)$.
3. The random vectors $\mathbf{g}^{(1)}, \ldots, \mathbf{g}^{(p)}$ has limiting normal distribution with

$$\lim_{n\to\infty} \mathrm{Cov}[\mathbf{g}^{(i)}, \mathbf{g}^{(j)}] = \begin{cases} \displaystyle\sum_{k=1, k\neq i}^{p} \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \boldsymbol{\beta}^{(k)} {\boldsymbol{\beta}^{(k)}}^\top, & i = j \\ \displaystyle -\frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \boldsymbol{\beta}^{(j)} {\boldsymbol{\beta}^{(i)}}^\top, & i \neq j \end{cases}$$

Asymptotic Distributions

One treats l_i as approximately normal with mean λ_i and variance $2\lambda_i^2/n$. Since l_i is a consistent estimate of λ_i, the limiting distribution of

$$\frac{\sqrt{n}\,(l_i - \lambda_i)}{\sqrt{2}\,l_i}$$

is $\mathcal{N}(0,1)$.

1. A two-tailed test of hypothesis $\lambda_i = \lambda_i^0$ has (asymptotic) acceptance region

$$-z(\epsilon) \le \frac{\sqrt{n}\,(l_i - \lambda_i^0)}{\sqrt{2}\,\lambda_i^0} \le z(\epsilon)$$

   where the value of the $\mathcal{N}(0,1)$ distribution beyond $z(\epsilon)$ is $\epsilon/2$.

2. The confidence interval for λ_i with confidence $1-\epsilon$ is

$$\frac{l_i}{1+\sqrt{2/n}\,z(\epsilon)} \le \lambda_i \le \frac{l_i}{1-\sqrt{2/n}\,z(\epsilon)}.$$

Exact Confidence Limits on the Characteristic Roots

Let $n\mathbf{S} \sim \mathcal{W}(\boldsymbol{\Sigma}, n)$, then

$$\frac{n\boldsymbol{\beta}^{(1)\top}\mathbf{S}\boldsymbol{\beta}^{(1)}}{\lambda_1} \quad \text{and} \quad \frac{n\boldsymbol{\beta}^{(p)\top}\mathbf{S}\boldsymbol{\beta}^{(p)}}{\lambda_p}.$$

are independently distrusted as χ^2-distribution with n degrees of freedom.

Let l and u be two numbers such that

$$1 - \epsilon = \Pr\left\{nl \leq \chi_n^2\right\} \Pr\left\{\chi_n^2 \leq nu\right\}.$$

Then a confidence interval for the characteristic roots of $\boldsymbol{\Sigma}$ with confidence at least $1 - \epsilon$ is

$$\frac{l_p}{u} \leq \lambda_p \leq \lambda_1 \leq \frac{l_1}{l}.$$

Outline

Canonical Correlations

We still consider random vector $\mathbf{x}$ of p components has zero means and the covariance matrix $\mathbf{\Sigma} \succ \mathbf{0}$.

We partition $\mathbf{x}$ into two subvectors of p_1 and p_2 components $(p_1 \leq p_2)$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}.$$

The covariance matrix is partitioned into p_1 and p_2 rows and columns

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}.$$

Here we shall develop a transformation of $\mathbf{x}^{(1)}$ and another transformation of $\mathbf{x}^{(2)}$ to a new system that exhibit clearly the intercorrelations between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$.

Canonical Correlations

Consider linear combinations

$$u = \boldsymbol{\alpha}^\top \mathbf{x}^{(1)} \quad \text{and} \quad v = \boldsymbol{\gamma}^\top \mathbf{x}^{(2)}.$$

We ask for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that maximize the correlation between u and v.

1. We require $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ such that

$$1 = \mathbb{E}\left[u^2\right] = \mathbb{E}\left[\boldsymbol{\alpha}^\top \mathbf{x}^{(1)} \mathbf{x}^{(1)\top} \boldsymbol{\alpha}\right] = \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha},$$
$$1 = \mathbb{E}\left[v^2\right] = \mathbb{E}\left[\boldsymbol{\gamma}^\top \mathbf{x}^{(2)} \mathbf{x}^{(2)\top} \boldsymbol{\gamma}\right] = \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\gamma}.$$

2. The correlation between u and v is

$$\mathbb{E}[uv] = \mathbb{E}\left[\boldsymbol{\alpha}^\top \mathbf{x}^{(1)} \mathbf{x}^{(2)\top} \boldsymbol{\gamma}\right] = \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{12} \boldsymbol{\gamma}.$$

3. Then the problem is

$$\max_{\substack{\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha} = 1 \\ \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\gamma} = 1}} \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{12} \boldsymbol{\gamma}.$$

Canonical Correlations

The solution of

$$\max_{\substack{\boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}=1 \\ \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}=1}} \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}.$$

must satisfy

$$\begin{bmatrix} -\lambda\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda\boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} = \mathbf{0},$$

where λ is the root of

$$\det\left(\begin{bmatrix} -\lambda\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda\boldsymbol{\Sigma}_{22} \end{bmatrix}\right) = 0.$$

Denote the largest root and the corresponds vectors be λ_1, $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\gamma}^{(1)}$

Canonical Correlations

Then we consider $u = \boldsymbol{\alpha}^\top \mathbf{x}^{(1)}$ and $v = \boldsymbol{\gamma}^\top \mathbf{x}^{(2)}$ for $\mathbf{x}^{(2)}$ with maximum correlation, such that u is uncorrelated with $u_1 = {\boldsymbol{\alpha}^{(1)}}^\top \mathbf{x}^{(1)}$ and v is uncorrelated with $v_1 = {\boldsymbol{\gamma}^{(1)}}^\top \mathbf{x}^{(2)}$.

This procedure is continued. At r-th step, we have

$$u_1 = {\boldsymbol{\alpha}^{(1)}}^\top \mathbf{x}^{(1)}, \dots, u_r = {\boldsymbol{\alpha}^{(r)}}^\top \mathbf{x}^{(1)}$$
$$v_1 = {\boldsymbol{\gamma}^{(1)}}^\top \mathbf{x}^{(2)}, \dots, v_r = {\boldsymbol{\gamma}^{(r)}}^\top \mathbf{x}^{(2)}$$

and each of them are uncorrelated. Let the correlation between u_i and v_i be λ_i.

We obtain $\boldsymbol{\alpha}^{r+1}$ and $\boldsymbol{\gamma}^{(r+1)}$ by maximizing the correlation between $u = \boldsymbol{\alpha}^\top \mathbf{x}^{(1)}$ and $v = \boldsymbol{\gamma}^\top \mathbf{x}^{(2)}$ such that u is uncorrelated with $u_1, \dots, u_r$ and v is uncorrelated with $v_1, \dots, v_r$.

Canonical Correlations

Let $\mathbf{A} = [\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(p_1)}]$, $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2] = [\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(p_2)}]$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{p_1} \end{bmatrix}.$$

All of conditions can be summarized as

$$\begin{aligned} \mathbf{A}^\top \boldsymbol{\Sigma}_{11} \mathbf{A} =& \mathbf{I}, \\ \mathbf{A}^\top \boldsymbol{\Sigma}_{12} \boldsymbol{\Gamma}_1 =& \boldsymbol{\Lambda}, \\ \boldsymbol{\Gamma}_1^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\Gamma}_1 =& \mathbf{I}, \\ \boldsymbol{\Gamma}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\Gamma}_1 =& \mathbf{0}, \\ \boldsymbol{\Gamma}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\Gamma}_2 =& \mathbf{I}. \end{aligned}$$

Canonical Correlations

Each $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\gamma}^{(i)}$ can be obtained by solving

$$\begin{bmatrix} -\lambda_i \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda_i \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} = \mathbf{0},$$

where λ_i is the i-th largest root of

$$\det\left(\begin{bmatrix} -\lambda \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) = 0.$$

This can be written as generalized eigenvalue problems

$$\left(\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} - \lambda^2 \boldsymbol{\Sigma}_{11} \right) \boldsymbol{\gamma} = \mathbf{0}$$

and

$$\left(\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} - \lambda^2 \boldsymbol{\Sigma}_{22} \right) \boldsymbol{\alpha} = \mathbf{0}.$$

Canonical Correlations

Let

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}$$

be a random vector where $\mathbf{x}^{(1)}$ has p_1 components and $\mathbf{x}^{(2)}$ has p_2 components.

In the r-th pair of canonical variates is the pair of linear combinations

$$u_r = \boldsymbol{\alpha}^{(r)\top}\mathbf{x}^{(1)} \quad \text{and} \quad v_r = \boldsymbol{\gamma}^{(r)\top}\mathbf{x}^{(2)},$$

each of unit variance and uncorrelated with the first $r-1$ pairs of canonical variates and having maximum correlation.

The correlation between u_r and v_r is the r-th canonical correlation.

Canonical Correlations

The canonical correlations are invariant with respect to transformations

$$\begin{cases} \mathbf{x}^{*(1)} = \mathbf{C}_1 \mathbf{x}^{(1)}, \\ \mathbf{x}^{*(2)} = \mathbf{C}_2 \mathbf{x}^{(2)}, \end{cases}$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are non-singular. Additionally, any function of $\boldsymbol{\Sigma}$ that is invariant (under any such transformation) is a function of the canonical correlations.

Estimation of Canonical Correlation

Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ be N observations from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let each $\mathbf{x}^{(\alpha)}$ be partitioned into suh-vectors of p_1 and p_2 components

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}.$$

The maximum likelihood of $\boldsymbol{\Sigma}$ is

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} \end{bmatrix} \\ &= \frac{1}{N} \begin{bmatrix} \sum_{\alpha=1}^{N} (\mathbf{x}_{\alpha}^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_{\alpha}^{(1)} - \bar{\mathbf{x}}^{(1)})^{\top} & \sum_{\alpha=1}^{N} (\mathbf{x}_{\alpha}^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_{\alpha}^{(2)} - \bar{\mathbf{x}}^{(2)})^{\top} \\ \sum_{\alpha=1}^{N} (\mathbf{x}_{\alpha}^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_{\alpha}^{(1)} - \bar{\mathbf{x}}^{(1)})^{\top} & \sum_{\alpha=1}^{N} (\mathbf{x}_{\alpha}^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_{\alpha}^{(2)} - \bar{\mathbf{x}}^{(2)})^{\top} \end{bmatrix}. \end{aligned}$$

The maximum likelihood estimators $\hat{\boldsymbol{\Lambda}}$, $\hat{\mathbf{A}}$ and $\hat{\boldsymbol{\Gamma}}$ of $\boldsymbol{\Lambda}$, $\mathbf{A}$ and $\boldsymbol{\Gamma}$ involve applying the algebra of the previous slides to $\hat{\boldsymbol{\Sigma}}$.

Sample Canonical Variates and Correlations

We could define the sample canonical variates and correlations in terms of

$$\mathbf{S} = \frac{1}{N-1}\sum_{\alpha=1}^{N}(\mathbf{x}_{\boldsymbol{\alpha}} - \bar{\mathbf{x}})(\mathbf{x}_{\boldsymbol{\alpha}} - \bar{\mathbf{x}})^\top = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}.$$

Let $\mathbf{a}^{(j)} = \sqrt{(N-1)/N}\,\hat{\boldsymbol{\alpha}}^{(j)}$, $\mathbf{c}^{(j)} = \sqrt{(N-1)/N}\,\hat{\gamma}^{(j)}$ and l_j satisfies

$$\begin{aligned} \mathbf{S}_{12}\mathbf{c}^{(j)} &= l_j\mathbf{S}_{11}\mathbf{a}^{(j)}, \\ \mathbf{S}_{21}\mathbf{c}^{(j)} &= l_j\mathbf{S}_{22}\mathbf{c}^{(j)}, \\ {\mathbf{a}^{(j)}}^\top\mathbf{S}_{11}\mathbf{a}^{(j)} &= 1, \\ {\mathbf{c}^{(j)}}^\top\mathbf{S}_{22}\mathbf{c}^{(j)} &= 1. \end{aligned}$$

We call linear combinations ${\mathbf{a}^{(j)}}^\top\mathbf{x}_\alpha^{(1)}$ and ${\mathbf{c}^{(j)}}^\top\mathbf{x}_\alpha^{(2)}$ the sample canonical variates.