# Principal component analysis

Ching-Lin Huang (Andy)

Institute of Ecology and Evolutionary Biology, NTU

March 25, 2022

# Content

1. Matrix algebra
2. Projection
3. Mean and variance
4. Principal component analysis

# Notation

$\mathbb{R}$ is the set of real number.
$x$ is a constant.
$\mathbf{x}$ is a "column" vector, and $\mathbf{x}^T$ is a "row" vector.
$\mathbf{X}$ is a matrix.

Definition

1. Let $\mathbf{x} = (x_1, \ldots, x_n)^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be a column vector in $\mathbb{R}^n$ if $x_1, \ldots, x_n \in \mathbb{R}$. We denote $\mathbf{x} \in \mathbb{R}^n$.
   $\mathbf{x}^T = (x_1, \ldots, x_n)$ is a row vector.
2. Let $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \ldots & \ldots & x_{np} \end{pmatrix}$ be a matrix where $x_{ij} \in \mathbb{R}$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

# Notation

Matrix can be expressed in three ways:

$$\mathbf{X} = (x_{ij})_{n\times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \dots & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x_1} & \dots & \mathbf{x_p} \end{pmatrix} = \begin{pmatrix} \mathbf{x_1}^T \\ \vdots \\ \mathbf{x_n}^T \end{pmatrix}$$

where $\mathbf{x_i}$ in the second form are the $i$-th column vector in $\mathbf{X}$ and $\mathbf{x_j}^T$ in the third form are the $j$-th row vector in $\mathbf{X}$.
e.g.
$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. The first column vector is $\mathbf{x_1} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and the second row vector is $\mathbf{x_2}^T = (3, 4)$.

# Calculation of vectors and matrices

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$ be vectors, $\mathbf{A} = (a_{ij})_{n\times p}$ and $\mathbf{B} = (b_{ij})_{n\times p}$ be two matrices and $c \in \mathbb{R}$ be a constant. Then we define

1. $c\mathbf{x} + \mathbf{y} = (cx_1 + y_1, cx_2 + y_2, \ldots, cx_n + y_n)^T$
2. $c\mathbf{A} + \mathbf{B} = (ca_{ij} + b_{ij})_{n\times p}$

e.g.

Let $\mathbf{x} = (1, 2, 3)^T$, $\mathbf{y} = (-1, 0, 1)^T$, $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $c = 2$.

# Calculation of vectors and matrices

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$ be vectors, $\mathbf{A} = (a_{ij})_{n \times p}$ and $\mathbf{B} = (b_{ij})_{n \times p}$ be two matrices and $c \in \mathbb{R}$ be a constant. Then we define

3. $\mathbf{x}^T\mathbf{y} = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$

   is said to be the inner product of vector $\mathbf{x}$ and $\mathbf{y}$.

4. $\mathbf{x}\mathbf{y}^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} y_1 & \ldots & y_n \end{pmatrix} = (x_iy_j)_{n \times n}$

   is said to be the tensor product of vector $\mathbf{x}$ and $\mathbf{y}$.

e.g.
Let $\mathbf{x} = (1, 2, 3)^T$, $\mathbf{y} = (-1, 0, 1)^T$.

# Calculation of vectors and matrices

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_p)^T$ be vectors and $\mathbf{A} = (a_{ij})_{n \times p}$ . Then we define

5.

$$\mathbf{x}^T\mathbf{A} = (x_1, \ldots, x_n)\begin{pmatrix} a_{11} & \ldots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{np} \end{pmatrix}$$

$$= \mathbf{x}^T \begin{pmatrix} \mathbf{a_1} & \ldots & \mathbf{a_p} \end{pmatrix} = (\mathbf{x}^T\mathbf{a_1}, \ldots, \mathbf{x}^T\mathbf{a_p})$$

$$= (x_1, \ldots, x_n)\begin{pmatrix} \mathbf{a_1}^T \\ \vdots \\ \mathbf{a_n}^T \end{pmatrix} = x_1\mathbf{a_1}^T + \cdots + x_n\mathbf{a_n}^T$$

is a $1 \times n$ row vector.

# Calculation of vectors and matrices

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_p)^T$ be vectors and $\mathbf{A} = (a_{ij})_{n \times p}$ . Then we define

6.

$$\mathbf{Ay} = \begin{pmatrix} a_{11} & \ldots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{a_1} & \ldots & \mathbf{a_p} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \mathbf{a_1} y_1 + \cdots + \mathbf{a_p} y_p$$

$$= \begin{pmatrix} \mathbf{a_1}^T \\ \vdots \\ \mathbf{a_n}^T \end{pmatrix} \mathbf{y} = \begin{pmatrix} \mathbf{a_1}^T \mathbf{y} \\ \vdots \\ \mathbf{a_n}^T \mathbf{y} \end{pmatrix}$$

is a $p \times 1$ column vector.

# Calculation of vectors and matrices

Let $\mathbf{A} = (a_{ij})_{n \times p}$ be a matrix. Then we define

7. $\mathbf{A}^T = (a_{ji})_{p \times n}$ is said to be the transpose of matrix $\mathbf{A}$.

<u>Remark</u>

If $\mathbf{A}$ is a squared matrix and $\mathbf{A} = \mathbf{A}^T$, then $\mathbf{A}$ is said to be a symmetric matrix.

e.g.

Let $\mathbf{x} = (1, 2)^T$, $\mathbf{y} = (-1, 0)^T$, $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

# Calculation of vectors and matrices

Let $\mathbf{A} = (a_{ij})_{n\times p}$ and $\mathbf{B} = (b_{ij})_{n\times p}$ be a matrix.

8.

$$\mathbf{A}\mathbf{B}^T = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1p} & \dots & b_{np} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1p}b_{1p} & \dots & a_{11}b_{n1} + \cdots + a_{1p}b_{np} \\ \vdots & \ddots & \vdots \\ a_{n1}b_{11} + \cdots + a_{np}b_{1p} & \dots & a_{n1}b_{n1} + \cdots + a_{np}b_{np} \end{pmatrix}$$

e.g.

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$

# Calculation of vectors and matrices

Let $\mathbf{A} = (a_{ij})_{n \times p}$ and $\mathbf{B} = (b_{ij})_{n \times p}$ be a matrix.

8.

$$
\begin{aligned}
\mathbf{AB}^T &= \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1p} & \dots & b_{np} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{a_1} & \dots & \mathbf{a_p} \end{pmatrix} \begin{pmatrix} \mathbf{b_1}^T \\ \vdots \\ \mathbf{b_p}^T \end{pmatrix} = \mathbf{a_1 b_1}^T + \cdots + \mathbf{a_p b_p}^T \\
&= \begin{pmatrix} \mathbf{a_1}^T \\ \vdots \\ \mathbf{a_n}^T \end{pmatrix} \begin{pmatrix} \mathbf{b_1} & \dots & \mathbf{b_n} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1}^T \mathbf{b_1} & \dots & \mathbf{a_1}^T \mathbf{b_n} \\ \vdots & \ddots & \vdots \\ \mathbf{a_n}^T \mathbf{b_1} & \dots & \mathbf{a_n}^T \mathbf{b_n} \end{pmatrix}
\end{aligned}
$$

# Calculation of vectors and matrices

Let $\mathbf{A} = (a_{ij})_{n\times p}$ and $\mathbf{B} = (b_{ij})_{n\times p}$ be a matrix.

8.

$$\begin{aligned}
\mathbf{A}^T\mathbf{B} &= \begin{pmatrix} a_{11} & \dots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1p} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{np} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{a_1}^T \\ \vdots \\ \mathbf{a_p}^T \end{pmatrix} \begin{pmatrix} \mathbf{b_1} & \dots & \mathbf{b_p} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1}^T\mathbf{b_1} & \dots & \mathbf{a_1}^T\mathbf{b_p} \\ \vdots & \ddots & \vdots \\ \mathbf{a_p}^T\mathbf{b_1} & \dots & \mathbf{a_p}^T\mathbf{b_p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{a_1} & \dots & \mathbf{a_n} \end{pmatrix} \begin{pmatrix} \mathbf{b_1}^T \\ \vdots \\ \mathbf{b_n}^T \end{pmatrix} = \mathbf{a_1}\mathbf{b_1}^T + \dots + \mathbf{a_n}\mathbf{b_n}^T
\end{aligned}$$

# Calculation of vectors and matrices

Remark

$$(\mathbf{A}^T\mathbf{B})^T = \mathbf{B}^T(\mathbf{A}^T)^T = \mathbf{B}^T\mathbf{A}$$

Transpose each matrices (or vectors) and flip the order!

# Calculation of vectors and matrices

Definition

The squared matrix $\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{n\times n}$ is called the identity matrix, denoted by $\mathbf{I}_n$.

9. For any matrix $\mathbf{A} = (a_{ij})_{n\times p}$, or any vector $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{A} = \mathbf{I}_n\mathbf{A} = \mathbf{A}\mathbf{I}_p$$

$$\mathbf{v} = \mathbf{I}_n\mathbf{v}$$

# Linear combination

Definition

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ and $c_1, \ldots, c_n$ are constants. Then,

$$\mathbf{v} = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$$

is said to be a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Definition

The hyperplane spanned by vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is defined as the collection of all the points which are pointed by some vector which is a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from the origin.

# Trace

Definition

Let $\mathbf{A} = (a_{ij})_{n \times n}$ be a squared matrix. The trace of $\mathbf{A}$, denoted by $\mathrm{tr}(\mathbf{A})$, is defined as the sum of the diagonal entries, i.e.

$$\mathrm{tr}(\mathbf{A}) = a_{11} + \cdots + a_{nn}$$

Property

Let $\mathbf{A}$ be a $m \times n$ matrix, $\mathbf{B}$ be a $n \times m$ matrix and $c$ be a constant, then

1. $\mathrm{tr}(c\mathbf{A} + \mathbf{B}) = c\,\mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ if $n = m$
2. $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ (verify by yourself!)

# Length

Definition

Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ be a vector. The length of $\mathbf{x}$ is defined as

$$|\mathbf{x}| = \sqrt{x_1^2 + \cdots + x_p^2} = (\mathbf{x}^T\mathbf{x})^{\frac{1}{2}}$$

Remark

$\mathbf{x}^T\mathbf{x}$ is said to be a quadratic form of $\mathbf{x}$.

Definition

A unit vector is a vector with length 1.
e.g. $(1, 0)$, $(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$.

# Projection I

Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ and $\mathbf{v} = (v_1, \ldots, v_p)^T$ are vectors in $\mathbb{R}^p$.
The projection of $\mathbf{x}$ on $\mathbf{v}$ is a vector defined as

$$\left(\frac{\mathbf{x}^T\mathbf{v}}{|\mathbf{v}|}\right) \cdot \frac{\mathbf{v}}{|\mathbf{v}|}$$

Remark
If $\mathbf{v}$ is a unit vector, then the projection of $\mathbf{x}$ on $\mathbf{v}$ is $(\mathbf{x}^T\mathbf{v})\mathbf{v}$.

# Orthogonality

Definition

Let $\mathbf{v} = (v_1, \ldots, v_p)^T$ and $\mathbf{w} = (w_1, \ldots, w_p)^T$ are vectors in $\mathbb{R}^p$, then

1. $\mathbf{v}$ and $\mathbf{w}$ are said to be orthogonal if $\mathbf{v}^T\mathbf{w} = 0$.
2. $\mathbf{v}$ and $\mathbf{w}$ are said to be orthonormal if $\mathbf{v}^T\mathbf{w} = 0$ and both $\mathbf{v}$ and $\mathbf{w}$ are unit vectors, i.e. $|\mathbf{v}| = |\mathbf{w}| = 1$.
3. A $n \times n$ matrix $\mathbf{V}$ is said to be orthogonal if $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$

# Projection II

Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ and $\mathbf{V} = \begin{pmatrix} \mathbf{v_1} & \ldots & \mathbf{v_q} \end{pmatrix}$ where $\mathbf{v_i}$ are vectors in $\mathbb{R}^p$ and $\mathbf{V}$ is orthogonal, i.e. $\mathbf{v_i}$ are mutually orthonormal. Then, the projection of $\mathbf{x}$ on the hyperplane spanned by $\mathbf{v_1}, \ldots, \mathbf{v_q}$ is defined as

$$
\begin{aligned}
(\mathbf{x}^T\mathbf{V})\mathbf{V}^T &= \mathbf{x}^T \begin{pmatrix} \mathbf{v_1} & \ldots & \mathbf{v_q} \end{pmatrix} \begin{pmatrix} \mathbf{v_1}^T \\ \vdots \\ \mathbf{v_q}^T \end{pmatrix} \\
&= (\mathbf{x}^T\mathbf{v_1})\mathbf{v_1}^T + \cdots + (\mathbf{x}^T\mathbf{v_q})\mathbf{v_q}^T
\end{aligned}
$$

Remarks

$(\mathbf{x}^T\mathbf{V})\mathbf{V}^T$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_q$

# Examples

1. Let $\mathbf{x} = (1, 2, 3)^T$ and $\mathbf{V} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

# Examples

1. Let $\mathbf{x} = (-1, 0, 1, 2)^T$ and $\mathbf{V} = \begin{pmatrix} 0 & \frac{3}{5} \\ -1 & \frac{4}{5} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

## Eigenvalues and eigenvectors

Let $\mathbf{A}_{p \times p}$ be a matrix and $\mathbf{y}$ be a non-zero $p \times 1$ vector such that

$$\mathbf{A}\mathbf{y} = \lambda \mathbf{y}$$

where $\lambda$ is a constant (scale).
Then, $\mathbf{y}$ is called a eigenvector of $\mathbf{A}$ and $\lambda$ is the corresponding eigenvalue of $\mathbf{A}$.

# Eigenvalues and eigenvectors

<u>Property</u>

$\lambda$ is an eigenvalue of $\mathbf{A}_{p \times p}$ if and only if $\det(\mathbf{A} - \lambda \mathbf{I}_p) = 0$

<u>Property</u>

If $\mathbf{y}$ is an eigenvector of a matrix $\mathbf{A}$, then for any constant $c$, vector $c\mathbf{y}$ is also an eigenvector of a matrix $\mathbf{A}$ with the same associated eigenvalue.

<u>Remark</u>

Any eigenvector can be scaled to be a unit vector.

# Eigendecomposition

**Spectral decomposition theorem**

Any symmetric matrix $\mathbf{A}_{p\times p}$ can be expressed as

$$\mathbf{A} = \Gamma\Lambda\Gamma^T = \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_p \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_p^T \end{pmatrix}$$

$$= \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

where
$\Lambda$ is a diagonal matrix of eigenvalues $\lambda_i$ of $\mathbf{A}$ and
$\Gamma$ is an **orthogonal** matrix whose column vectors $\mathbf{v}_i$ are corresponding eigenvectors.

# Mean and variance

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{1}_n = (1, \ldots, 1)$.

The mean of vector $\mathbf{x}$ is defined as

$$E(\mathbf{x}) = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}_n^T\mathbf{x}$$

The variance of vector $\mathbf{x}$ is defined as

$$\begin{aligned}
Var(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}(x_i - E(\mathbf{x}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\mathbf{1}_n^T\mathbf{x})^2 \\
&= \frac{1}{n}(\mathbf{x} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x})^T(\mathbf{x} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x}) \\
&= \frac{1}{n}\mathbf{x}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{x} \\
&= \frac{1}{n}\mathbf{x}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{x}
\end{aligned}$$

# Mean and variance

Let $\mathbf{X} = \begin{pmatrix} \mathbf{x_1} & \dots & \mathbf{x_p} \end{pmatrix}_{n \times p}$ be a matrix (data).

The mean of matrix $\mathbf{X}$ is defined as

$$\begin{aligned} E(\mathbf{X}) &= (E(\mathbf{x}_1), \dots, E(\mathbf{x}_p)) \\ &= (\frac{1}{n}\mathbf{1}_n^T \mathbf{x_1}, \dots, \frac{1}{n}\mathbf{1}_n^T \mathbf{x_p}) \\ &= \frac{1}{n}\mathbf{1}_n^T \mathbf{X} \end{aligned}$$

# Mean and variance

Let $\mathbf{X} = \begin{pmatrix} \mathbf{x_1} & \dots & \mathbf{x_p} \end{pmatrix}_{n \times p}$ be a matrix (data).

The covariance matrix of matrix $\mathbf{X}$ is defined as

$$var(\mathbf{X}) = \mathbf{S}_n^2 = \frac{1}{n}\mathbf{X}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{X}$$
$$= \begin{pmatrix} var(\mathbf{x_1}) & cov(\mathbf{x_1}, \mathbf{x_2}) & \dots & cov(\mathbf{x_1}, \mathbf{x_p}) \\ cov(\mathbf{x_2}, \mathbf{x_1}) & var(\mathbf{x}_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ cov(\mathbf{x_p}, \mathbf{x_1}) & \dots & \dots & var(\mathbf{x}_p) \end{pmatrix}$$

where

$$cov(\mathbf{x_i}, \mathbf{x_j}) = \frac{1}{n}(\mathbf{x_i} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x_i})^T(\mathbf{x_j} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x_j})$$
$$= \frac{1}{n}\mathbf{x_i}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{x_j}$$

# Properties of covariance matrix

Properties

1. $\mathbf{S}_n^2$ is a positive semidefinite matrix, i.e. the eigenvalues of $\mathbf{S}_n^2$ are all non-negative.
2. $\mathbf{S}_n^2$ can apply spectual decomposition theorem

$$\mathbf{S}_n^2 = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$$

   Moreover, $\lambda_1, \ldots, \lambda_p$ are non-negative and $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are orthonormal pairwisely.
3. The **total variance** is the sum of the diagonal elements of the covaiance matrix, i.e. $tr(\mathbf{S}_n^2)$. By the property of trace,

$$tr(\mathbf{S}_n^2) = \lambda_1 + \cdots + \lambda_p$$

Remark

Any $\mathbf{v} \in \mathbb{R}^p$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, i.e.

$$v = a_1\mathbf{v}_1 + \cdots + a_p\mathbf{v}_p$$

for some $a_1, \ldots, a_p$ are constants.

# Principal component analysis I

Property

Let $\mathbf{X} = \begin{pmatrix} \mathbf{x_1} & \dots & \mathbf{x_p} \end{pmatrix}_{n \times p}$ be data, then for any $\mathbf{v} \in \mathbb{R}^p$ is a unit vector,

$$var(\mathbf{X}\mathbf{v_1}) \geq var(\mathbf{X}\mathbf{v})$$

where $\mathbf{v}_1$ is the eigenvector of $\mathbf{S}_n^2$ associated with the largest eigenvalue $\lambda_1$.

Moreover, $var(\mathbf{X}\mathbf{v}_1) = \lambda_1$.

# Principal component analysis I

Proof

Claim: $var(\mathbf{X}\mathbf{v_1}) \geq var(\mathbf{X}\mathbf{v})$

First,

$$\begin{aligned} var(\mathbf{X}\mathbf{v}) &= \frac{1}{n}(\mathbf{X}\mathbf{v})^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{X}\mathbf{v} \\ &= \frac{1}{n}\mathbf{v}^T\mathbf{X}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{X}\mathbf{v} \\ &= \mathbf{v}^T\mathbf{S}_n^2\mathbf{v} \end{aligned}$$

# Principal component analysis I

Second, by eigendecomposition,

$$\mathbf{S}_n^2 = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

say that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$.
Let $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_p\mathbf{v}_p$ be a unit vector (since $\mathbf{v}$ must be the linear combination of $\mathbf{v_1}, \ldots, \mathbf{v_p}$), i.e. $a_1^2 + \cdots + a_p^2 = 1$.

$$\begin{aligned} var(\mathbf{Xv}) &= \mathbf{v}^T(\sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^T)\mathbf{v} = \sum_{i=1}^{p} \lambda_i (\mathbf{v}^T \mathbf{v}_i)^2 \\ &= \sum_{i=1}^{p} \lambda_i (a_i \mathbf{v}_i^T \mathbf{v}_i)^2 \quad \text{since orthogonality of } \mathbf{\Gamma} \\ &= \sum_{i=1}^{p} \lambda_i (a_i)^2 \leq \lambda_1 \quad \text{since } a_1^2 + \cdots + a_p^2 = 1 \end{aligned}$$

# Principal component analysis I

Note that

$$var(\mathbf{Xv}) \leq \lambda_1$$

So the $var(\mathbf{Xv})$ has the maximum, i.e. $var(\mathbf{Xv}) = \lambda_1$, if the equality holds. The equality holds as $\mathbf{v} = \mathbf{v}_1$.

Thus, data projected on the eigenvector corresponding to the largest eigenvalue has the maximal variance, and the variance is the eigenvalue. $\square$

# Principal component analysis I

Remarks

1. The data projected on the eigenvector associated with the largest eigenvalue has the largest variance than the data projected on other direction.
2. Note that if the data has been centralized, i.e. $\mathbf{1}_n^T \mathbf{x}_i =$ sum of the values in the $i$-th colum vector $= 0$, the covariance matrix $\mathbf{S}_n^2 = \frac{1}{n} \mathbf{X}^T \mathbf{X}$.
3. The first principal component (site score) of $j$-th data $\mathbf{x}_j^T$ is defined as the projection value of the $j$-th data on the first eigenvector of the covariance matrix, i.e. $\mathbf{x}_j^T \mathbf{v}_1$.
4. The first species score of $i$-th species is defined as the projection of $i$-th axis on the first eigenvector of the covariance matrix, i.e. $\mathbf{e}_i^T \mathbf{v}_1 = v_{1i}$.

# Principal component analysis II

Property

Let $\mathbf{X}_{n \times p}$ be data after centralized and $\mathbf{Y}_{n \times p}$ is a orthogonal projection of $\mathbf{X}$ on hyperplane of dimension $q$ in $\mathbb{R}^p$, i.e.

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{B}^T$$

for some $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_q \end{pmatrix}_{p \times q}$ with $\mathbf{B}^T\mathbf{B} = \mathbf{I}_q$. Then, the mean squared error which is defined as

$$MSE(\mathbf{X}, \mathbf{Y}) = \frac{1}{n} \sum_{i=1}^{n} |\mathbf{x}_i^T - \mathbf{y}_i^T|^2$$

is minimized for $\mathbf{Y} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$ where $\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_q \end{pmatrix}$ and $\mathbf{v}_i$ is the eigenvector of $\mathbf{S}_n^2$ associated with the $i$-th largest eigenvalues.

# Principal component analysis II

Moreover,

$$MSE(\mathbf{Y},\mathbf{X}) = (\sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{q} \lambda_i)/n$$

where $\lambda_i$ is the $i$-th largest eigenvalues.

# Conclusions

Consider a species composition matrix (site $\times$ species) $\mathbf{X}_{n \times p}$. The plots are in the $p$-dimensional species space. If we want to project the plots on a **lower two dimensional plane**,

1. The projection on the plane that spanned by the eigenvectors corresponding to the two largest eigenvalues of the covariance matrix of species composition matrix **preserve the most variation**.
2. The **explained variation** of the two PCA axes are the sum of the two largest eigenvalues of the covariance matrix. The **total variance** is the sum of all the eigenvalues of the covariance matrix.
3. The **site score** of $j$-th data is the projection value of the $j$-th data on the PCA axes, i.e. the eigenvectors of the covariance matrix.
4. The **species score** of the $i$-th species is the projection value of the original $i$-th axis on the PCA axes, i.e. the eigenvectors of the covariance matrix.