Principal component analysis

Ching-Lin Huang (Andy)

Institute of Ecology and Evolutionary Biology, NTU

March 24, 2022

Content

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

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

$$\underbrace{\left(\frac{\mathbf{x}^T\mathbf{v}}{|\mathbf{v}|}\right)}_{\text{length}} \cdot \underbrace{\frac{\mathbf{v}}{|\mathbf{v}|}}_{\text{direction}}$$

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

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 $\underbrace{\mathbf{V}^T\mathbf{V}}_{\text{quadratic form}} = \mathbf{I}_n$

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

$$
\begin{aligned}
(\mathbf{x}^T\mathbf{V})\mathbf{V}^T &= \mathbf{x}^T \begin{pmatrix} \mathbf{v_1} & \ldots & \mathbf{v_q} \end{pmatrix} \begin{pmatrix} \mathbf{v_1}^T \\ \vdots \\ \mathbf{v_q}^T \end{pmatrix} \\
&= \underbrace{(\mathbf{x}^T\mathbf{v_1})\mathbf{v_1}^T}_{\text{proj of } x \text{ on } v_1} + \cdots + \underbrace{(\mathbf{x}^T\mathbf{v_q})\mathbf{v_q}^T}_{\text{proj of } x \text{ on } v_q}
\end{aligned}
$$

Remarks

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

$c_1 v_1^T + \cdots + c_q v_q^T$

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 $\mathbf{V} = \begin{pmatrix} 0 & \frac{3}{5} \\ -1 & \frac{4}{5} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

$v_1 \quad v_2$

proj of x on the plane spanned by v_1 and v_2

$= (x^T v_1) v_1 + (x^T v_2) v_2 = 0 \cdot v_1 + \left(-\frac{3}{5}\right) v_2$

$= -\frac{3}{5}\left(\frac{3}{5}, \frac{4}{5}, 0, 0\right)$

$x^T v_1 = (-1 \ 0 \ 1 \ 2)\begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} = 0$

$x^T v_2 = (-1 \ 0 \ 1 \ 2)\begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 0 \\ 0 \end{pmatrix} = -\frac{3}{5}$

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

matrix value

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

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

$A = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix}$ claim the eigenvalues are $\left.\begin{matrix} \lambda_1 = 2 \\ \lambda_2 = -5 \end{matrix}\right]$

$y = (y_1, y_2)^T$ $\quad Ay = 2y$ or $Ay = -5y$

$$\begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 + 2y_2 \\ 3y_1 - 4y_2 \end{pmatrix} = \begin{pmatrix} 2y_1 \\ 2y_2 \end{pmatrix} \Rightarrow \begin{cases} 2y_2 = y_1 \\ 3y_1 = 6y_2 \end{cases}$$

Say $y_1 = 1$, then $y_2 = 0.5$, $y = (1, 0.5) \Rightarrow Ay = 2y$

$y_1 = 2$, then $y_2 = 1$ $y = (2, 1)$ ← eigenvector corresponding to $\lambda_1 = 2$

Eigenvalues and eigenvectors

determinant

characteristic function

Property

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

Property

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.

Remark

Any eigenvector can be scaled to be a unit vector.

$$A=\begin{pmatrix}1 & 2\\ 3 & -4\end{pmatrix}.\quad A-\lambda I=\begin{pmatrix}1 & 2\\ 3 & -4\end{pmatrix}-\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}=\begin{pmatrix}1-\lambda & 2\\ 3 & -4-\lambda\end{pmatrix}$$

$$\det\begin{pmatrix}1-\lambda & 2\\ 3 & -4-\lambda\end{pmatrix}=(1-\lambda)(-4-\lambda)-6=\lambda^2+4\lambda-\lambda-4-6$$

$$=\lambda^2+3\lambda-10=(\lambda+5)(\lambda-2)=0$$

$$\Rightarrow \lambda=-5 \text{ or } 2$$

Eigendecomposition

$A = A^T$ $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$

By definition
$A v_1 = \lambda_1 v_1$

Spectral decomposition theorem

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

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

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

rank

$v_i \perp v_j$ if $i \neq j$

$$\Gamma^T \Gamma = \begin{pmatrix} v_1^T \\ \vdots \\ v_p^T \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

Lambda

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

Gamma

Mean and variance

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

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

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

$$(1\ 1 \cdots 1)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

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

$$\begin{aligned}
Var(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}(x_i - E(\mathbf{x}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\mathbf{1}_n^T\mathbf{x})^2 = \frac{1}{n}\left((x_1 - \frac{1}{n}1_n^T x)^2 + \cdots + (x_n - \frac{1}{n}1_n^T x)^2\right) \\
&= \frac{1}{n}(\mathbf{x} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x})^T(\mathbf{x} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{x}) \\
&= \frac{1}{n}\mathbf{x}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{x} \\
&= \frac{1}{n}\mathbf{x}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{x}
\end{aligned}$$

$$(x_1 - \frac{1}{n}1_n^T x, \ldots, x_n - \frac{1}{n}1_n^T x)^T$$

$$\frac{1}{n}1_n 1_n^T x = \left(\frac{1}{n}1_n^T x, \ldots, \frac{1}{n}1_n^T x\right)$$

idompotent matrix.

Mean and variance

species

site abundance

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

species

site abun.

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

covariance

$$var(\mathbf{X}) = \boxed{\mathbf{S}_n^2} = \frac{1}{n}\mathbf{X}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{X}$$

symmetric

$$= \begin{pmatrix} var(\mathbf{x_1}) & cov(\mathbf{x_1}, \mathbf{x_2}) & \dots & cov(\mathbf{x_1}, \mathbf{x_p}) \\ cov(\mathbf{x_2}, \mathbf{x_1}) & var(\mathbf{x}_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ cov(\mathbf{x_n}, \mathbf{x_1}) & \dots & \dots & var(\mathbf{x_p}) \end{pmatrix}$$

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

for any vector y $y^T S_n^2 y \ge 0 \Rightarrow S_n^2$ p.s.d.

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

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

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$$

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,

trace

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

$$tr(AB) = tr(BA)$$

$$tr(\Gamma\Lambda\Gamma^T) = tr(\underbrace{\Gamma^T\Gamma}_{I_p}\Lambda) = tr(\Lambda)$$

Remark

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

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

e.g. $A = (\ldots)$, $y_1 = (\ldots) \leftrightarrow \lambda_1 = 2$, $y_2 = (\ldots) \leftrightarrow \lambda_2 = 5$

$v = a_1 y_1 + a_2 y_2 = (\ldots) = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

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

Principal component analysis I

$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ proj on a vector v $\begin{pmatrix} x_1^T v \\ \vdots \\ x_n^T v \end{pmatrix} = Xv$

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,

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

where $\mathbf{v}_1$ is the eigenvector of $\mathbf{S}_n^2$ associated with the largest eigenvalue λ_1.

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

Principal component analysis I

any vector v.

Proof

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

First,

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

$S_n^2 = \frac{1}{n} X^T(I_n - \frac{1}{n} 1_n 1_n^T) X$

Principal component analysis I

Second, by eigendecomposition,

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

→ eigenvalue

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_p \end{pmatrix}$$

$$\Gamma = (v_1 \cdots v_p)$$

→ eigenvectors

say that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$.

Let $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_p\mathbf{v}_p$ (since $\mathbf{v}$ must be the linear combination of $\mathbf{v_1}, \ldots, \mathbf{v_p}$)

unit vector, $v_i \perp v_j$

S_n

$$var(\mathbf{X}\mathbf{v}) = \frac{1}{n}\mathbf{v}^T\left(\boxed{\sum_{i=1}^{p}\lambda_i\mathbf{v}_i\mathbf{v}_i^T}\right)\mathbf{v} = \frac{1}{n}\sum \lambda_i (v^T v_i)(v_i^T v)$$

$v_i^T v_j = 0$

$$= \frac{1}{n}\sum_{i=1}^{p}\lambda_i\boxed{(\mathbf{v}^T\mathbf{v}_i)^2} = \frac{1}{n}\sum \lambda_i \left[(a_1 v_1 + \cdots + a_p v_p)^T v_i\right]^2$$

$$\left[(a_1 v_1 + \cdots + a_p v_p)^T v_i\right]$$

$$= \left[a_1 v_1^T v_i + \cdots + a_p v_p^T v_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{p}\lambda_i(a_i\mathbf{v}_i^T\mathbf{v}_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{p}\lambda_i(a_i)^2 \leq \lambda_1 \quad \text{since } a_1^2 + \cdots + a_p^2 = 1$$

Principal component analysis I

Note that

$$var(\mathbf{X}\mathbf{v}) \leq \lambda_1$$

So the $var(\mathbf{X}\mathbf{v})$ has the maximum, i.e. $var(\mathbf{X}\mathbf{v}) = \lambda_1$, 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 $\mathbf{Y} = \mathbf{\Gamma}\mathbf{\Gamma}^T$ where $\mathbf{\Gamma} = \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_q \end{pmatrix}$ 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 λ_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.