Numerical Methods in Bioinformatics

Principal Component Analysis

Xiuxia Du

`Xiuxia.Du@uncc.edu`

Department of Bioinformatics & Genomics

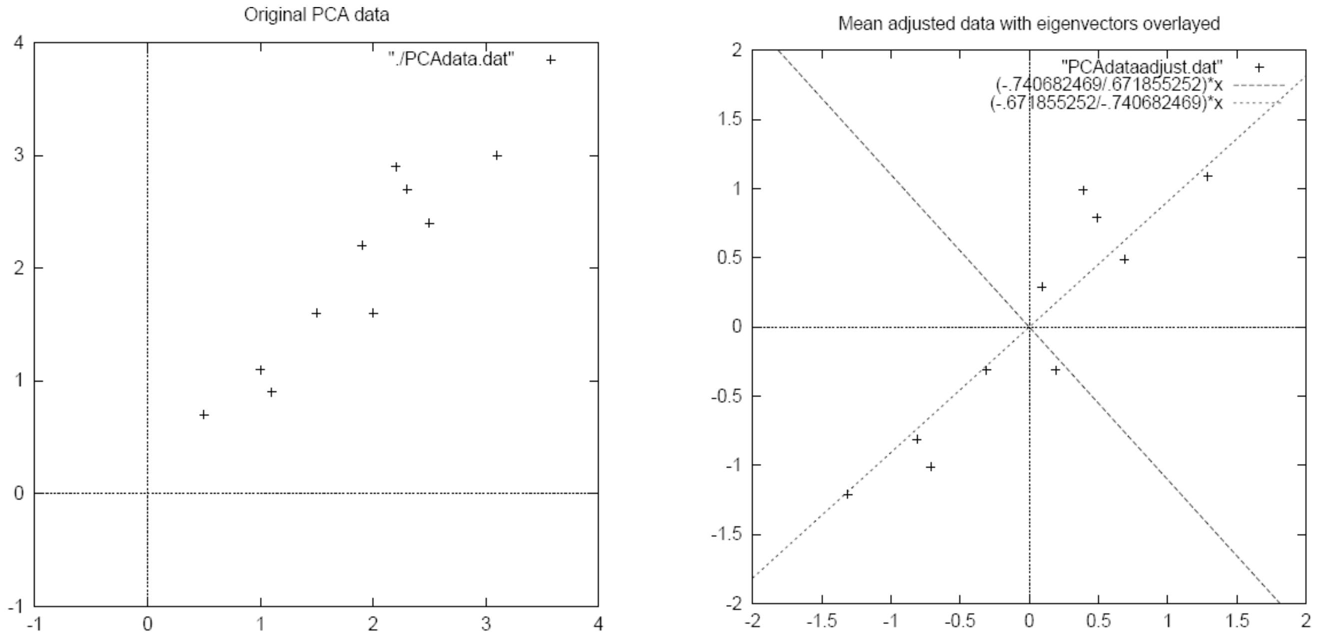
University of North Carolina at Charlotte

Charlotte, NC 28223, USA

Motivation

Case 1: Given the following data [a]:

x	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.1
y	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9



We can see that x and y are correlated.

---

[a] A tutorial on Principal Component Analysis, Lindsay I Smith

Motivation

Case 2: Given the following micro-array data:

gene	assay 1	assay 2	assay 3	$\cdots$
gene 1	2.0	2.1	1.9	$\cdots$
gene 2	4.3	3.9	4.5	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
gene n	8.9	9.1	8.8	$\cdots$

Question: Can we find the most meaningful basis to re-express the data?

Answer: Principal Component Analysis (PCA).

PCA

- Transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components.. PCA is also called Karhunen-Loeve transform or Hotelling transform.

PCA involves:

- covariance matrix
- linear transformation
- dot product
- basis of a vector space
- eigenvalues and eigenvectors
- matrix diagonalization

Review: Random Variable

- mean: expectation of a random variable.

$$\mu_x = \mathbb{E}\left(\boldsymbol{X}\right)$$

- variance:

$$\textit{var}\left(\boldsymbol{X}\right) = \mathbb{E}\left[\left(\boldsymbol{X} - \mu_x\right)^2\right] = \sigma_x^2$$

- covariance:

$$\begin{aligned}\textit{cov}\left(\boldsymbol{X}, \boldsymbol{Y}\right) &= \mathbb{E}\left[\left(\boldsymbol{X} - \mu_x\right)\left(\boldsymbol{Y} - \mu_y\right)\right] \\ &= \mathbb{E}\left(\boldsymbol{X}\boldsymbol{Y}\right) - \mathbb{E}\left(\boldsymbol{X}\right)\mathbb{E}\left(\boldsymbol{Y}\right)\end{aligned}$$

Random variables whose covariance is zero are called uncorrelated.

Review: Covariance Matrix

Given a random vector $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n]$, the covariance between all possible pairs of the randaom variables can be written in a covariance matrix

$$\mathit{cov}\,(\boldsymbol{X}) = \begin{bmatrix} \mathit{var}\,(\boldsymbol{X}_1) & \mathit{cov}\,(\boldsymbol{X}_1, \boldsymbol{X}_2) & \cdots & \mathit{cov}\,(\boldsymbol{X}_1, \boldsymbol{X}_n) \\ \mathit{cov}\,(\boldsymbol{X}_2, \boldsymbol{X}_1) & \mathit{var}\,(\boldsymbol{X}_2) & \cdots & \mathit{cov}\,(\boldsymbol{X}_2, \boldsymbol{X}_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathit{cov}\,(\boldsymbol{X}_n, \boldsymbol{X}_1) & \mathit{cov}\,(\boldsymbol{X}_n, \boldsymbol{X}_2) & \cdots & \mathit{var}\,(\boldsymbol{X}_n) \end{bmatrix}$$

From the covariance matrix, we can derive a transformation matrix that can be used to de-correlate the data. This approach is called PCA.

Review: Discrete Random Variable

- mean:

$$\mu_x = \sum_i x_i p_i, \qquad \bar{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- variance:

$$\sigma_x^2 = \sum_i p_i \left(x_i - \mu_x\right)^2, \qquad s_x^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{\boldsymbol{X}}\right)^2}{n-1}$$

- covariance:

$$s_{xy} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{\boldsymbol{X}}\right)\left(y_i - \bar{\boldsymbol{Y}}\right)}{n-1}$$

Linear Transformation

Let S and T be two vector spaces over the same field. A function $f : S \rightarrow T$ is said to be a linear map or linear transformation if for any two vectors x_1 and x_2 in S and any scalar, the following two conditions are satisfied:

$$f(x_1 + x_2) = f(x_1) + f(x_2)$$
$$f(ax_1) = af(x_1)$$

This is equivalent to requiring that for any vectors $x_1, x_2, \cdots, x_m$ and scalars $a_1, a_2, \cdots, a_m$:

$$f(a_1x_1 + a_2x_2 + \cdots + a_mx_m) = a_1f(x_1) + a_2f(x_2) + \cdots + a_mf(x_m)$$

Linear Transformation

Examples:

1. The map $x \longmapsto 2x$ is linear.
2. The map $x \longmapsto x^2$ is nonlinear.
3. The integral yields a linear map from the space of all real-valued integrable functions on some interval to $\mathbb{R}$.

$$\int_a^b (\alpha f(t) + \beta g(t))\, dt = \alpha \int_a^b f(t)dt + \beta \int_a^b g(t)dt$$

4. Differentiation is a linear map from the space of all differentiable functions to the space of all functions.

$$\frac{d}{dt}(\alpha f(t) + \beta g(t)) = \alpha f'(t) + \beta g'(t)$$

Linear Transformation

- An $m \times n$ matrix $\boldsymbol{A}$ defines a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ by sending the column vector $\boldsymbol{X} \in \mathbb{R}^n$ to the column vector $\boldsymbol{AX} \in \mathbb{R}^m$.
- In linear algebra, every linear transformation between finite-dimensional spaces can be expressed as a matrix.

Example

$$(a).\ \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & 8 & 6 & 4 \\ 3 & 10 & 8 & 8 \\ 4 & 12 & 10 & 6 \end{bmatrix}, \quad (b).\ \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & 8 & 6 & 4 \\ 3 & 10 & 8 & 8 \end{bmatrix}, \quad (c).\ \begin{bmatrix} 1 & 2 & 4 \\ 2 & 8 & 6 \\ 3 & 10 & 8 \\ 4 & 12 & 10 \end{bmatrix}.$$

Linear Transformation

Special cases of linear transformations in $\mathbb{R}^2$.

- rotation by θ degrees clockwise

$$\boldsymbol{A} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

- scaling by 2 in all directions

$$\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- projection onto the y axis

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Dot Product

Given two vectors $\boldsymbol{a} = [a_1, a_2, \cdots, a_n]$ and $\boldsymbol{b} = [b_1, b_2, \cdots, b_n]$ in the Euclidean space, their dot product is defined as:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

Geometric interpretation:

- $$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\, |\boldsymbol{b}| \cos(\theta)$$

  where θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$. Clearly,

  $$\boldsymbol{a} \perp \boldsymbol{b} \Leftrightarrow \boldsymbol{a} \cdot \boldsymbol{b} = 0$$

  When $\boldsymbol{a} \perp \boldsymbol{b}$, we say that they are orthogonal to each other.

Dot Product

- Scalar projection

$$|\boldsymbol{a}| = |\boldsymbol{b}| = 1 \Rightarrow \boldsymbol{a} \cdot \boldsymbol{b} = \cos(\theta)$$

$$|\boldsymbol{b}| = 1 \Rightarrow \boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}| \cos(\theta)$$

projection of $\boldsymbol{a}$ in the direction of $\boldsymbol{b}$

Example

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} [a_{11}, a_{12}, a_{13}] \cdot [x_1, x_2, x_3] \\ [a_{21}, a_{22}, a_{23}] \cdot [x_1, x_2, x_3] \\ [a_{31}, a_{32}, a_{33}] \cdot [x_1, x_2, x_3] \end{bmatrix}$$

Basis of a Vector Space

A basis of a vector space $\boldsymbol{V}$ is a linearly independent subset of $\boldsymbol{V}$ that spans $\boldsymbol{V}$. Specifically, the basis satisfies the following two properties (assume that the vector space $\boldsymbol{V}$ is over the field of $\mathbb{R}$):

- The linear independence property: for all $a_1, \cdots, a_n \in \mathbb{R}$, if $a_1\boldsymbol{v_1} + \cdots + a_n\boldsymbol{v}_n = 0$, then $a_1 = \cdots = a_n = 0$.
- The spanning property: for every $\boldsymbol{x}$ in $\boldsymbol{V}$ it is possible to choose $a_1, \cdots, a_n \in \mathbb{R}$ such that $\boldsymbol{x} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$.

The numbers $a_i, i = 1, \cdots, n$ are called the coordinates of the vector $\boldsymbol{x}$ with respect to the basis and they are uniquely determined.

Basis of a Vector Space

The number of vectors that the basis contains is the dimension of the vector space.

Example

- For $\mathbb{R}^2$, $\boldsymbol{e}_1 = (1, 0)$ and $\boldsymbol{e}_1 = (1, 0)$ form a basis and the dimension of $\mathbb{R}^2$ is 2.
- For $\mathbb{R}^n$, $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ are linearly independent, generate $\mathbb{R}^n$, and thus form a basis. The dimension of $\mathbb{R}^n$ is n.
- For the vector space of polynomials, a basis is given by $1, x, x^2, \cdots$ and the dimension is countably infinite.

Orthonormal Basis

For a basis $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n]$, if:

$$\begin{aligned} \boldsymbol{v}_i \perp \boldsymbol{v}_j, \quad & \text{for } i, j = 1, \cdots, n \\ \|\boldsymbol{v}_i\| = 1 & \end{aligned}$$

then the basis is orthonormal.

Normalization of a vector: $\boldsymbol{x} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$

In the Euclidean space, the second requirement is $|\boldsymbol{v}_i| = 1$.

Orthonormal Basis

- Let $\boldsymbol{V}_1 = [\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]$ and $\boldsymbol{V}_2 = [\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}]$ be two different orthonormal bases of the same space $\mathbb{R}^3$.
- $\boldsymbol{a}_1 = [a_x, a_y, a_z]$ represents vector $\boldsymbol{a}$ in terms of $\boldsymbol{V}_1$.
- $\boldsymbol{a}_2 = [a_r, a_s, a_t]$ represents the same vector $\boldsymbol{a}$ in terms of $\boldsymbol{V}_2$.

Then $\boldsymbol{a}_2$ can be obtained from $\boldsymbol{a}_1$ by rotation using a rotation matrix $\boldsymbol{P}$:

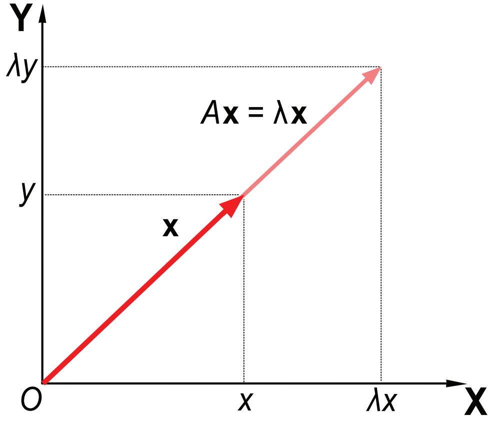
$$\boldsymbol{a}_2 = \boldsymbol{P}\boldsymbol{a}_1 = \begin{bmatrix} r_x & r_y & r_z \\ s_x & s_y & s_z \\ t_x & t_y & t_z \end{bmatrix} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} \boldsymbol{r} \cdot \boldsymbol{a}_1 \\ \boldsymbol{s} \cdot \boldsymbol{a}_1 \\ \boldsymbol{t} \cdot \boldsymbol{a}_1 \end{bmatrix} = \begin{bmatrix} a_r \\ a_s \\ a_t \end{bmatrix}$$

Eigenvalue and Eigenvector

Given a linear transformation $\boldsymbol{A}$, a non-zero vector $\boldsymbol{x}$ is defined to be an eigenvector of the transformation if it satisfies the equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

for some scalar λ. In this situation, the scalar λ is called an eigenvalue of $\boldsymbol{A}$ corresponding to the eigenvector $\boldsymbol{x}$.



Eigenvalue and Eigenvector

- The eigenvector $\boldsymbol{x}$ has the property that its direction is NOT changed by the transformation $\boldsymbol{A}$. It is scaled by a factor of λ.
- Vectors that are not eigenvectors will change direction and magnitude under the transformation. Thus eigenvectors and eigenvalues are special.
- For the identity matrix, all non-zero vectors are eigenvectors.

Eigenvalue and Eigenvector

$$\begin{aligned} \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} &\Rightarrow \boldsymbol{A} - \lambda \boldsymbol{I}\boldsymbol{v} = \boldsymbol{0} \\ &\Rightarrow \left(\boldsymbol{A} - \lambda \boldsymbol{I}\right)\boldsymbol{v} = \boldsymbol{0} \\ &\Rightarrow \text{det}\left(\boldsymbol{A} - \lambda \boldsymbol{I}\right) = 0 \end{aligned}$$

The determinant requirement is called the characteristic equation and the left-hand-side is called the characteristic polynomial.

Eigenvalue and Eigenvector

Let $p(\lambda)$ represent the characteristic polynomial, then $p(\lambda)$ can be factorized as

$$p(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k} = 0$$

where

$$\sum_{i=1}^{k} n_i = n$$

The set of eigenvalues is called the spectrum of $\boldsymbol{A}$. For each eigenvalue λ_i, we have a specific eigenvalue equation

$$(\boldsymbol{A} - \lambda_i \boldsymbol{I})\, \boldsymbol{v} = \mathbf{0}, \quad i = 1 \cdots k$$

Eigenvalue and Eigenvector

There will be $1 \leq m_i \leq n_i$ linearly independent solutions to each eigenvalue equation. The interger n_i and m_i are termed the algebraic and geometric multiplicity of λ_i, respectively.

$$\sum_{i=1}^{k} m_i \leq n$$

Eigenvalue and Eigenvector

Example Compute the eigenvalues and eigenvectors of the following matrix

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Result

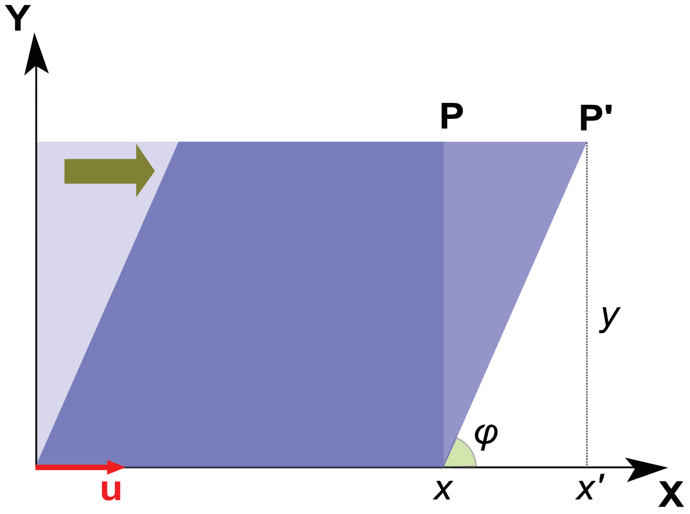
$$\lambda_1 = 1, \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 3, \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Eigenvalue and Eigenvector

Example Shear:

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

Interpretation: $$\lambda = 1, \quad \boldsymbol{x} = \begin{bmatrix} x \\ 0 \end{bmatrix}$$



Eigenvalue and Eigenvector

- The eigenvectors corresponding to different eigenvalues are linearly independent, that is, none of the eigenvectors can be written as a linear combination of other eigenvectors.
- Each eigenvector is associated with one eigenvalue, but one eigenvalue can be associated with an infinite number of eigenvectors.

Matrix Diagonalization

Let $\boldsymbol{A}$ be a square $n \times n$ matrix with n linearly independent eigenvectors $\boldsymbol{p}_i$, $i = 1, \cdots n$. Then $\boldsymbol{A}$ can be factorized as

$$\boldsymbol{A} = \boldsymbol{P}\Lambda\boldsymbol{P}^{-1}$$

where

$$\boldsymbol{P} = [\boldsymbol{p}_1,\ \boldsymbol{p}_2,\ \cdots\ ,\ \boldsymbol{p}_n]$$

and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues.

The eigenvectors are usually normalized, but they need not be.

Matrix Diagonalization

- matrix inverse via eigendecomposition / matrix diagonalization

  $$\boldsymbol{A}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^{-1} \quad \text{and} \quad [\Lambda^{-1}]_{ii} = \frac{1}{\lambda_i}$$

  $\boldsymbol{A}$ can be inverted if and only if $\lambda_i \neq 0$ for any i.

- determinant

  $$\text{det}(\boldsymbol{A}) = \prod_{i=1}^{k} \lambda_i^{n_i}$$

  Note that each eigenvalue is raised to the power n_i, the algebraic multiplicity.

Matrix Diagonalization

- power of a matrix

$$\boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1}, \quad \left[\Lambda^k\right]_{ii} = \lambda_i^k$$

- trace

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} n_i\lambda_i$$

Note that each eigenvalue is multiplied by n_i.

Matrix Diagonalization

- For an $n \times n$ matrix $\boldsymbol{A}$, if there exists n independent eigenvectors, then:
  - these eigenvectors can form a basis for the vector space $\mathbb{R}^n$.
  - matrix $\boldsymbol{A}$ is diagonalizable.
- For an $n \times n$ matrix $\boldsymbol{A}$, if there exists $< n$ independent eigenvectors, then:
  - these eigenvectors can NOT form a basis for the vector space $\mathbb{R}^n$.
  - matrix $\boldsymbol{A}$ is NOT diagonalizable.

Matrix Diagonalization

Example Find a transformation matrix $\boldsymbol{P}$ that can diagonalize

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & -4 & 2 \end{bmatrix}$$

Solution

$$\lambda_1 = 3, \boldsymbol{v}_1 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad \lambda_2 = 2, \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \lambda_3 = 3, \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

Let $\boldsymbol{P}$ be the matrix with these eigenvectors as its columns:

$$\boldsymbol{P} = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & 0 \\ 2 & 1 & 2 \end{bmatrix}$$

Matrix Diagonalization

Then $\boldsymbol{P}$ diagonalizes $\boldsymbol{A}$:

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 0 & -1 & 0 \\ 2 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & -4 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & 0 \\ 2 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix Diagonalization

A matrix $\boldsymbol{P}$ is called orthogonal if

$$\boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}$$

If the transformation matrix $\boldsymbol{P}$ is an orthogonal matrix, then the linear transformation $\boldsymbol{Y} = \boldsymbol{P}^T\boldsymbol{X}$ sends every vector into a new vector space spanned by the orthogonal basis (i.e. the columns of $\boldsymbol{P}$).

$$\boldsymbol{Y} = \boldsymbol{P}^T\boldsymbol{X} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n]^T \boldsymbol{X}$$

$$= \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \vdots \\ \boldsymbol{p}_n \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} \boldsymbol{p}_1 \cdot \boldsymbol{X} \\ \boldsymbol{p}_2 \cdot \boldsymbol{X} \\ \vdots \\ \boldsymbol{p}_n \cdot \boldsymbol{X} \end{bmatrix}$$

Matrix Diagonalization

For any symmetric matrix $\boldsymbol{A}$, there exists an orthonormal matrix that can diagonalize $\boldsymbol{A}$. The columns of the orthonormal matrix consists of the eigenvectors of $\boldsymbol{A}$. More precisely:

A matrix is symmetric if and only if it has an orthonormal basis of eigenvectors.

The covariance matrix is symmetric!

Similar Matrices

Two $n \times n$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are similar if there exists matrix $\boldsymbol{P}$ such that

$$\boldsymbol{Y} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}$$

Similar matrices have the same

- rank
- determinant
- trace
- eigenvalue
- characteristic polynomial

Derivation of PCA

Let $\boldsymbol{X}$ be a n-dimensional random vector expressed as a column vector. Assume that $\boldsymbol{X}$ has zero empirical mean. We want to find a $n \times n$ orthonormal transformation matrix $\boldsymbol{P}$ such that

$$\boldsymbol{Y} = \boldsymbol{P}^T \boldsymbol{X}$$

such that cov$(\boldsymbol{Y})$ is a diagonal matrix.

$$\begin{aligned} cov(\boldsymbol{Y}) &= \mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^T\right] = \mathbb{E}\left[\left(\boldsymbol{P}^T\boldsymbol{X}\right)\left(\boldsymbol{P}^T\boldsymbol{X}\right)^T\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{P}^T\boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{P}\right)\right] \\ &= \boldsymbol{P}^T\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\boldsymbol{P} = \boldsymbol{P}^T cov\left(\boldsymbol{X}\right)\boldsymbol{P} \end{aligned}$$

Thus, we need to diagonalize $cov\left(\boldsymbol{X}\right)$.

Derivation of PCA

Since *cov* $(\boldsymbol{X})$ is symmetric, the transformation matrix $\boldsymbol{P}$ can be formed using its eigenvectors. Let

$$\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n]$$

where each column $\boldsymbol{p}_i$ is an eigenvectors. Then

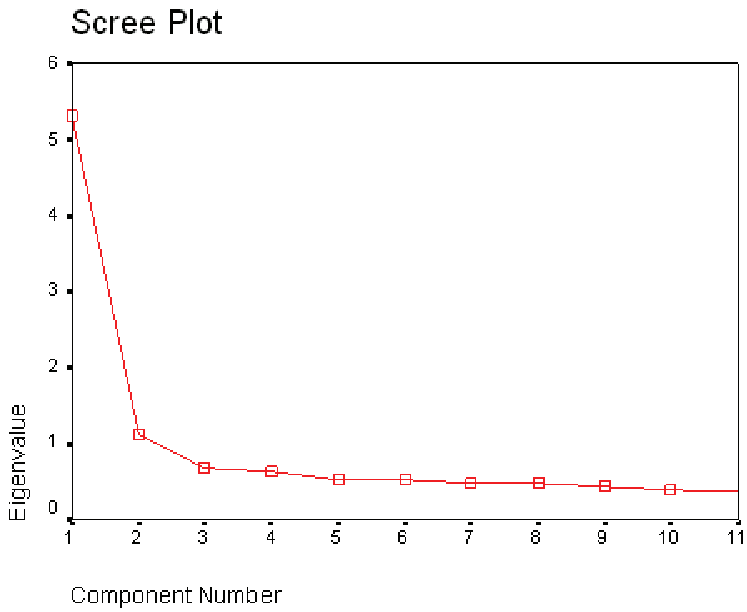
$$\boldsymbol{Y} = \boldsymbol{P}^T \boldsymbol{X} = \begin{bmatrix} \boldsymbol{p}_1 \cdot \boldsymbol{X} \\ \boldsymbol{p}_2 \cdot \boldsymbol{X} \\ \vdots \\ \boldsymbol{p}_n \cdot \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1^{new} \\ x_2^{new} \\ \vdots \\ x_n^{new} \end{bmatrix}$$

So, $\boldsymbol{X}$ is transformed into the same vector space. But the new coordinates $x_i^{new}, i = 1, \cdots, n$ are with respect to the eigenvectors.

PCA

- The eigenvectors are all orthogonal to each other and form an orthonormal basis (when they are normalized.)
- $[y_1, y_2, \cdots, y_n]$ are the projection of the original variables onto the eigenvectors.
- The order of the columns of $\boldsymbol{P}$ can be ordered in the descending order of the corresponding eigenvalues.
- The ordered eigenvectors are called the $1^{\text{st}}, 2^{\text{nd}}, \cdots, n^{\text{th}}$ principal components.

PCA



- The new variables have a variance equal to their eigenvalues.
- Small $\lambda_i \Leftrightarrow$ small variance $\Leftrightarrow$ data change little in the direction of $\boldsymbol{p}_i$.
- The relative variance that is explained by each principal component is $\lambda_i / \sum_i \lambda_i$.

PCA

If the eigenvectors that correspond to small eigenvalues are discarded, then we get the transformation of $\boldsymbol{X}$ in a vector space of reduced dimension without losing much information:

$$\boldsymbol{Y}_{reduced} = \boldsymbol{P}^T_{reduced}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{p}_1 \cdot \boldsymbol{X} \\ \boldsymbol{p}_2 \cdot \boldsymbol{X} \\ \vdots \\ \boldsymbol{p}_k \cdot \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} x_1^{new} \\ x_2^{new} \\ \vdots \\ x_k^{new} \end{bmatrix}$$

where $k < n$.

Example

PCA

- The new variables $[y_1, y_2, \cdots, y_n]$ are called the scores. Write $\boldsymbol{P}$ as

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

then

$$y_1 = \boldsymbol{p}_1 \cdot \boldsymbol{X} = p_{11}x_1 + p_{21}x_2 + \cdots + p_{n1}x_n$$

so, the new variables $y_i, i = 1, \cdots, n$ are weighted sum of the original variables. The transformation matrix provides the weights.

PCA

How to get the original data back?

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$$

For a reduced dimension, we get

$$\boldsymbol{X}_{reduced} = \boldsymbol{P}_{reduced}\boldsymbol{Y}_{reduced} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_k] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$$

PCA

PCA is very useful for

- finding more informative and uncorrelated features
- reducing dimensionality by rejecting low-variance features.

Limitations of PCA

- PCA is only powerful if the biological question is related to the highest variance in the dataset.
- If not, then other techniques are more useful, such as Independent Component Analysis (ICA)