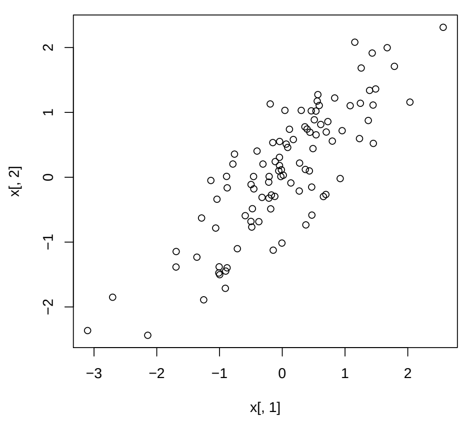
Principal Component Analysis

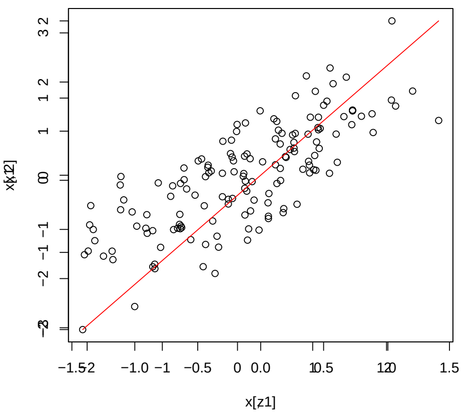
1 Introduction

Principal component analysis is a dimension reduction technique in multivariate analysis. The principal idea of reducing the dimension of $X = (X_1, X_2, \ldots, X_p)'$ is achieved through linear combination. Low dimensional linear combinations are easier to interpret and serve as an intermediate step as in a more complex analysis. More precisely one looks for linear combination which create largest spread among the values of X. In other words, one is searching for linear combinations of variables with largest variances.

Consider an the following example

Let $(X_1, X_2)' \sim N_2((0,0)', \mathbf{\Sigma})$ where $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$. The scatterplot of 100 sample drawn from this distribution is as follows





Let us consider the following transformation $\mathbf{X} \to \mathbf{Y}$

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 \\ Y_2 &= a_{21}X_1 + a_{22}X_2 \end{aligned}$$

or

$$\mathbf{Y} = \mathbf{AX}$$

The components of A are chosen such that $Var(Y_1)$ is maximized. The vectors $a_i' = (a_{i1}, a_{i2})'$ for $i = 1, 2$ are chosen such that $a_i a_j' = 0$ and $a_i a_i' = 1$, that is the row vectors of coefficient matrix are orthonormal.

Let

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2 \\ Y_2 &= \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2 \end{aligned}$$

Then

$$V(Y_1) = \frac{1}{2}\left(V(X_1) + V(X_2) + 2\text{Cov}(X_1, X_2)\right) = 1.8$$

and

$$V(Y_2) = \frac{1}{2}\left(V(X_1) + V(X_2) - 2\text{Cov}(X_1, X_2)\right) = 0.2.$$

So the linear combination Y_1 explains 90% of total variability $V(Y_1)+V(Y_2)$. Y_1 is first principal component. It can be verified that

$$V(X_1)+V(X_2) = V(Y_1)+V(Y_2)$$

General Model

$$\begin{aligned}
Y_1 &= a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p = \mathbf{a}_1'\mathbf{X} \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p = \mathbf{a}_2'\mathbf{X} \\
\vdots & \quad \vdots \\
Y_p &= a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p = \mathbf{a}_p'\mathbf{X} \\
\text{or} & \\
\mathbf{Y} &= \mathbf{AX}
\end{aligned}$$

$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)'$ are the principal components. Y_j is the j^{th} principal component.
The model is defined once the elements of A are estimated. Firstly $\mathbf{a}_1', \mathbf{a}_2' \ldots \mathbf{a}_p'$ needs to be orthogonal that is $\mathbf{a}_j'\mathbf{a}_k = 0$ and $\mathbf{a}_j'\mathbf{a}_j = 1$.
Choose $\mathbf{a}_1$ such that $Var(Y_1) = Var(\mathbf{a}_1'\mathbf{X})$ is maximized with respect to $\mathbf{a}_1$. $\mathbf{a}_2$ such that $Var(Y_2) = Var(\mathbf{a}_2'\mathbf{X})$ is maximized with respect to $\mathbf{a}_2$ subject to the orthogonality condition and so on. Here we have $\text{cov}(Y_j, Y_k) = 0$ and

$$Var(Y_1) \geq Var(Y_2) \geq \cdots \geq Var(Y_p)$$

Also,

$$\sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} Var(Y_i).$$

For the first $Y_1, Y_2, \ldots, Y_q, q < p$ we compute

$$\frac{\sum\limits_{i=1}^{q} Var(Y_i)}{\sum\limits_{i=1}^{p} Var(Y_i)} \times 100$$

If it is reasonably large we can consider these q variables instead of p original ones.

2 Population principal components

Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ be a p variate random vector with $E(\mathbf{X}) = \mu$ and known covariance matrix Σ. We shall consider cases which Σ is positive semidefinite matrix. Since we shall be concerned with variances and covariances of X we shall assume that $\mu = 0$. The first principal component the normalized linear combination $Y_1 = \alpha' X$where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ is such that $\alpha'\alpha = 1$ and

$$Var(\alpha'\mathbf{X}) = \max_{l} Var(l'\mathbf{X})$$

with $l' \in \mathbb{R}^p$ satisfying $L'L = 1$.
Now

$$V(l'X) = l'\Sigma l.$$

Thus to find first principal component $\alpha' X$ we need to the α such that maximizes $l'\Sigma L$ for all choices of $l \in \mathbb{R}^p$ subject to the restriction $l'l = 1$. Using the Lagrange's multiplier λ we need to find the α that maximizes

$$\phi_1(l) = l'\Sigma l - \lambda(l'l - 1)$$

for all choices of $l \in \mathbb{R}^p$ satisfying $l'l = 1$. Now

$$\begin{aligned} \frac{\partial \phi_1}{\partial l} &= 0 \\ \text{or} \quad 2\Sigma l - 2\lambda l &= 0 \\ \text{or} \quad (\Sigma - \lambda I)\, l &= 0 \end{aligned}$$

So, α satisfies the equation $(\Sigma - \lambda I)\,\alpha = 0 \ldots (1)$. Since $\alpha \neq 0$ as a consquence of $\alpha'\alpha = 1$, the equation (1) has a solution if $|\Sigma - \lambda I| = 0$.

That is λ is the characteristic root of Σ and α is the corresponding characteristic vector. Since dimension of Σ is $p \times p$ there are p values of λ. Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$$

denote the ordered characteristic roots and

$$\alpha_1 = (\alpha_{11}, \ldots, \alpha_{1p})' \ldots \alpha_p = (\alpha_{p1}, \ldots, \alpha_{pp})'$$

denote the characteristic vectors of Σ. Now, Σ may have zero characteristic root or some the roots may have multiplicity greater than unity.
Now, $(\Sigma - \lambda I)\,\alpha = 0$ gives

$$Var(\alpha' X) = \alpha' \Sigma \alpha = \lambda \alpha' \alpha = \lambda.$$

where λ is the characteristic root Σ corresponding to α. Thus to maximize $Var(\alpha' X)$ we need to choose $\lambda = \lambda_1$, the largest characteristic root of Σ and $\alpha = \alpha_1$ is the characteristic root corresponding to λ_1.

So the first principal component is the normalized linear function

$$Y_1 = \alpha_1' X = \sum_{i=1}^{p} \alpha_{1i} X_i$$

where α_1 is the normalized characteristic vector of Σ corresponding to largest characteristic root λ_1 is called first principal component of X.

The second principal component is the normalized linear function $\alpha' X$ having maximum variance among all normalized linear functions $l' X$ that are uncorrelated with Y_1.

$$\begin{aligned} \operatorname{cov}\left(l'X, Y_1\right) = E\left(l'XY_1\right) = E\left(l'Xl_1'\right) &= E\left(l'XX'\alpha_1\right) \\ &= l'\Sigma\alpha_1 = l'\lambda_1\alpha_1 = \lambda_1 l'\alpha_1 = 0 \end{aligned}$$

This implies l and α_1 are otrthogonal. The second principal component will be linear combination $\alpha' X$ that has maximum variance among all normalized linear combination $l'X, l \in \mathbb{R}^p$ which is uncorrelated with Y_1. Again by using Lagrange's multiplier we will get the second principal component as

$$Y_2 = \alpha_2' X$$

where α_2 is the characteristic vector corresponding to characteristic root λ_2.

Continuing in this way we will get the rth principal component as

$$Y_r = \alpha_r' X$$

where α_r is the characteristic vector corresponding to chracteristic root λ_r.

Now define matrices

$$A = (\alpha_1, \alpha_2, \ldots, \alpha_p), \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ are ordered characteristic roots and $\alpha_1, \ldots, \alpha_p$ are the corresponding normalized characteristic vectors. Since $AA' = I$ and $\Sigma A = A\Lambda$ we conclude that $A'\Sigma A = \Lambda$.

So there exists an orthogonal transformation

$$Y = A'X$$

such that $D(Z) = \Lambda$ a diagonal matrix with diagonal elements $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ the ordered roots of $|\Sigma - \lambda I| = 0$.

3 Sample principal component

In practice the covariance matrix is usually unknown. If sample obsrevations on a multivarite random vector is given we have to replace Σ by an estimate of covariance matrix Σ. Now we assume that $\mathbf{X} \sim N_p(\mu, \Sigma)$ where Σ is positive definite matrix.
Let $x^\alpha = (x_{\alpha 1}, x_{\alpha 2}, \ldots, x_{\alpha p})'$, $\alpha = 1, 2, \ldots, N, (N > p)$ be a sample of size N from the distribution of X which is univariate normal with mean vector μ and dispersion matrix Σ.

Let

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{\alpha=1}^{N} x^\alpha \qquad \mathbf{s} = \sum_{\alpha=1}^{N} (x^\alpha - \bar{x})(x^\alpha - \bar{x})'.$$

The maximumlikelihood estimate of Σ is $\frac{\mathbf{s}}{N}$ and that of μ is $\bar{\mathbf{x}}$.

Theorem: The maximum likelihood estimates of ordered characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_p$ of Σ and corresponding characteristic vector $\alpha_1, \alpha_2, \ldots, \alpha_p$ are , respectively the ordered characteristic roots $r_1, r_2, \ldots, r_p$ and characteristic vector $a_1, a_2, \ldots, a_p$ of $\mathbf{s}/N$.

Proof Omitted

The estimate of total system variance is given by

$$\text{trace}\left(\frac{\mathbf{s}}{N}\right) = \sum_{i=1}^{p} r_i$$

and is called total sample variance. The importance of i^{th} principal component is measured by

$$\frac{r_i}{\sum_{i=1}^{p} r_i}.$$

If the estimates of the principal components are obtained by from sample correlation matrix

$$R = (r_{ij}) \qquad r_{ij} = \frac{s_{ij}}{(s_{ii}s_{jj})^{1/2}}$$

with $s = (s_{ij})$ then the total sample variance will be $p = \text{trace}(R)$.

If first k principal components explain large amount of total sampel variance, they may be used in place of original vector $\mathbf{X}$.

Exercise

1. Let

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{pmatrix} \right)$$

Obtain the first principal component and obtain the percentage of variability explained by first principal component.

2. Let $(X_1, X_2, X_3)'$ is trivariate random vector with correlation matrix

$$\begin{pmatrix} 1 & 0.8944 & 0.7071 \\ 0.8944 & 1 & 0.6325 \\ 0.7071 & 0.6325 & 1 \end{pmatrix}$$

Find the first principal component and obtain the percentage of variability explained by first principal component.

3. Find principal components of the data "mtcars" in R. (Use "prcomp" command)