UNIT 04

Principal Components

A Principal Component analysis is Concerned with explaining the variance Co-variance structure of a set of variables through a few linear Combinations of these variables and its main objectives are as follows -

(a) Data reduction and

(b) Interpretation.

In practice the method of principal Components is used to find the linear Combinations with large variance. In many exploratory ~~st~~ Studies the number of variables under Consideration is too large to handle the data. Since it is a deviations in these Studies that are of interest, a way of reducing the number of variables to be treated is to discard the linear Combinations which have Small variances and study only those with large variances.

In multivariate analysis, the dimension of x often Causes the problem our aim is obtaining the Suitable Statistical technique to analyse the set of repeated observations on x. For this, it is natural to rearranging the data with a little loss of informations as possible then the dimension of a problem is ~~poss~~ possibly reduced, Such a technique is Called principal Componeor.

Principal Components in Population

Principal Components are particular linear combinations of the p-dimensional random variable x. Geometrically these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $x_1, x_2, \ldots x_p$ as the coordinate axes. The ~~new~~ new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

Let the random vector $x = (x_1\ x_2 \ldots x_p)'$ have the covariance matrix Σ with eigen values $\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \geq \lambda_p \geq 0$.

Consider the linear combinations

$$z_1 = \beta_1' x = \beta_{11} x_1 + \beta_{12} x_2 + \ldots + \beta_{1p} x_p$$

$$z_2 = \beta_2' x = \beta_{21} x_1 + \beta_{22} x_2 + \ldots + \beta_{2p} x_p$$

$$\vdots$$

$$z_p = \beta_p' x = \beta_{p1} x_1 + \beta_{p2} x_2 + \ldots + \beta_{pp} x_p$$

By the properties of linear transformation,

$$Var(z_i) = \beta_i' \Sigma \beta_i \quad ; \quad i = 1, 2, \ldots p$$

$$Cov(z_i, z_k) = \beta_i' \Sigma \beta_k \quad ; \quad k = 1, 2 \ldots p$$

The principal components are those uncorrelated linear combinations $z_1, z_2 \ldots z_p$ whose variances are large as possible.

First principal component is the linear combination $\beta_1' x$ that maximizes $Var(\beta_1' x)$ subject to $\beta_1' \beta_1 = 1$

Second principal component is the linear combination $\beta_2' x$ that maximizes $Var(\beta_2' x)$ subject to $\beta_2' \beta_2 = 1$

and $Cov(\beta_1' X, \beta_2' X) = 0$

$\vdots$

at the i^{th} principal Component is the linear Combination of $\beta_i' x$ that maximizes $Var(\beta_i' x)$ Subject to $\beta_i' \beta_i = 1$

and $Cov(\beta_i' x, \beta_k' x) = 0$ for $k < i$

—a—

<u>First Principal Component</u>

Let $X = (x_1\ x_2 \cdots x_p)'$ be a 'p' variate random vector with $E(X) = \mu$ and known Covariance matrix Σ. We shall consider only variance, covariance of X so that we shall assume $\mu = 0$.

Let β be the p ~~variate~~ Component Column vector such that $\beta'\beta = 1$ then the first principal Component is an normalised linear Combination say $z_1 = \beta' x$,

where $\beta = (\beta_1\ \beta_2 \cdots \beta_p)'$

$$Var(\beta' x) = \cancel{E(\beta' x)}\ E(\beta' x)^2 - [E(\beta' X)]^2$$

$$= E[(\beta' x)(\beta' x)'] - [\beta' E(x)]^2 \quad \{\because E(x) = \mu = 0$$

$$= E[\beta' x x' \beta] - 0$$

$$= \beta' E(x x') \beta$$

$$Var(\beta' x) = \beta' \Sigma \beta \longrightarrow (1)$$

Thus we find the first Principal Component which maximizes equation (1) subject to the Condition that $\beta'\beta = 1$.

using, Lagrange's multiplier, let us define

$$\phi_1 = \beta' \Sigma \beta - \lambda(\beta'\beta - 1) \longrightarrow (2)$$

Scanned by CamScanner

differentiate w.r.t β and equating to zero

we get,

$$\frac{\partial \phi_1}{\partial \beta} = 2\Sigma\beta - 2\lambda\beta = 0$$

$$\Rightarrow (\Sigma - \lambda)\beta = 0 \longrightarrow (3)$$

Since $\beta \neq 0$ and $(\Sigma - \lambda)$ is singular

$\therefore$ λ must satisfy $|\Sigma - \lambda I| = 0 \longrightarrow (4)$

ie λ is the characteristic root of Σ and β is the corresponding characteristic vector.

Since Σ has p-dimension, the function $|\Sigma - \lambda I|$ is a polinomial of degree p. The ordered characteristic roots of Σ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$

Let $\beta_1 = (\beta_{11} \ \beta_{12} \ \cdots \ \beta_{1p})'$

$\beta_2 = (\beta_{21} \ \beta_{22} \quad \beta_{2p})'$

$\vdots$

$\beta_p = (\beta_{p1} \ \beta_{p2} \quad \beta_{pp})'$ denotes the corresponding characteristic vector of Σ.

Pre multiplying both sides of (3) by β'

$$(3) \Rightarrow \beta'(\Sigma - \lambda)\beta = 0$$

$$\beta'\Sigma\beta - \beta'\lambda\beta = 0$$

$$\beta'\Sigma\beta = \beta'\lambda\beta$$

$$= \lambda\beta'\beta$$

$$\beta'\Sigma\beta = \lambda \qquad \{\because \beta'\beta = 1$$

Thus $Var(\beta' x) = \lambda \longrightarrow (5)$

Thus for maximum variance we should use the equation 3, the largest root is λ_1.

Let $\beta_{(1)}$ be a normalised solution of ~~$|\Sigma - \lambda I| \beta = 0$~~

$|\Sigma - \lambda I| \beta = 0$ then $z_1 = \beta_{(1)}' X$ is an normalised linear combination with maximum variance, which is the first Principal Component.

Second Principal Component

For the second principal component, let us find the normalising linear combination $\beta' X$ which has the maximum variance of all linear combination uncorrelated with z_1.

$$E[(\beta' x) z_1] = 0 = E[(\beta' x) z_1']$$

$$E[(\beta' x) z_1'] = 0$$

&

$$E[(\beta' x)(\beta_{(1)}' x)'] = 0 \qquad \{\because z_1 = \beta_{(1)}' x$$

$$E[(\beta' x)(x' \beta_{(1)})] = 0$$

$$E[\beta' (x x') \beta_{(1)}] = 0$$

$$\beta' E(x x') \beta_{(1)} = 0$$

$$\beta' \Sigma \beta_{(1)} = 0$$

$$\beta' \lambda_1 \beta_{(1)} = 0 \qquad \{\because \text{The } \overset{\text{highest}}{\cancel{\text{maximum}}} \text{ root of } \Sigma \text{ is } \lambda_1$$

$$\lambda_1 \beta' \beta_{(1)} = 0 \quad \cancel{(4)}$$

$$\boxed{\beta' \beta_{(1)} = 0} \longrightarrow (1) \qquad \{\because \lambda_1 \neq 0$$

$$(\Sigma - \lambda I) \beta_{(1)} = 0 \qquad \{\because \text{By first principal Component}$$

$$\Sigma \beta_{(1)} - \lambda I \beta_{(1)} = 0$$

$$\boxed{\Sigma \beta_{(1)} = \lambda \beta_{(1)}} \longrightarrow (2) \qquad \text{where } I \text{ is the unit matrix}$$

By Lagrange's multiplier, consider the transformation

$$\phi_2 = \beta'\Sigma\beta - \lambda(\beta'\beta - 1) - 2v_1(\beta'\Sigma\beta_{(1)}) \longrightarrow ③$$

where λ and v_1 are Lagrange's multipliers

Now, ϕ

$$\frac{\partial\phi_2}{\partial\beta} = 2\Sigma\beta - \lambda(2\beta - 0) - 2v_1(\Sigma\beta_{(1)})$$

Put $\frac{\partial\phi_2}{\partial\beta} = 0$ we get

$$2\Sigma\beta - 2\lambda\beta - 2v_1\Sigma\beta_{(1)} = 0$$

Pre Multiplying both sides by $\beta'_{(1)}$ we get

$$2\beta'_{(1)}\Sigma\beta - 2\beta'_{(1)}\lambda\beta - 2v_1\beta'_{(1)}\Sigma\beta_{(1)} = 0$$

$$2\cancel{\beta'_{(1)}\lambda\beta} - 2\cancel{\beta'_{(1)}\lambda\beta} - 2v_1\beta'_{(1)}\Sigma\beta_{(1)} = 0 \qquad \{\because \text{from } ②$$

$$\boxed{-2v_1\beta'_{(1)}\Sigma\beta_{(1)} = 0} \longrightarrow ④$$

From the first Principal Component,

we have, $(\Sigma - \lambda_1)\beta_{(1)} = 0$

Premultiplying both sides by $\beta'_{(1)}$

$$\beta'_{(1)}(\Sigma - \lambda_1)\beta_{(1)} = 0$$

$$\beta'_{(1)}\Sigma\beta_{(1)} - \lambda_1\beta'_{(1)}\beta_{(1)} = 0 \qquad \{\because \beta_{(1)}\beta'_{(1)} = 1$$

$$\beta'_{(1)}\Sigma\beta_{(1)} - \lambda_1 = 0$$

$$\boxed{\beta'_{(1)}\Sigma\beta_{(1)} = \lambda_1} \longrightarrow ⑤$$

Substitute ⑤ in ④ we get,

$$-2v_1\lambda_1 = 0$$

Scanned by CamScanner

Since $\lambda_1 \neq 0$,

$\therefore$ $v_1 = 0$ and β must satisfy the condition

$$(\Sigma - \lambda I)\beta = 0$$

$\therefore$ λ must satisfy $|\Sigma - \lambda I| = 0$

Let λ_2 be the maximum root such that the vector β satisfying $(\Sigma - \lambda_2 I)\beta = 0$ and corresponding linear combination $Z_2 = \beta_{(2)}' X$, which is called second principal component.

Maximum likelihood estimator of principal components and their variances

The statistical inference in principal component analysis is to estimate the vectors $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(p)}$ and the scalars $\lambda_1 \lambda_2 \dots \lambda_p$.

Theorem Let $x_1, x_2 \dots x_N$ be N observations $(N > p)$ from the $N(\mu, \Sigma)$ where Σ is a matrix with p different characteristic roots. Then a set of maximum likelihood estimators of $\lambda_1 \lambda_2 \dots \lambda_p$ and $\beta^{(1)}, \beta^{(2)} \dots \beta^{(p)}$ consists of the roots $k_1 \geq k_2 \geq \dots \geq k_p$ of $|\hat{\Sigma} - kI| = 0 \rightarrow (1)$ and a set of corresponding vectors $b^{(1)}, b^{(2)}, \dots, b^{(p)}$ satisfying

$$(\hat{\Sigma} - k_i I) b^{(i)} = 0 \longrightarrow (2)$$

$$b^{(i)\prime} b^{(i)} = 1 \longrightarrow (3)$$

where $\hat{\Sigma}$ is the M.L.E of Σ.

Proof

When the roots of $|\Sigma - \lambda I| = 0$ are different, each vector $\beta^{(i)}$ is uniquely defined except that $\beta^{(i)}$ can be replaced by $-\beta^{(i)}$. If we require that the first nonzero component of $\beta^{(i)}$ be positive.

then $\beta^{(i)}$ is uniquely defined and μ, Λ, β is a single valued functions of μ, Σ. The set of MLE of μ, Λ, β is the same function of $\hat{\mu}, \hat{\Sigma}$.

Since the function $|\Sigma - \lambda I|$ is a polynomial in λ of ~~deg~~ degree p.

$$\beta' \Sigma \beta = \Lambda$$

$$\Sigma \beta = \beta \Lambda$$

$$\Sigma = \beta \Lambda \beta' \qquad \{\because |\Sigma| \neq 0$$

$$\Sigma = \sum \lambda_i \beta^{(i)} \beta^{(i)'}$$

Similarly $\hat{\Sigma} = \sum k_i \, b^{(i)} \, b^{(i)'}$.

Replacing $b^{(i)}$ by $-b^{(i)}$ clearly doesnot change $\sum k_i \, b^{(i)} \, b^{(i)'}$. Since the likelihood function depends only on $\hat{\Sigma}$. Since the likelihood function depends only on $\hat{\Sigma}$, the maximum of the likelihood function is attained by taking any set of solutions to (2) and (3).

Suppose that we assume $|\Sigma - \lambda I| = 0$ has one root of multiplicity p. Let it be λ_1. By the theorem $\Sigma - \lambda_1 I$. is of rank zero.

ie., $\Sigma - \lambda_1 I = 0$

$$\Sigma = \lambda_1 I$$

If X is distributed according to $N(\mu, \Sigma)$

ie $X \sim N(\mu, \lambda_1 I)$ the components of X are independently ~~distributed~~ distributed with variance λ_1.

Thus the MLE of λ_1 is $\hat{\lambda}_1$

$$\hat{\lambda}_1 = \frac{1}{pN} \sum_{i=1}^{p} \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)^2 \quad \text{and} \quad \hat{\Sigma} = \hat{\lambda}_1 I$$

and β can be any ~~other~~ orthogonal matrix.

Canonical Correlation and Canonical Variables

We consider two sets of variates with a joint distribution, and we analyse the correlations between the variables of one set and those of the other set. We find a new coordinate system in the space of each set of variates in such a way that the new coordinates display unambiguously the system of correlation. Clearly, we find linear combinations of variables in the sets that have maximum correlation; these linear combinations are the first coordinates in the new system. Then a second linear combination in each set is sought such that the correlation between these is the maximum of correlations between such linear combinations as are uncorrelated with the first linear combinations. This method is continued until the two new coordinate systems are completely specified.

The investigator may have two large sets of variates and may wish to study the interrelations. If the two sets are very large, he may wish to consider only a few linear combinations of each it. One set of variables may be measurements of physical characteristics, such as various lengths ~~of~~ and breadths of skulls; the other variables may be measurements of mental characteristics, such as scores on intelligence test. If the investigator is interested on these, he may find that the interrelation is almost completely described by the correlation between first ~~first~~ few canonical variates.

Canonical Correlation and variates in the Population

Let X be a random vector with p Components $x_1, x_2 \cdots x_p$ and has the Covariance matrix Σ. Assume $E(x) = 0$ when treating the population. we partition X into ~~to~~ two Sub vectors with p_1 and p_2 Components,, $X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$.

we shall assume $p_1 \leq p_2$. The Covariance matrix is partitioned Similarly into p_1 and p_2 rows and Columns.

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$