Statistical Inference and Multivariate Analysis (MA324)

LECTURE SLIDES
Lecture 33

Principal Component Analysis

Indian Institute of Technology Guwahati

Jan-May 2023

Principal Component Analysis

- Concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables.
- Objectives
  - Data Reduction.
  - Interpretation.
- Serves as intermediate steps in a larger study (Multiple Linear Regression, Cluster Analysis).
- Wide application areas: including image analysis, finance, health sector, cryptography, data-privacy etc.

Population Pricipal Components

- Let $X_1, X_2, ..., X_p$ be p-random variables with variance-covariance matrix Σ.

- First we want to find a linear combination of $\underset{\sim}{X} = (X_1, X_2, ..., X_p)$, say $\underset{\sim}{a}_1' \underset{\sim}{X}$, such that $Var(\underset{\sim}{a}_1' \underset{\sim}{X})$ is maximum.

- Note that $Var(\underset{\sim}{a}_1' \underset{\sim}{X}) = \underset{\sim}{a}_1' \Sigma \underset{\sim}{a}_1$. Now by multiplying $\underset{\sim}{a}_1$ by a constant, we can increase $\underset{\sim}{a}_1' \Sigma \underset{\sim}{a}_1$ arbitrarily.

- Therefore, a more precise aim is:

$$\max_{\underset{\sim}{a}_1} Var(\underset{\sim}{a}_1' \underset{\sim}{X}) = \max_{\underset{\sim}{a}_1} \underset{\sim}{a}_1' \Sigma \underset{\sim}{a}_1,$$

subject to $\underset{\sim}{a}_1' \underset{\sim}{a}_1 = 1$

- Thus, 1^{st} principal component = linear combination $\underset{\sim}{a}_1'\underset{\sim}{X}$ which maximizes $Var(\underset{\sim}{a}_1'\underset{\sim}{X})$ subject to $\underset{\sim}{a}_1'\underset{\sim}{a}_1 = 1$
- Next, we want to find another linear cobination, say $\underset{\sim}{a}_2'\underset{\sim}{X}$, such that $Var(\underset{\sim}{a}_2'\underset{\sim}{X})$ is maximum subject to $\underset{\sim}{a}_2'\underset{\sim}{a}_2 = 1$ and $Cov(\underset{\sim}{a}_1'\underset{\sim}{X}, \underset{\sim}{a}_2'\underset{\sim}{X}) = 0$.
- We proceed in the following manner
    - i^{th} principle component = linear combination $\underset{\sim}{a}_i'\underset{\sim}{X}$ that maximizes $Var(\underset{\sim}{a}_i'\underset{\sim}{X})$ subject to $\underset{\sim}{a}_i'\underset{\sim}{a}_i = 1$ and $Cov(\underset{\sim}{a}_k'\underset{\sim}{X}, \underset{\sim}{a}_i'\underset{\sim}{X}) = 0$ for $k < i$.
- Notice that $Cov(\underset{\sim}{a}_k'\underset{\sim}{X}, \underset{\sim}{a}_i'\underset{\sim}{X}) = \underset{\sim}{a}_i'\Sigma\underset{\sim}{a}_k$, and $Var(\underset{\sim}{a}_i'\underset{\sim}{X}) = \underset{\sim}{a}_i'\Sigma\underset{\sim}{a}_i$.

Theorem: Maximization of Quadratic Forms

Let $B_{p\times p}$ be a symmetric non-negative definite matrix with eigen-values $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$ and associated normalized eigen-vectors $\underset{\sim}{e_1}, \underset{\sim}{e_2}, ..., \underset{\sim}{e_p}$.

Then,

$$\max_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}' B \underset{\sim}{x}}{\underset{\sim}{x}' \underset{\sim}{x}} = \lambda_1 \left(= \frac{\underset{\sim}{e_1}' B \underset{\sim}{e_1}}{\underset{\sim}{e_1}' \underset{\sim}{e_1}} \right) \; attained \; at \; \underset{\sim}{x} = \underset{\sim}{e_1},$$

$$\min_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}' B \underset{\sim}{x}}{\underset{\sim}{x}' \underset{\sim}{x}} = \lambda_p \; attained \; at \; \underset{\sim}{x} = \underset{\sim}{e_p}$$

Moreover,

$$\max_{\underset{\sim}{x} \neq \underset{\sim}{0}, \underset{\sim}{x} \perp \underset{\sim}{e_1}, ..., \underset{\sim}{e_{k-1}}} \frac{\underset{\sim}{x}' B \underset{\sim}{x}}{\underset{\sim}{x} \underset{\sim}{x}'} = \lambda_k \; attained \; at \; \underset{\sim}{x} = \underset{\sim}{e_k}, k = 1, 2, ..., p$$

Theorem

Let $\Sigma_{p\times p}$ be the variance-covariance matrix of $\underset{\sim}{X}_{p\times 1} = (X_1, ..., X_p)$. Let Σ have the eigen-value-eigen-vector pair $(\lambda_1, e_1), (\lambda_2, e_2), ..., (\lambda_p, e_p)$ where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Then the i^{th} principal component is given by,

$$Y_i = \underset{\sim}{e_i}' \underset{\sim}{X}, \quad i = 1, 2, ..., p.$$

With these choices $Var(Y_i) = \lambda_i$ and $Cov(Y_i, Y_j) = 0 \; for \; i \neq j$. Moreover, if some λ_i are equal, the choices of e_i are not unique, and hence Y_i are not unique.

Theorem

$$\sum_{i=1}^{p} Var(X_i) = \sigma_{11} + \sigma_{22} + ... + \sigma_{pp} = tr(\Sigma) = \lambda_1 + \lambda_2 + ... + \lambda_p = \sum_{i=1}^{p} Var(Y_i).$$

- Therefore, the proportion of total variance explained by i^{th} principal component is $\frac{\lambda_i}{\lambda_1 + ... + \lambda_p}$, $i = 1, 2, ..., p$.
- e_{ik} measures the importance of X_k to the i^{th} principal component.
- Particularly, e_{ik} is proportional to the correlation coefficient between Y_i and X_k.

Theorem

If $Y_1 = \underset{\sim}{e_1}' \underset{\sim}{X}, Y_2 = \underset{\sim}{e_2}' \underset{\sim}{X}, \cdots, Y_p = \underset{\sim}{e_p}' \underset{\sim}{X}$ are the principal components obtained from the covariance matrix Σ, and ρ_{Y_i, X_k} denotes the correlation coefficient between the Y_i and X_k, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}},$$

where $(\lambda_i, e_i), i = 1, \cdots, p$ are the eigenvalue-eigenvector pairs for Σ.