Multivariate Statistics

Sudipta Das

Assistant Professor,
Department of Data Science,
Ramakrishna Mission Vivekananda University, Kolkata
Slides adapted from Jhonson & Winchern

Outline I

Introduction I

- A principal component analysis (PCA) is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables.
- Objectives
    - Data reduction
    - Data interpretation

Introduction II

- By PCA we select k principal components from a set for $p(\geq k)$ initial variables such that the total system variability is retained as much as possible.
- Data-set of size $(n \times p) \xRightarrow{PCA}$ Data-set of size $(n \times k)$
- Note
  - To retain the total system variability, we need to retain all the p principal components

Population Principal Components I

- Principal components are particular linear combinations of the p random features/variables $X_1, X_2, \ldots, X_p$.
  - These linear combinations represents selection of new coordinate system obtained by rotating the original system with $X_1, X_2, \ldots, X_p$ as the coordinate axes
    - the new axes represent the direction with maximum variability and
    - provide a simpler and more parsimonious description of the covariance structure
- Note:
  - Principal components depends solely on the covariance matrix Σ of $X_1, X_2, \ldots, X_p$.
  - Their development does not require a multivariate normal assumption.
  - However, standard results on inference can be used if the samples are assumed to be coming from normal population

Population Principal Components II

- Formal definition
  - First principal component
    - is the linear combination of $\mathbf{a_1}'\mathbf{X}$ that maximizes $Var(\mathbf{a_1}'\mathbf{X})$ subject to $\mathbf{a_1}'\mathbf{a_1} = 1$
  - Second principal component
    - is the linear combination of $\mathbf{a_2}'\mathbf{X}$ that maximizes $Var(\mathbf{a_2}'\mathbf{X})$ subject to $\mathbf{a_2}'\mathbf{a_2} = 1$ and $Cov(\mathbf{a_2}'\mathbf{X}, \mathbf{a_1}'\mathbf{X}) = 0$
  - $\cdots\cdots$
  - *i*th principal component
    - is the linear combination of $\mathbf{a_i}'\mathbf{X}$ that maximizes $Var(\mathbf{a_i}'\mathbf{X})$ subject to $\mathbf{a_i}'\mathbf{a_i} = 1$ and $Cov(\mathbf{a_i}'\mathbf{X}, \mathbf{a_k}'\mathbf{X}) = 0$ for all $k < i$

Population Principal Components III

- Result: Let Σ be the covariance matrix associated with random vector $\mathbf{X} = [X_1, X_2, \dots, X_p]'$. Let Σ have the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}), \dots, (\lambda_p, \mathbf{e_p})$, where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. Then the *i*th principal component is given by

$$Y_i = \mathbf{e_i}'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p, \text{ for } i = 1, \dots, p$$

  - With these choices,

$$Var(Y_i) = \mathbf{e_i}'\Sigma\mathbf{e_i} = \lambda_i, \text{ for } i = 1, \dots, p$$

$$Cov(Y_i, Y_k) = \mathbf{e_i}'\Sigma\mathbf{e_k} = 0, \text{ for } i \neq k$$

  - Note: If some λ_i are equal then the choices of the corresponding coefficient vectors $\mathbf{e}_i$, and hence Y_i, are not unique.

Population Principal Components IV

- Sketch of proof:
  To get the first principal component, we need

$$\max_{\mathbf{a}}(Var(\mathbf{a}'\mathbf{X})) \text{ s. t. } \mathbf{a}'\mathbf{a} = 1 \Rightarrow \max_{\mathbf{a}} \left(\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}} \right)$$

  Thus (*Lemma: Maximization of Quadratic Forms for Points on the Unit Sphere*),

$$\max_{\mathbf{a}} \left(\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}} \right) = \lambda_1$$

  and maximum is attained at $\mathbf{a} = \mathbf{e_1}$.
  Hence, $Y_1 = \mathbf{e}_1'\mathbf{X}$ and $Var(Y_1) = \mathbf{e}_1'\Sigma\mathbf{e_1} = \lambda_1$

Population Principal Components V

- Sketch of proof (contd.):
  To get the *i*th principal component, we need

$$\max_{\mathbf{a}}(Var(\mathbf{a}'\mathbf{X})) \text{ s. t. } \mathbf{a}'\mathbf{a} = 1 \text{ and } Cov(\mathbf{a}'\mathbf{X}, \mathbf{a_k}'\mathbf{X}) = 0 \text{ for all } k < i$$

$$\Rightarrow \quad \max_{\mathbf{a}}\left(\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}}\right) \text{ s. t. } Cov(\mathbf{a}'\mathbf{X}, \mathbf{e_k}'\mathbf{X}) = 0 \text{ for all } k < i$$

$$\Rightarrow \quad \max_{\mathbf{a}\perp\mathbf{e_1},\ldots,\mathbf{e_{i-1}}}\left(\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}}\right), [\text{ since } \mathbf{a}'\Sigma\mathbf{e_k} = \mathbf{a}'\lambda_k\mathbf{e_k} = 0 \Rightarrow \mathbf{a}\perp\mathbf{e_k}]$$

Thus (*Lemma: Maximization of Quadratic Forms for Points on the Unit Sphere*),

$$\max_{\mathbf{a}\perp\mathbf{e_1},\ldots,\mathbf{e_{i-1}}}\left(\frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}}\right) = \lambda_i$$

and maximum is attained at $\mathbf{a} = \mathbf{e_i}$.
Hence, $Y_i = \mathbf{e_i}'\mathbf{X}$ and $Var(Y_i) = \mathbf{e_i}'\Sigma\mathbf{e_i} = \lambda_i$
Also,

$$Cov(Y_i, Y_k) = Cov(\mathbf{e_i}'\mathbf{X}, \mathbf{e_k}'\mathbf{X}) = \mathbf{e_i}'\Sigma\mathbf{e_k} = 0$$

Population Principal Components VI

- Result: Let the random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]'$ have covariance matrix Σ, with the eigenvalue- eigenvector pairs $(\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}), \ldots, (\lambda_p, \mathbf{e_p})$, where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. Let $Y_1 = \mathbf{e_1}'\mathbf{X}$, $Y_2 = \mathbf{e_2}'\mathbf{X}, \ldots, Y_p = \mathbf{e_p}'\mathbf{X}$ be the principal components. Then

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

- Proportion of total population variance explained by kth principal component:
$$\frac{\lambda_k}{\lambda_1 + \ldots + \lambda_k + \ldots + \lambda_p}$$
- Proportion of total population variance explained by first k principal components:
$$\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_k + \ldots + \lambda_p}$$

Population Principal Components VIII

- Result: If $Y_1 = \mathbf{e_1}'\mathbf{X}$, $Y_2 = \mathbf{e_2}'\mathbf{X}, \ldots, Y_p = \mathbf{e_p}'\mathbf{X}$ are the principal components obtained from the covariance matrix Σ, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \text{ for } i, k = 1, 2, \ldots, p$$

  are the correlation coefficients between the components Y_i and the variables X_k. Here $(\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}), \ldots, (\lambda_p, \mathbf{e_p})$ are the eigenvalue- eigenvector pairs for Σ.
  - The magnitude of e_{ik} measures the importance of kth variable (X_k) to the ith principal component (Y_i)

Population Principal Components IX

- Sketch of proof:

$$
\begin{aligned}
\rho_{Y_i,X_k} &= Cor(Y_i, X_k) \\
&= \frac{Cov(e_i'\mathbf{X}, [0\ 0 \ldots \underset{k^{th}}{1} \ldots 0]\mathbf{X})}{\sqrt{Var(Y_i)Var(X_k)}} \\
&= \frac{[0\ 0 \ldots \underset{k^{th}}{1} \ldots 0]\Sigma e_i}{\sqrt{\lambda_i \sigma_{kk}}} \\
&= \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i \sigma_{kk}}} = \frac{\sqrt{\lambda_i} e_{ik}}{\sqrt{\sigma_{kk}}}
\end{aligned}
$$

Principal Components on Standardized Variable I

- Given the vector X, the standardized vector can be obtained as

$$\mathbf{Z} = V^{-\frac{1}{2}}(\mathbf{X} - \mu),$$

recall $v = \begin{bmatrix} \sigma_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{pp} \end{bmatrix}$.

- Note:
  - $E(\mathbf{Z}) = \mathbf{0} = [0 \dots 0]'$
  - $Cov(\mathbf{Z}) = \rho = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots & \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{bmatrix}$.

Principal Components on Standardized Variable II

- Result: The ith principal component of the standardized variables $\mathbf{Z} = [Z_1\ Z_2 \dots Z_p]'$ with $Cov(\mathbf{Z}) = \rho$, is given by

$$Y_i = e_i'\mathbf{Z}, \text{ for } i = 1, 2, \dots, p.$$

Moreover,

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

In this case, $(\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}), \dots, (\lambda_p, \mathbf{e_p})$ are the eigenvalue-eigenvector pairs for ρ, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

Principal Components on Standardized Variable III

- Proportion of total population variance explained by kth principal component:

$$\frac{\lambda_k}{p}$$

- Proportion of total population variance explained by first k principal components:

$$\frac{\lambda_1 + \ldots + \lambda_k}{p}$$

Summarizing Sample Variations by Principal Components I

- Result: Let $\mathbf{X}$ be the observation on the variables $X_1, X_2, \ldots, X_p$ with the corresponding sample covariance matrix $S_{p\times p}$. Then the ith sample principal component is given by

$$\hat{\mathbf{Y}}_i = \hat{\mathbf{e}}_i'\mathbf{X} = \hat{e}_{i1}X_1 + \cdots + \hat{e}_{ip}X_p \text{ for } i = 1, 2 \ldots, p,$$

where $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ are the eigenvalue-eigenvector pairs for S, with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. Also,

$$Var(\hat{Y}_i) = \hat{\lambda}_i, \text{ for } i = 1, 2, \ldots, p$$

and

$$Cov(\hat{Y}_i, \hat{Y}_k) = 0, \text{ for } i \neq k.$$

In addition,

$$\text{Total Sample Variance } = \sum_{i=1}^{p} s_{ii} = \sum_{i=1}^{p} \hat{\lambda}_i$$

Summarizing Sample Variations by Principal Components II

- Result: Let $\mathbf{Z}$ be the observation on the variables $Z_i \left(= \frac{X_i - \bar{X}_i}{\sqrt{s_{ii}}}\right)$s, $i = 1, \ldots, p$, with the corresponding sample covariance matrix $R_{p \times p}$. Then the ith sample principal component is given by

$$\hat{\mathbf{Y}}_i = \hat{\mathbf{e}}_i' \mathbf{Z} = \hat{e}_{i1} Z_1 + \cdots + \hat{e}_{ip} Z_p \text{ for } i = 1, 2 \ldots, p,$$

where $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ are the eigenvalue-eigenvector pairs for R, with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. Also,

$$Var(\hat{Y}_i) = \hat{\lambda}_i, \text{ for } i = 1, 2, \ldots, p$$

and

$$Cov(\hat{Y}_i, \hat{Y}_k) = 0, \text{ for } i \neq k.$$

In addition,

$$\text{Total Sample Variance } = \sum_{i=1}^{p} \hat{\lambda}_i = p$$

Summarizing Sample Variations by Principal Components III

- How many principal components to be retained?
- No definite answer.
- Subjectively, we deice on
  - the relative size of the eigenvalues and the amount of sample variation explained
  - subject-matter interpretations of the components is also important
- Visual aid: Scree Plot
  - Plot of $\hat{\lambda}_i$ vs i
  - To determine the appropriate number of components we look for an elbow (bend) in the scree plot
    - The number of components is taken to be the point at which the remaining eigenvalues are relatively small and all about the same size.