CdL in Scienze Statistiche ed Economiche - Università degli Studi di Milano-Bicocca

Esercitazione : Analisi delle Componenti Principali

Esercitatrice: Chiara Gaia Magnani

Example 0.1. *1. Determinare le componenti principali relative alla matrice di varianze e covarianze*

$$S = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

2. Si calcoli la proporzione di varianza spiegata dalla prima componente principale.

Example 0.2. *Partendo dalla matrice di varianze e covarianze dell'esercizio precedente, calcolare nella matrice di correlazione R.*
1. Determinare le componenti principali a partire da R e la proporzione di varianza spiegata dalla prima componente.
2. Confrontare i risultati con quelli ottenuti nell'esercizio precedente e commentare la risposta.
3. Si calcoli la correlazione tra il k-simo vettore dei punteggi $\underset{n\times 1}{y_k}$, per $k = 1, 2$, e la j-sima colonna di Z, i.e. $\underset{n\times 1}{z_j}$, per $j = 1, 2$.

Example 0.3. *Si supponga che la matrice dei dati X consista di due colonne $\underset{n\times 1}{x_1}$ e $\underset{n\times 1}{x_2}$ tali che $x_2 = 2x_1$. Determinare autovalori e autovettori della matrice di correlazione R di X. Qual è la percentuale di varianza spiegata dalla prima componente principale?*

Example 0.4. *Supponiamo che alla matrice dei dati $\underset{n\times p}{X}$ sia associata la varianze/covarianze $S = diag(s_{11}, \ldots, s_{pp})$ con $s_{11} \geq \ldots \geq s_{pp} > 0$. Per questa particolare matrice di varianze/covarianze, ha senso effettuare l'analisi delle componenti principali? E l'analisi delle componenti principali basata sulla corrispondente matrice di correlazione R? Giustificare le risposte.*

Example 0.5. *Supponiamo di aver ortogonalizzato i dati attraverso la trasformazione di Mahalanobis, ottenendo così la matrice $\tilde{Z}$. Ha senso considerare l'analisi delle componenti principali sui dati ortogonalizzati $\tilde{Z}$? Giustificare la risposta.*

Example 0.6. *Si consideri la seguente matrice dei dati*

$$X = [\underset{5\times 1}{x_1} \;\; \underset{5\times 1}{x_2}] = \begin{bmatrix} -3 & 1 \\ 1 & -0.5 \\ 0 & -1 \\ -1 & -0.5 \\ 3 & 1 \end{bmatrix}$$

Calcolare il vettore dei punteggi $\underset{5\times 1}{y_1}$ *relativo alla prima componente principale basata sulla matrice di varianza/covarianza di* X *e la corrispondente proporzione di varianza spiegata.*

Example 0.7. *Alla matrice dei dati standardizzati* $\underset{n\times p}{Z}$ *+ associata la seguente matrice di correlazione:*

$$\underset{p\times p}{R} = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \cdots & \cdots & \cdots & \cdots \\ r & r & \cdots & 1 \end{bmatrix}$$

Questa matrice di correlazione descrive p *variabili ugualmente correlate.*
Per $r > 0$, *gli autovalori-autovettori di* R *risultano*

$$\lambda_1 = 1 + (p-1)r, \quad v_1 = (1/\sqrt{p}, \ldots, 1/\sqrt{p})'$$

e

$$\lambda_j = 1 - r, \quad v_j = (\underbrace{\frac{1}{\sqrt{(j-1)j}}, \ldots, \frac{1}{\sqrt{(j-1)j}}}_{j-1}, \frac{-(j-1)}{\sqrt{(j-1)j}}, \underbrace{0, \ldots, 0}_{p-j})', \quad j = 2, \ldots, p$$

Scrivere l'equazione della prima componente principale di Z. *i.e.* $\underset{n\times 1}{y_1} = Zv_1$ *e calcolare la proporzione di varianza spiegata. Per quali valori di* r *e* p *la proporzione di varianza spiegata dalla prima componente principale risulta elevata? Fornire un esempio numerico.*

Example 0.8. *Si consideri l'analisi delle componenti principali sui dati standardizzati* $\underset{n\times 2}{Z}$ *con la seguente matrice di correlazione*

$$R = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

dove $0 < r < 1$. *Si consideri ora una trasformazione di scala:* $\underset{n\times 1}{y_1} = c\,\underset{n\times 1}{z_1}$ *e* $\underset{n\times 1}{y_2} = \underset{n\times 1}{z_2}$ *per* $c > 0$.
Determinare la matrice di varianze/covarianze di $\underset{n\times 2}{Y} = [y_1 \ y_2]$ *e i relativi autovalori.*