

ESERCIZIO PCA

Dato il dataset X , con matrice di covarianza Σ , otteniamo la matrice degli autovettori U e il vettore degli autovalori S . Utilizziamo la tecnica della PCA, per ridurre la dimensionalità di X mantenendo almeno il 60% della varianza del dataset originale, per ricostruire i dati originali.

DATI:

dataset

$$X = \begin{pmatrix} 1 & 5 & 3 \\ 3 & 4 & 8 \\ 5 & 1 & 6 \end{pmatrix}$$

matrice di covarianza

$$\Sigma = \begin{pmatrix} 4 & 1 & -4 \\ 1 & 7 & 3.5 \\ -4 & 3.5 & 7 \end{pmatrix}$$

Autovettori

$$U = \begin{pmatrix} -0.34517975 & 0.54357965 & 0.76509614 \\ -0.62904883 & -0.73898939 & 0.2412307 \\ -0.69652603 & 0.39804488 & -0.59702231 \end{pmatrix}$$

autovalori

$$S = (\underbrace{11.2912878}_{\lambda_1}, \underbrace{6.70871215}_{\lambda_2}, \underbrace{3.9171372510 \times 10^{-16}}_{\lambda_3})$$

SOLGIMENTO:

1 VARIANZA TOTALE = $\lambda_1 + \lambda_2 + \lambda_3 \Rightarrow$ SOMMA AUTOVALEORI

$$= 11.2912878 + 6.70871215 + 3.9171372510 \times 10^{-16}$$

$$= 18.99999...6 \approx \textcircled{18}$$

2 CALCOLO RAPPORTI DI VARIANZA \Rightarrow N.B. \Rightarrow MANTENERE ALMENO 60% DEL DATASET ORIGINALE

$K=1$

$$\frac{\lambda_1}{\text{VARIANZA TOTALE}} = \frac{11.2912878}{18} = 0.63 \geq 0.6 \checkmark \Rightarrow \text{permette } K=1$$

GARANTITO 60% \checkmark di mantenere il 60%.

$K=2$

$$\frac{\lambda_1 + \lambda_2}{\text{VAR TOTALE}} = \frac{11.2912878 + 6.70871215}{18} = 1$$

N.B.
Scegliere numero colonne in base a quanto la varianza rappresenta il dataset.
 \rightarrow Se la varianza 60% preserva 2 colonne

3 SCELTA DEL NUMERO MINIMO DI COMPONENTI NECESSARIE

Con λ_1 si spiega il 60%, perciò $K=1 \Rightarrow$ dunque considero la PRIMA COLONNA di U

$K=1 \rightarrow$ SOLO UNA COLONNA DI U

$U_{\text{reduce}} = \begin{pmatrix} -0.34517975 \\ -0.62904883 \\ -0.69652603 \end{pmatrix}$ // si prende la 1ª colonna

4 PROIEZIONE DEL DATASET ORIGINALE SU U_{reduce}

$$Z = U_{\text{reduce}} \cdot X = \begin{bmatrix} -0.34517975 & -0.62904883 & -0.69652603 \end{bmatrix} \begin{pmatrix} 1 & 5 & 3 \\ 3 & 4 & 8 \\ 5 & 1 & 6 \end{pmatrix}$$

righe * colonne (somma)

$$= \begin{pmatrix} -5.11 & -4.84 & -10.25 \end{pmatrix}$$

DEPENDENT ANALYSIS

RICOSTRUIRE DATI ORIGINALI

$$\tilde{X} = U_{reduce} \cdot Z = \begin{pmatrix} -0.34514945 \\ -0.62904883 \\ -0.69652603 \end{pmatrix}$$

riga x column
(1) (1)

$$\begin{pmatrix} -3.41 & -4.84 & -10.25 \end{pmatrix}$$

$$= \begin{pmatrix} 1.97 & 1.07 & 3.59 \\ 3.59 & 3.10 & 6.64 \\ 3.57 & 3.63 & 7.13 \end{pmatrix} \approx \begin{pmatrix} 2 & 1 & 9 \\ 9 & 3 & 7 \\ 4 & 3 & 7 \end{pmatrix}$$

↓
matrice
di covarianza
60%
della varianza.