

Lezione : Analisi delle componenti principali

Docente: Aldo Solari

1 Trasformazioni lineari

Riduzione della dimensionalità:

$$\underset{n \times p}{X} \mapsto \underset{n \times q}{Y} \quad q \leq p$$

Restringiamo l'attenzione alle trasformazioni lineari, cercando di preservare più possibile la struttura dei dati originali. La *trasformazione lineare* di $\underset{n \times p}{X}$

$$\underset{n \times q}{Y} = \underset{n \times p}{X} \underset{p \times q}{A'} + \underset{n \times 1}{1} \underset{1 \times q}{b'}$$

è definita da

- la matrice $\underset{q \times p}{A}$
- il vettore $\underset{q \times 1}{b}$

Proposition 1.1. Il vettore delle medie $\underset{q \times 1}{\bar{y}}$ delle trasformazione lineare $\underset{n \times q}{Y} = \underset{n \times p}{X} \underset{p \times q}{A'} + \underset{n \times 1}{1} \underset{1 \times q}{b'}$ è dato da

$$\underset{q \times 1}{\bar{y}} = \underset{q \times p}{A} \underset{p \times 1}{\bar{x}} + \underset{q \times 1}{b}$$

Dimostrazione.

$$\underset{q \times 1}{\bar{y}} = \frac{1}{n} \underset{q \times n}{Y'} \underset{n \times 1}{1} = \frac{1}{n} \underset{q \times p}{A} \underset{p \times n}{X} \underset{n \times 1}{1} + \frac{1}{n} \underset{q \times 1}{b} \underset{1 \times n}{1'} \underset{n \times 1}{1} = \underset{q \times p}{A} \underset{p \times 1}{\bar{x}} + \underset{q \times 1}{b}$$

□

Proposition 1.2. La matrice di varianze/covarianze $\underset{q \times q}{S^Y}$ della trasformazione lineare $\underset{n \times q}{Y} = \underset{n \times p}{X} \underset{p \times q}{A'} + \underset{n \times 1}{1} \underset{1 \times q}{b'}$ è data da

$$\underset{q \times q}{S^Y} = \underset{q \times p}{A} \underset{p \times p}{S} \underset{p \times q}{A'}$$

Dimostrazione.

$$S^Y = \frac{1}{n} \tilde{Y}' \tilde{Y} = \frac{1}{n} A \tilde{X}' \tilde{X} A' = A S A'$$

dove

$$\tilde{Y} = H Y = H X A' + H \mathbf{1} b' = H X A' = \tilde{X} A'$$

□

Trasformazioni lineari note:

| | | | |
|-----|----------------------------|--|--|
| q | A $q \times p$ | b $q \times 1$ | $Y = X A' + \mathbf{1} b'$ $n \times q \quad n \times p \times q \quad n \times 1 \times q$ |
| | | | |
| p | I $p \times p$ | $-\bar{x}$ $p \times 1$ | \tilde{X} $n \times p$ |
| p | $D^{-1/2}$ $p \times p$ | $-D^{-1/2} \bar{x}$ $p \times p \quad p \times 1$ | Z $n \times p$ |
| p | $S^{-1/2}$ $p \times p$ | $-S^{-1/2} \bar{x}$ $p \times p \quad p \times 1$ | \tilde{Z} $n \times p$ |
| | | | |

1.1 Combinazione lineari

La combinazione lineare di X
 $n \times p$

$$y = X a = \begin{bmatrix} \sum_{j=1}^p a_j x_{1j} \\ \vdots \\ \sum_{j=1}^p a_j x_{ij} \\ \vdots \\ \sum_{j=1}^p a_j x_{nj} \end{bmatrix}$$

è un caso particolare di trasformazione lineare con $q = 1$, $A = a'$ e $b = 0$
 $q \times p \quad 1 \times p \quad q \times 1$

- $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = a' \bar{x}$
 $1 \times p \times p \times 1$
- $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = a' S a$
 $1 \times p \times p \times p \times 1$

Consideriamo ora la combinazione lineare di \tilde{X}

$$y = \tilde{X} a$$

- $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 0$

- $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = a' S a$

Qual è il vettore a che massimizza la varianza $a' S a$?

La varianza di y dipende dalla lunghezza del vettore a :

$$a' S a = \|a\|^2 \cdot v' S v$$

dove $v = \frac{a}{\|a\|}$ ha lunghezza unitaria $\|v\| = 1$

Di conseguenza, la varianza di una combinazione lineare $y = \tilde{X} a$ può essere resa grande/piccola a piacere cambiando la lunghezza di a .

Per questo motivo andremo a considerare solo vettori v di lunghezza unitaria $\|v\| = 1$, e diremo

che $y = \tilde{X} v$ è una *combinazione lineare normalizzata*

Theorem 1.3. Sia S la matrice di varianze/covarianze di \tilde{X} .

Il vettore v di lunghezza unitaria $\|v\| = 1$ che massimizza $v' S v$ è l'autovettore normalizzato v_1 (con segno $+$ o $-$) di S

$$\pm v_1 = \arg \max_{v: \|v\|=1} v' S v$$

e il massimo di $v' S v$ è pari all'autovalore più grande λ_1 di S

$$\lambda_1 = \max_{v: \|v\|=1} v' S v = v_1' S v_1 = (-v_1)' S (-v_1)$$

La combinazione lineare normalizzata

$$y_1 = \tilde{X} v_1$$

(oppure $-y_1$ con $-v_1$) è detta **prima componente principale** di \tilde{X} .

Dimostrazione. Sia $v = V w$, con w di lunghezza unitaria $\|w\| = 1$, tale che $\|v\| = \sqrt{v'v} = \sqrt{w'V'Vw} = \sqrt{w'w} = 1$, e V ha come colonne gli autovettori v_1, \dots, v_p di S .

$$v'Sv = w'V'SVw = w'V'V\Lambda V'Vw = w'\Lambda w = \sum_{j=1}^p w_j^2 \lambda_j$$

Il problema di massimo si riduce alla somma pesata degli autovalori $\lambda_1 \geq \dots \geq \lambda_p > 0$ con pesi pari a w_1^2, \dots, w_p^2 . Il problema ha soluzione nel vettore w che dà tutto il peso al primo autovalore λ_1 .

Il massimo si ottiene per $w_1 = \pm 1, w_2 = 0, \dots, w_p = 0$, quindi

$$V w = V \begin{bmatrix} \pm 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \pm v_1$$

□

2 Analisi delle componenti principali

Sia \tilde{X} con $\text{rango}(\tilde{X}) = p$.

Le p componenti principali di \tilde{X} sono le p colonne della trasformazione lineare

$$\begin{bmatrix} y_1 & y_2 & \dots & y_p \end{bmatrix} = Y = \tilde{X} V$$

dove le colonne di V sono gli autovettori normalizzati di S .

Per ridurre la dimensionalità di \tilde{X} basta considerare le prime $q < p$ componenti principali

$$\begin{bmatrix} y_1 & y_2 & \dots & y_q \end{bmatrix} = Y_q = \tilde{X} V_q$$

Si noti che la soluzione $-Y = \tilde{X} (-V)$ è equivalente a Y .

La derivazione delle componenti principali avviene sequenzialmente:

- si cerca la combinazione lineare normalizzata con varianza massima
- poi si cerca una seconda combinazione lineare normalizzata con varianza massima con il vincolo che sia incorrelata con la precedente;

- poi si cerca una terza combinazione lineare normalizzata con varianza massima e che sia incorrelata con le precedenti;
- e così via, determinando un numero di componenti principali pari al rango di \tilde{X}

Prima componente principale:

- I *pesi* (*loadings* in inglese) della prima componente principale di \tilde{X} sono gli elementi di

$$v_1 = \arg \max_{\substack{p \times 1 \\ v: \|v\|=1}} v' S v \quad 1 \times p p \times p p \times 1$$

dove v_1 è l'autovettore normalizzato di S associato a λ_1

- I *punteggi* (*scores* in inglese) della prima componente principale di \tilde{X} sono i valori della combinazione lineare normalizzata

$$y_1 = \tilde{X} v_1 \quad n \times 1 \quad n \times p p \times 1$$

- La *varianza spiegata* dalla prima componente principale di \tilde{X} è

$$\lambda_1 = v_1' S v_1 \quad 1 \times p p \times p p \times 1$$

Seconda componente principale:

- I *pesi* della seconda componente principale di \tilde{X} sono gli elementi di

$$v_2 = \arg \max_{\substack{p \times 1 \\ v: \|v\|=1, \\ v'v_1=0}} v' S v \quad 1 \times p p \times p p \times 1$$

dove v_2 è l'autovettore normalizzato di S associato a λ_2

- I *punteggi* della seconda componente principale di \tilde{X} sono i valori della combinazione lineare normalizzata

$$y_2 = \tilde{X} v_2 \quad n \times 1 \quad n \times p p \times 1$$

- La *varianza spiegata* dalla seconda componente principale di \tilde{X} è

$$\lambda_2 = v_2' S v_2 \quad 1 \times p p \times p p \times 1$$

j -sima componente principale:

- I pesi della j -sima componente principale di \tilde{X} sono gli elementi di

$$v_j = \underset{\substack{v: \|v\|=1, \\ v'v_k=0, k=1, \dots, j-1}}{\arg \max} \quad v' S v$$

dove v_j è l'autovettore normalizzato di S associato a λ_j

- I punteggi della j -sima componente principale di \tilde{X} sono i valori della combinazione lineare normalizzata

$$y_j = \tilde{X} v_j$$

- La varianza spiegata dalla j -sima componente principale di \tilde{X} è

$$\lambda_j = v_j' S v_j$$

2.1 Proprietà delle componenti principali

1. Il vettore delle medie di $Y = \tilde{X}V$ è nullo:

$$\frac{1}{n} Y' \mathbf{1} = \frac{1}{n} V' \tilde{X}' \mathbf{1} = V' \mathbf{0} = \mathbf{0}$$

2. La matrice di varianze/covarianze di $Y = \tilde{X}V$ è

$$S^Y = \frac{1}{n} Y' Y = \frac{1}{n} V' \tilde{X}' \tilde{X} V = V' S V = V' V \Lambda V' V = \Lambda$$

dove $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, ovvero y_1, \dots, y_p hanno varianze pari a $\lambda_1 \geq \dots \geq \lambda_p$ e sono tra loro incorrelati

3. Varianza totale di S^Y :

$$\text{tr}(S^Y) = \text{tr}(\Lambda) = \sum_{j=1}^p \lambda_j = \text{tr}(S)$$

coincide con la varianza totale di S

4. Proporzione di varianza spiegata dalla j -sima componente principale

$$\frac{\lambda_j}{\text{tr}(S)} = \frac{\lambda_j}{\sum_{k=1}^p \lambda_k}$$

5. Varianza generalizzata di S^Y :

$$\det(S^Y) = \det(\Lambda) = \prod_{j=1}^p \lambda_j = \det(S)$$

coincide con la varianza generalizzata di S

6. La correlazione tra la j -sima colonna \tilde{x}_j di \tilde{X} e i punteggi $y_k = \tilde{X}v_k$ della k -sima componente principale di \tilde{X} è pari a

$$\frac{v_{jk}\sqrt{\lambda_k}}{\sqrt{s_{jj}}}$$

Dimostrazione. Possiamo scrivere $\tilde{x}_j = \tilde{X} a_j$ dove a_j ha valore 1 in posizione j -sima e 0 altrove. La covarianza tra \tilde{x}_j e y_k è

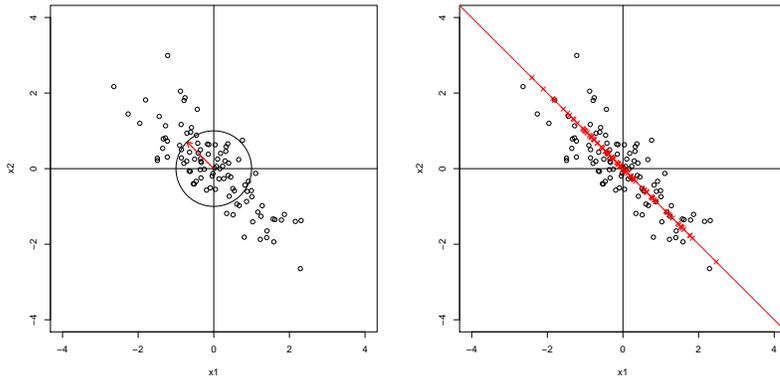
$$\frac{1}{n} \sum_{i=1}^n \tilde{x}_{ji} y_{ki} = \frac{1}{n} \tilde{x}'_j y_k = \frac{1}{n} a'_j \tilde{X}' \tilde{X} v_k = a'_j S v_k = a'_j \lambda_k v_k = \lambda_k v_{jk}$$

dove abbiamo utilizzato $S v_k = V \Lambda V' V a_k = V \Lambda a_k = V \lambda_k a_k = \lambda_k v_k$. La correlazione risulta quindi $\frac{v_{jk}\lambda_k}{\sqrt{\lambda_k}\sqrt{s_{jj}}} = \frac{v_{jk}\sqrt{\lambda_k}}{\sqrt{s_{jj}}}$. \square

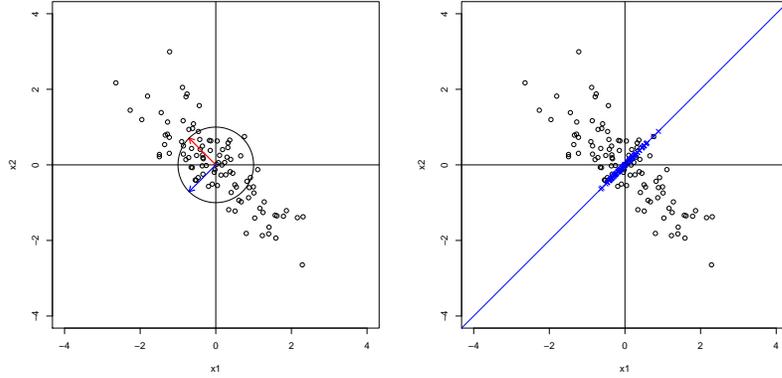
2.2 Interpretazione geometrica

La proiezione delle righe di \tilde{X} sul vettore v_1 è

$$\tilde{X} v_1 v_1' = y_1 v_1'$$



$p = 2$: vettore v_1 e proiezione delle righe di \tilde{X} su v_1



$p = 2$: vettore v_2 e proiezione delle righe di \tilde{X} su v_2

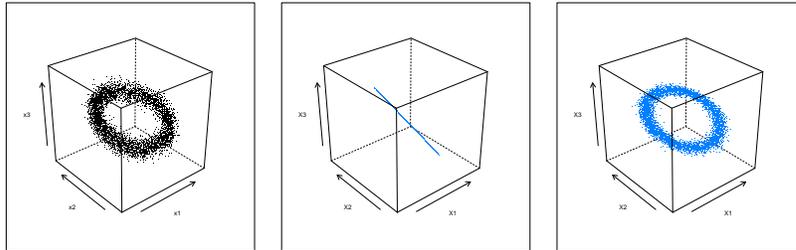
La proiezione delle righe di \tilde{X} sullo spazio generato da v_1, \dots, v_q , con $q \leq p$, è

$$\tilde{X} V_q V_q' = Y_q V_q'$$

$n \times p$ $p \times q$ $q \times p$ $n \times p$ $n \times q$ $q \times p$

dove

$$V_q = \begin{bmatrix} v_1 & \cdots & v_q \\ p \times 1 & & p \times 1 \end{bmatrix}$$



$p = 3$: proiezione di \tilde{X} su v_1 e sullo spazio generato da v_1 e v_2

Theorem 2.1 (Teorema di Eckart-Young). *La miglior approssimazione di rango $q \leq p$ della matrice \tilde{X} è data dalla matrice A di rango q definita da*

$$A = Y_q V_q' = \tilde{X} V_q V_q' = \arg \min_{B : \text{rango}(B)=q} \sum_{i=1}^n \sum_{j=1}^p (\tilde{x}_{ij} - b_{ij})^2$$

$n \times p$ $n \times q$ $q \times p$ $n \times p$ $p \times q$ $q \times p$ $n \times p$

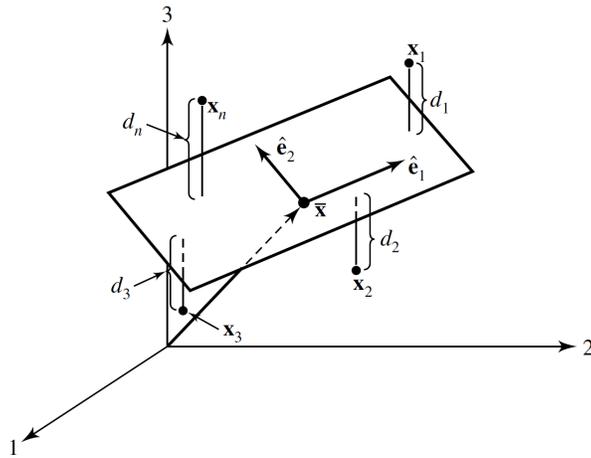
che minimizza l'errore di approssimazione

$$\sum_{i=1}^n \sum_{j=1}^p (\tilde{x}_{ij} - a_{ij})^2$$

rispetto a qualsiasi altra matrice B di rango q , i.e.

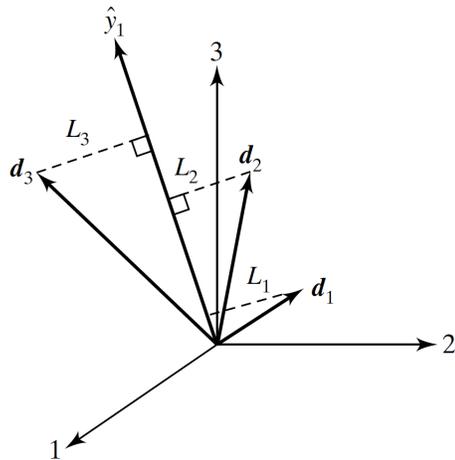
$$\sum_{i=1}^n \sum_{j=1}^p (\tilde{x}_{ij} - a_{ij})^2 \leq \sum_{i=1}^n \sum_{j=1}^p (\tilde{x}_{ij} - b_{ij})^2$$

Spazio delle variabili:



$p = 3$: il piano bidimensionale identificato da v_1 e v_2 minimizza la distanza al quadrato dai punti \tilde{x}'_i (le righe di \tilde{X})

Spazio delle osservazioni:



$n = 3$: il vettore y_1 minimizza le distanze al quadrato dai vettori scarto dalla media \tilde{x}_j (le colonne di \tilde{X})

3 Analisi delle componenti principali per dati standardizzati

L'analisi delle componenti principali non è invariante rispetto a trasformazioni lineari, e in particolare di scala.

Essendo le componenti principali costruite sulla base della matrice varianze/covarianze un cambiamento di scala che non sia omogeneo su tutte le variabili produce un cambiamento nelle varianze

col risultato di aumentare il peso nelle componenti principali di quelle variabili la cui varianza è aumentata.

Questo implica, ad esempio, che un cambiamento di unità di misura operato su una sola delle variabili modifica il risultato.

Queste considerazioni vanno tenute presenti quando si effettua un'analisi per decidere se partire da \tilde{X} o da Z ; la scelta andrà fatta caso per caso e non si danno regole generali.

L'analisi delle componenti principali di Z equivale a considerare la matrice di correlazione: $S^Z = R$.

- Le p componenti principali sono
$$Y_{n \times p} = Z_{n \times p} V_{p \times p}$$
- I pesi v_j della j -sima componente principale è il j -simo autovettore normalizzato di R associato al j -simo autovalore λ_j ; in generale (v_j, λ_j) di R sono diversi da quelli di S
- I punteggi della j -sima componente principale sono
$$y_j = Z_{n \times 1} v_j_{p \times 1}$$
- Poichè $\text{tr}(R) = p$, la proporzione di varianza spiegata dalla j -sima componente principale è λ_j/p
- La correlazione tra la j -sima colonna z_j di Z e i punteggi $y_k = Z v_k$ della k -sima componente principale di Z è pari a $v_{jk} \sqrt{\lambda_k}$

3.1 Caso $p = 2$

Consideriamo i dati standardizzati Z . La matrice di varianze e covarianze per Z :

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

con $r \geq 0$ I due autovalori di R sono

$$\lambda_1 = 1 + r, \quad \lambda_2 = 1 - r$$

I due autovettori normalizzati di R sono

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

I punteggi delle due componenti principali sono

$$y_{i1} = \frac{1}{\sqrt{2}}(z_{i1} + z_{i2}), \quad y_{i2} = \frac{1}{\sqrt{2}}(z_{i1} - z_{i2})$$

Se noti che se $r < 0$, l'ordine degli autovalori e quindi delle componenti principali è invertito