Aspetti algebrici delle componenti principali

Marco Riani,[*] Dipartimento di Scienze Economiche e Aziendale
and Interdepartmental Centre for Robust Statistics,
Università di Parma, 43100 Parma, Italy

23 ottobre 2020

1 Concetti propedeutici

Date 3 matrici conformabili A, B e C

$$(ABC)' = C'B'A'$$

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$$

$$det(ABC) = |ABC| = |A||B||C|$$

Il sistema lineare non omogeneo $Ax = b$ ammette una sola soluzione se la matrice A è non singolare (ossia se il suo determinante è diverso da zero). In tal caso $x = A^{-1}b$.

[*] e-mail: mriani@unipr.it

Data una matrice A quadrata, $tr(A)$ è la somma degli elementi sulla diagonale principale. L'operatore traccia è commutativo nel senso che $tr(ABC) = tr(CBA)$.

Il rango di una matrice $rank(A)$ di dimensione $n \times p$ è il massimo numero di righe (o colonne) linearmente indipendenti. Il rango gode della seguente proprietà:

- $rank(A) = rank(A'A) = rank(AA')$;

- $rank(AB) \leq min(rank(A), rank(B))$

Una matrice H di dimensione $n \times n$ si dice idempotente se $H \times H = H$.

Una matrice quadrata di dimensione V si dice ortogonale se la sua trasposta è uguale alla sua inversa: $V' = V^{-1}$. Il determinante di una matrice ortogonale è 1 oppure -1.

La norma $||x||$ (Euclidea) di un vettore $x = (x_1, x_2, \ldots, x_n)'$ di dimensione n (anche detta modulo o lunghezza di x) è definita dalla seguente espressione.

$$||x|| = \sqrt{\sum_i^n x_i^2} = \sqrt{x'x}$$

Due vettori colonna di lunghezza n si dicono ortogonali (perpendicolari) se $x'y = 0$. Ad esempio è facile verificare che i due vettori colonna v_1 e v_2 definiti come $v_1 = (a_1, 0)'$ e $v_2 = (0\ a_2)'$ dove a_1 e a_2 sono due numeri reali qualsiasi sono perpendicolari in quanto $v_1'v_2 = 0$.

2 Notazione

$X_{n\times p}$ = matrice dei dati originale. X può essere scritta come come insieme di p vettori colonna $X_1,\dots,X_p$. X_j, vettore di dimensione di dimensione $n\times 1$ che contiene i valori (modalità) della variabile j, con $j = 1,2,\dots,p$. $X = (X_1,\dots,X_j,\dots,X_p)$

X può anche essere scritta come come insieme di n vettori riga $x'_1\dots x'_n$ dove $x'_i = (x_{i1}, x_{i2},\dots,x_{ip})$.

$$X = \begin{pmatrix} x'_1 \\ x'_2 \\ \dots \\ x'_n \end{pmatrix}$$

$\tilde{X}_{n\times p}$ = matrice degli scostamenti dalla media

$1_{n\times 1}$ = vettore di lunghezza n con tutti gli elementi uguali ad 1

$1'_{n\times 1}\tilde{X} = 0_{1\times p}$ dove $0_{1\times p}$ è un vettore riga con elementi tutti uguali a zero di lunghezza p.

I_n = matrice identità di ordine n

$H = (I_n - 1_{n\times 1}1'_{n\times 1})/n$. Centering matrix. Matrice simmetrica e idempotente (v. zona D43:S58 del file di Excel `formulemat(out).xlsx`)

$H = H'$ e $H = H\times H$

$\tilde{X} = HX$

(v. zona D61:F76)

$S_{p\times p}$= matrice di covarianze. $S = \tilde{X}'\tilde{X}/(n-1) = X'HX/(n-1)$ (v. zona J62:L64). Il generico elemento i,i di questa matrice è $var(X_i) = s_i^2$, $i = 1,2,\dots,p$. Il generico elemento i,j di questa matrice è $cov(X_i,X_j)$,$i = 1,2,\dots,p$,

$j = 1, 2, \ldots, p$.

$D_{p \times p}$ = matrice diagonale che contiene sulla diagonale principale gli scostamenti quadratici medi (campionari) delle variabili originarie $D = diag(s_1, \ldots, s_p)$ (v. zona P62:R64).

$Z_{n \times p}$ = matrice che contiene gli scostamenti standardizzati. $Z = (Z_1, \ldots, Z_p)$. $var(Z_j) = 1$, e $1'_{n \times 1} Z_j = 0$. $Z = \tilde{X} D^{-1}$

(v. zona D81:F96).

$R_{p \times p}$ =matrice di correlazione. Il generico elemento i, i di questa matrice è 1, $i = 1, 2, \ldots, p$. Il generico elemento i, j di questa matrice è $r(X_i, X_j)$ ossia la correlazione tra la variabile X_i e la variabile X_j.

$$r(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{var(X_i) var(X_j)}}$$

$R = Z'Z/(n-1)$

(v. zona K79:M81).

$R = D^{-1} S D^{-1}$

(v. zona P79:R81).

3 Autovalori e autovettori

Data una generica matrice A di dimensione $m \times m$ e x un vettore colonna di lunghezza m, dato uno scalare λ (λ appartiene allo spazio dei numeri complessi) se si ha (per $x \neq 0$) che

$$Ax = \lambda x \tag{1}$$

allora λ è detto autovalore ed il corrispondente vettore x è detto autovettore della matrice A associato all'autovalore λ. L'autovettore, quindi, è un vettore non nullo che, moltiplicato per una matrice A, diventa un multiplo di sé stesso. Per una illustrazione grafica di questo concetto si veda il sito
https://blogs.mathworks.com/cleve/2013/07/08/eigshow-week-1/

Osservazione: gli autovettori di una matrice non sono unici: se x è un autovettore di A associato a λ anche cx, con c scalare qualsiasi è un autovettore di A associato a λ. $A(cx) = cAx = c\lambda x = \lambda(cx)$. Per fare in modo che l'autovettore sia definito unicamente si impone che la somma dei quadrati degli elementi di x sia uguale ad 1. In simboli: $x'x = 1$ (norma unitaria).

3.1 Polinomio caratteristico

L'obiettivo di questa sezione è capire come si calcolano in pratica gli autovalori ed i relativi autovettori. Da $Ax = \lambda x$, si ricava $(A - \lambda I_n)x = 0$, essendo I_n la matrice identità. Affinché esista $x \neq 0$ che soddisfa questa relazione, la matrice $A - \lambda I$ deve essere singolare cioè il suo determinante deve essere zero.

$$|A - \lambda I| = 0$$

Gli autovalori di una matrice A sono le radici del polinomiio caratteristico $p(\lambda)$ definito da

$$p(\lambda) = |A - \lambda I|$$

Ad esempio, se

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$$

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix}$$

il polinomio caratteristico è:

$$p(\lambda) = |A - \lambda I| = (2-\lambda)(3-\lambda) - 2 = \lambda^2 - 5\lambda + 4 \tag{2}$$

Risolvendo l'equazione di secondo grado si ottiene:

$$\lambda_1 = 4 \qquad \lambda_2 = 1$$

Sostituendo il valore di λ_1 nell'equazione (1) si ottiene

$$\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 4 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{cases} 2x_1 + 2x_2 = 4x_1 \\ 1x_1 + 3x_2 = 4x_2 \end{cases}$$

$$\begin{cases} 2x_2 = 2x_1 \\ x_1 = x_2 \end{cases}$$

Ovviamente il sistema è indeterminato e presenta infinite soluzioni. Per fare in modo che l'autovettore sia univocamente determinato si impone il vincolo che la somma dei quadrati dei suoi elementi sia uguali ad 1. In altri termini, si impone il vincolo di norma unitaria Ponendo quindi a sistema

$$\begin{cases} x_1 = x_2 \\ ||x||^2 = x_1^2 + x_2^2 = 1 \end{cases}$$

si ottiene che $x_1 = x_2 = 1/\sqrt{2}$. Il primo autovettore (che definiamo v_1) corrispondente all'autovalore $\lambda_1 = 4$ è dato da

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Con passaggi analoghi si ricava che il secondo autovettore (associato a $\lambda_2 = 1$) è dato dal vettore

$$v_2 = \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}$$

Osservazione: gli autovettori sono definiti a meno del segno nel senso che se x è un autovettore di A anche $-x$ è un autovettore di A in quanto l'equazione

$$Ax = \lambda x$$

è verificata anche da

$$A(-x) = \lambda(-x)$$

Osservazione 1: quando A è di ordine $n \times n$, la sua equazione caratteristica diventa un polinomio di grado n-esimo nella variabile λ che, quindi, ammette n soluzioni $\lambda_1, \ldots, \lambda_n$ reali o immaginarie a cui corrispondono n autovettori $x_1, \ldots, x_n$. Anche nel caso in cui la matrice A sia costituita da elementi reali, gli autovalori non sono necessariamente reali. Tuttavia, si può dimostrare che se la matrice A è simmetrica le radici caratteristiche sono sempre reali ed il numero di autovalori non nulli coincide con il rango di A.

Osservazione 2: se la matrice A è simmetrica e di ordine p, gli autovettori v_i e v_j relativi a due autovalori distinti λ_i e λ_j sono a due a due ortogonali ossia $v_i' v_j = 0$ $i \neq j = 1, 2, \ldots, p$. La matrice degli autovettori $V = (v_1, v_2, \ldots, v_p)$ in tal caso è ortogonale (ossia la sua inversa è uguale alla sua trasposta): $V'V = VV' = I_p$. Nel caso quindi in cui la matrice A sia uguale alla matrice di covarianze oppure alla matrice di correlazione (essendo queste matrici simmetriche) la matrice V è sempre ortogonale.

4 Calcolo degli autovalori/autovettori in Excel

La funzione `=xMCharPoly(Mat;"x";prec)` del componente aggiuntivo Xnumbers consente di trovare il polinomio caratteristico in funzione della variabile x.

`Mat` è la zona che contiene la matrice quadrata di partenza.

`prec` è uno scalare che consente di stabilire la precisione con cui si vogliono effettuare i calcoli.

Il polinomio caratteristico può essere inserito su una sola cella (in tal caso viene riportata l'equazione completa del polinomio caratteristico), oppure su più celle (in tal caso vengono riportati in celle diverse i coefficienti del polinomio caratteristico).
(v. ad esempio cella B107 oppure zona B111:B114).

La funzione `=PolySolve(input)` dove `input` è la cella che contiene il polinomio caratteristico o la zona che contiene i coefficienti del polinomio caratteristico, consente di trovare le radici del polinomio caratteristico ossia gli autovalori.

Se si parte da una matrice A di dimensione $m \times m$, per calcolare l'autovettore corrispondente all'autovalore occorre in una zona di dimensione $m \times m$ calcolare la matrice $A - \lambda I_m$ dove al posto di λ si deve sostituire il valore trovato dell'autovalore. In un'altra zona di dimensione $m \times 1$ occorre inserire un vettore di m elementi pari a 0 (vettore dei termini noti). Dato che il sistema è indeterminato una riga qualsiasi della matrice $A - \lambda I_m$ (senza perdita di generalità l'ultima) può essere modificata a piacimento. Similmente nella riga corrispondente del vettore di zeri possiamo inserire un numero a piacere (v. ad esempio zona G129:I132). A questo punto il sistema si risolve e si trova l'autovettore non normalizzato. Gli elementi dell'autovettore non normalizzato ovviamente non sono univoci in quanto dipendono da come è stata modificata l'ultima riga del sistema. L'autovettore normalizzato (quello in cui la somma dei quadrati degli elementi è pari a 1) al contrario è univoco ed è indipendente dalla modifica soggettiva che è stata fatta nell'ultima riga del sistema (v. ad esempio zona P130:P132).

Il procedimento di cui sopra va ripetuto per ogni altro autovalore. Se si parte da una matrice A di dimensione $m \times m$ gli autovalori sono m.

5 Calcolo degli autovalori/autovettori in MATLAB

La funzione MATLAB per trovare i coefficienti del polinomio caratteristico si chiama poly. Ad esempio `polinomiocar=poly([2 2; 1 3])` restituisce il vettore riga `polinomiocar` di lunghezza 3 contenente i numeri (v. i coefficienti dell'equazione di secondo grado riportata nell'equazione 2):

$$1 \quad -5 \quad 4$$

La funzione `roots(polinomiocar)` consente di trovare le radici del polinomio caratteristico, ossia gli autovalori .

La function `eig` calcola direttamente tutti gli autovalori e gli autovettori della matrice A. Più precisamente, `e=eig(A)` fornisce nel vettore colonna di output e tutti gli autovalori della matrice A. Al contrario, `[V,D]=eig(A)` fornisce la matrice diagonale D, contenente gli autovalori sulla diagonale principale, e la matrice V, contenente gli autovettori.

Osservazione la funzione `eig` non sempre restituisce gli autovalori ordinati di conseguenza è necessario ordinare gli autovalori ed i correspondenti autovalori utilizzando la funzione `sort`.

6 Scomposizione spettrale

Qualsiasi matrice quadrata simmetrica $S_{p \times p}$ può essere scomposta come segue:

$$S = V \Lambda V' \tag{3}$$

dove V è la matrice ortonormale ($V'V = I_p$) che contiene nelle colonne i p autovettori: $V = (v_1, v_2, \ldots, v_p)$ e Λ è la matrice diagonale che contiene sulla diagonale principale i p autovalori di S. Nel linguaggio dell'algebra lineare si dice che le colonne della matrice V formano una base ortonormale.
(v. zona N160:P162).

S può anche essere scritta come:

$$\begin{aligned}
S &= (v_1, v_2 \ldots, v_p) \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_p \end{pmatrix} \begin{pmatrix} v_1', \\ v_2' \\ \ldots, \\ v_p' \end{pmatrix} \\
&= (\lambda_1 v_1, \lambda_2 v_2 \ldots, \lambda_p v_p) \begin{pmatrix} v_1', \\ v_2' \\ \ldots, \\ v_p' \end{pmatrix} \\
S &= \sum_{i=1}^{p} \lambda_i v_i v_i'
\end{aligned}$$

L'implementazione della formula di cui sopra è ad esempio nella zona C190:E192

Vediamo ora un esempio di scomposizione spettrale della matrice

$$A = \begin{pmatrix} \frac{1}{3} & \frac{\sqrt{20}}{3} \\ \frac{\sqrt{20}}{3} & \frac{2}{3} \end{pmatrix}.$$

Risolvendo l'equazione caratteristica: $\lambda_i^2 - \lambda_i - 2 = 0$, otteniamo che $\lambda_1 = 2$ e $\lambda_2 = -1$. I corrispondenti autovettori normalizzati sono riportati nella matrice che segue

$$V = \begin{pmatrix} \frac{2}{3} & -\frac{\sqrt{5}}{3} \\ \frac{\sqrt{5}}{3} & \frac{2}{3} \end{pmatrix}.$$

La scomposizione spettrale è la seguente:

$$A = V\Lambda V' = \begin{pmatrix} \frac{1}{3} & \frac{\sqrt{20}}{3} \\ \frac{\sqrt{20}}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{\sqrt{5}}{3} \\ \frac{\sqrt{5}}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{\sqrt{5}}{3} \\ -\frac{\sqrt{5}}{3} & \frac{2}{3} \end{pmatrix}.$$

7 La prima componente principale come combinazione lineare delle variabili originarie

Obiettivo: data $\tilde{X}$ (matrice degli scostamenti dalla media) oppure Z matrice degli scostamenti standardizzati, si cerca il vettore a dimensione $p \times 1$ in modo tale che sia massima $var(\tilde{X}a)$ oppure la $var(Za)$. con il vincolo $a'a = 1$.

Proposizione: il vettore a che contiene i coefficienti della combinazione lineare delle varibili originarie che massimizza la varianza è il primo autovettore associato al primo autovalore della matrice di covarianze S (se si parte da $\tilde{X}$ oppure è il primo ed primo autovettore associato al primo autovalore della matrice di correlazione R (se si parte da Z).

8 Le prime k componenti principali come combinazioni lineari delle variabili originarie

In generale il nostro obiettivo è trovare k vettori a_1, $a_2, \ldots, a_k$ ciascuno combinazione lineare delle variabili originarie (senza perdità di generalità supponiamo di partire dalla matrice degli scostamenti standardizzati)

$$y_1 = Za_1 \qquad y_2 = Za_2, \qquad \ldots, \qquad y_r = Za_k$$

in termini matriciali

$$Y_{n\times k} = (y_1,\ y_2,\ \ldots, y_k) = Z_{n\times r}A_{p\times k}$$

in modo tale che la somma delle varianze della matrice Y sia la più grande possibile. In simboli, l'obiettivo è:

$$\max_A tr\ cov(Y) = \max_A tr\ cov(ZA)$$

con il vincolo che $A'A = I_r$. Questa espressione è massimizzata quando $A = V = (v_1, v_2, \ldots, v_k)$ ossia quando A contiene i primi k autovettori associati ai primi k autovalori della matrice R. Se invece che partire dalla matrice Z fossimo partiti dalla matrice $\tilde{X}$, allora l'espressione precedente sarebbe stata massimizzata quando A contiene i primi k autovettori associati ai primi k autovalori della matrice S.

8.1 Relazione tra autovalori traccia e determinante

L'obiettivo di questa sezione è quello di dimostrare la relazione che sussiste tra la traccia (somma degli elementi sulla diagonale principale) di una matrice quadrata e simmetrica, il suo determinante ed i suoi autovalori.

$$tr(A) = tr(V\Lambda V') = tr(\Lambda V'V) = tr(\Lambda) = \sum_{i=1}^{p} \lambda_i$$

di conseguenza, la somma degli autovalori di una matrice simmetrica è pari alla somma degli elementi sulla sua diagonale principale.

Osservazione: nel caso in cui la generica matrice A, su cui si vuole fare la scomposizione spettrale, sia la matrice di covarianze S di ordine p, si ha che

$$tr(S) = \sum_{i=1}^{p} var(X_i) = tr(\Lambda) = \sum_{i=1}^{p} \lambda_i$$

la somma degli autovalori, quindi, non è altro che la somma delle varianze delle variabili originarie. Di conseguenza, il rapporto:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} = \frac{\sum_{i=1}^{k} \lambda_i}{tr(S)} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} var(X_i)} \tag{4}$$

indica la quota di varianza delle p variabili originarie che viene spiegata dalle prime k componenti principali. Si noti che nel caso in cui si operi sulla matrice di correlazione

$$\sum_{i=1}^{p} var(X_i) = \sum_{i=1}^{p} \lambda_i = p$$

Se invece dell'operatore traccia applichiamo l'operatore determinante si ottiene che:

$$|A| = |V\Lambda V'| = |\Lambda V'V| = |\Lambda| = \prod_{i=1}^{p} \lambda_i$$

ossia il prodotto degli autovalori è esattamente uguale al determinante della matrice A. Se la matrice A è la matrice di covarianze allora dato che $|S|$ può interpretarsi come una varianza generalizzata, l'estrazione delle prime k componenti principali massimizza la varianza generalizzata che si può estrarre considerando k combinazioni lineari delle variabili originarie.

Osservazione: in questa sezione per dimostrare la relazione tra traccia, determinante e autovalori abbiamo utilizzato la scomposizione spettrale (cioè abbiamo supposto che la matrice fosse simmetrica). In realtà la relazione vista in questa sezione vale per qualsiasi matrice A non necessariamente simmetrica purché i suoi autovalori siano reali. E' facile verificare che nell'esempio visto nella sezione 3

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$$

la somma degli autovalori $1 + 4 = 5$ coincide con $tr(A) = 2 + 3 = 5$ e che il prodotto degli autovalori $1 \times 4 = 4$ coicide con il determinante $det(A) = |A| = 6 - 2 = 4$.

V. anche zone D124:D125 e zona I124:I125

9 La scomposizione in valori singolari (svd)

Qualsiasi matrice X di dimensioni $n \times p$ di rango r (con $r \leq min(n,p)$) può essere scomposta come

$$X_{n\times p} = U_{n\times r}\Gamma_{r\times r}V'_{r\times p} \tag{5}$$

dove

$U = (u_1, u_2, \ldots, u_r)$ è una matrice ortogonale $U'U = I_r$ che contiene gli autovettori di XX'.

$V = (v_1, v_2, \ldots, v_r)$ è una matrice ortogonale $V'V = I_r$ che contiene gli autovettori di $X'X$.

Γ è una matrice diagonale che contiene sulla diagonale principale i valori singolari (ossia le radici quadrate) degli r autovalori non nulli $\lambda_1, \lambda_2, \ldots, \lambda_r$ delle matrici $X'X$ oppure XX'.

$$\Gamma = \begin{pmatrix} \gamma_1 & 0 & \ldots & 0 \\ 0 & \gamma_2 & \ldots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \gamma_r \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_r} \end{pmatrix} = \Lambda^{0.5}$$

Si osservi che le colonne della matrice $Y = XV$ sono tra loro incorrelate, ossia la matrice di covarianze della matrice XV è diagonale:

$$cov(Y) = cov(XV) = (U\Gamma)'(U\Gamma)/(n-1) = \Lambda/(n-1)$$

Questo significa che nella scomposizione in valori singolari data una determinata matrice di partenza (diciamo A) e due generici vettori di norma unitaria (diciamo

x e y), dal punto di vista geometrico la svd si ha quando i due vettori Ax e Ay sono tra loro ortogonali (perpendicolari). Questo caso è esemplificato alla p. web https://blogs.mathworks.com/cleve/2013/08/05/eigshow-week-3
Nel nostro esempio precedente, x e y sono due generiche colonne della matrice degli autovettori V (ossia due vettori ortogonali, ossia due vettori che formano sempre nello spazio a due dimensioni un angolo di 90 gradi). I vettori Ax e Ay sono due nuovi vettori non necessariamente ortogonali. Se i due nuovi vettori Ax e Ay vengono scelti come ortogonali allora si ha la svd. Nel nostro contesto la matrice A non è altro che $\tilde{X}$ (oppure la matrice Z), i vettori Ax e Ay non solo altro che due generiche componenti principali $\tilde{X}v_i$ e $\tilde{X}v_j$ (nuovi vettori ortogonali) e identificano rispettivamente due nuovi assi. In presenza di due sole dimensioni, le prime due componenti principali identificano rispettivamente gli assi maggiore e minore dell'ellisse. La lunghezza dei semiassi dell'ellisse è esattamente uguale ai valori singolari della matrice A (ossia $\tilde{X}$ oppure Z). In termini algebrici la lunghezza di $\tilde{X}v_i/\sqrt{n-1}$, ossia la norma Euclidea della i-esima componente principale, ossia la radice quadrata della varianza della i-esima componente principale:

$$\begin{aligned}\sqrt{var(y_i)} &= \sqrt{var(\tilde{X}v_i)} = \sqrt{(\tilde{X}v_i)'(\tilde{X}v_i)/(n-1)} \\ &= \sqrt{v_i'(\tilde{X}'\tilde{X})/(n-1)v_i} = \sqrt{v_iSv_i} = \sqrt{v_i'V\Lambda V'v_i} \\ &= \sqrt{e_i'\Lambda e_i} = \sqrt{\lambda_i} = \gamma_i\end{aligned}$$

dove e_i è un vettore colonna composto da elementi tutti uguali a 0 eccetto per la posizione i-esima dove il valore che assume è 1. $v_i'V = v_i'(v_1, v_2, \ldots, v_p) = e_i'$ in quanto $v_i'v_j = 0$ per $i \neq j$ e $v_i'v_i = 1$ per $i = j$. Dato che, per definizione della prima componente principale la norma precedente viene massimizzata, la prima componente principale nel caso di due dimensioni identifica l'asse maggiore dell'ellisse. Dato che la seconda componente principale è ortogonale alla prima, essa identifica l'asse minore dell'ellisse.

Un modo alternativo di scrivere la s.v.d. è il seguente:

$$X = \sum_{i=1}^{r} \gamma_i u_i v_i'$$

che mostra come qualsiasi generica matrice X di rango r può essere scritta come somma r matrici di dimensioni $n \times p$ ciascuna di rango 1
(v. zona C269:E284).
La matrice $\gamma_i u_i v_i'$ è di rango 1, in quanto per le proprietà del rango:

$$\begin{aligned} rank(u_i v_i') &= rank((u_i v_i')' u_i v_i') = rank(v_i u_i' u_i v_i') \\ &= rank(v_i v_i') \leq min(rank(v_i), rank(v_i')) = min(1, 1) = 1 \end{aligned}$$

10 svd in Excel

La funzione `=xSVDU(Mat,DigitMax)` consente di calcolare la matrice U,
la funzione `=xSVDD(Mat,DigitMax)` consente di calcolare la matrice Γ,
la funzione `=xSVDV(Mat,DigitMax)` consente di calcolare la matrice V

11 svd in MATLAB

La funzione `[U,Gamma,V] = svd(A,'econ')` consente di calcolare nei 3 argomenti di output `U`, `Gamma` e `V` le 3 matrici viste sopra. E' facile verificare in Matlab che `A=U*Gamma*V'`.

12 Le prime k componenti principali come migliore rappresentazione di rango k delle variabili originarie

Il nostro obiettivo è quello di sostituire al posto della matrice $X_{n\times p}$ di partenza di rango r (il cui generico elemento è x_{ij}) una nuova matrice $\hat{X}_{n\times p}$ (il cui generico elemento è $\hat{x}_{ij}$) di rango ridotto, in modo tale che sia minima la somma dei quadrati delle differenze

$$\min_{\hat{x}_{ij}}(x_{ij}-\hat{x}_{ij})^2$$

Dato che la somma dei quadrati degli elementi di una matrice si può scrivere come:

$$\sum_{i=1}^{n}\sum_{j=1}^{p}x_{ij}^2 = tr(X'X)$$

si ottiene che

$$\min_{\hat{x}_{ij}}(x_{ij}-\hat{x}_{ij})^2 = tr[(X-\hat{X})'(X-\hat{X})]$$

Se per esempio cerchiamo come matrice $\hat{X}$ una matrice di rango 2 abbiamo che:

$$X - \hat{X} = \sum_{i=1}^{r} \gamma_i u_i v_i' - \sum_{i=1}^{2} \gamma_i u_i v_i' = \sum_{i=3}^{r} \gamma_i u_i v_i'$$

$$\begin{aligned}
tr\left((X-\hat{X})'(X-\hat{X})\right) &= tr\left(\sum_{i=3}^{r} \gamma_i u_i v_i' \sum_{i=3}^{r} \gamma_i u_i v_i'\right) \\
&= tr\left(\sum_{i=3}^{r} \sum_{j=3}^{r} v_i \gamma_i u_i' u_j \gamma_j v_j'\right) \\
&= tr\left(\sum_{i=3}^{r} \sum_{j=3}^{r} v_i \gamma_i \gamma_j v_j'\right) \\
&= tr\left(\sum_{i=3}^{r} \sum_{j=3}^{r} \gamma_i \gamma_j v_j' v_i\right) \\
&= tr\left(\sum_{i=3}^{r} \gamma_i^2 v_i' v_i\right) \\
&= tr\left(\sum_{i=3}^{r} \gamma_i^2\right) \\
&= \sum_{i=3}^{r} \gamma_i^2 \\
&= \sum_{i=3}^{r} \lambda_i
\end{aligned}$$

In altri termini, la miglior rappresentazione di rango 2 della matrice originaria (ossia quella che minimizza la somma dei quadrati dei residui) si ottiene quando si prendono i primi due autovettori associati ai primi due autovalori della matrice $X'X$ (XX') oppure della matrice $\tilde{X}'\tilde{X}$, $(\tilde{X}\tilde{X}')$.

Se si dividono le precedenti equazioni per $(n-1)$ si ha che

$$tr\left((X-\hat{X})'(X-\hat{X})\right)/(n-1)$$

è pari $\sum_{i=3}^{r}\lambda_i$ dove in questo caso i λ_i sono gli autovalori della matrice $\tilde{X}'\tilde{X}/(n-1)$ ossia gli autovalori della matrice di covarianze.
Per l'implementazione di questo concetto si veda ad esempio la cella I298.

Se come matrice X si sceglie la matrice $\tilde{X}$ la scomposizione svd può essere scritta come

$$\begin{aligned}\tilde{X} &= \sqrt{n-1}U\frac{\Gamma^*}{\sqrt{n-1}}V' \\ \tilde{X}/\sqrt{n-1} &= U\Gamma V' \\ \tilde{X} &= \sqrt{n-1}U\Gamma V'\end{aligned}$$

dove $\Gamma = \frac{\Gamma^*}{\sqrt{n-1}}$ è la matrice che contiene le radici quadrate degli autovalori della matrice $\tilde{X}'\tilde{X}/(n-1) = S$ ossia le radice quadrate degli autovalori della matrice di covarianze.
(v.zona K64:Q70 del foglio 'X(database originale)' del file di Excel benessere(out).xlsx).
La matrice $\sqrt{n-1}U\Gamma = \tilde{X}V = Y$ contiene i valori della prime r componenti principali.
(v. zona B4:H106 del foglio 'Y (matrice comp prin)' del file benessere(out).xlsx.
La matrice V contiene gli autovettori della matrice $\tilde{X}'\tilde{X}/(n-1) = S$ ossia gli

autovettori della matrice di varianze e covarianze S (oppure gli autovettori della matrice di correlazione se si parte dalla matrice Z anziché dalla matrice $\tilde{X}$).
v. zona K74:Q80 del foglio ‘X (database originale)’ del file benessere(out).xlsx.

Dato che:

$$tr(\tilde{X}'\tilde{X})/(n-1) = tr(S) = \sum_{i=1}^{p} var(X_i)$$

possiamo riassumere affermando che, se al posto della matrice originaria di dimensioni $n \times p$ di rango p si sostituisce la miglior approssimazione di rango k, la quota non spiegata di varianza delle variabili originarie è esattamente uguale a:

$$\frac{\sum_{i=k+1}^{p} \lambda_i}{\sum_{i=1}^{p} var(X_i)}$$

Si noti che questo risultato è esattamente uguale identico a quello ottenuto nella sezione 8, equation (4) in cui le prime k componenti principali erano state introdotte come combinazione lineare delle variabili originarie.

Si noti che

$$\begin{aligned}
cov(\sqrt{n-1}U\Gamma) &= cov(\tilde{X}V) = cov(Y) && (6)\\
&= (\tilde{X}V)'(\tilde{X}V)/(n-1) = V'\tilde{X}'\tilde{X}/(n-1)V && (7)\\
&= V'SV = V'V\Lambda V'V = \Lambda. && (8)
\end{aligned}$$

La matrice di varianze e covarianze della matrice delle componenti principali
(v. zona B4:H106 del foglio ‘Y (matrice comp princ)’ del file benessere(out).xlsx.
è diagonale (ossia le componenti principali sono tra loro incorrelate) e la varianza

della i-esima componente principale è pari all'i-esimo autovalore λ_i della matrice di covarianze.

(v. zona J7:P7 dello stesso foglio di lavoro).

$$
\begin{aligned}
cov(\sqrt{n-1}U) &= cov(ZV\Gamma^{-1}) \\
&= (ZV\Gamma^{-1})'(ZV\Gamma^{-1})/(n-1) \\
&= \Gamma^{-1}V'(Z'Z/(n-1))V\Gamma^{-1} \\
&= \Gamma^{-1}V'RV\Gamma^{-1} = \Gamma^{-1}V'V\Lambda V'V\Gamma^{-1} \\
&= \Gamma^{-1}\Lambda\Gamma^{-1} = \Gamma^{-1}\Gamma\Gamma\Gamma^{-1} = I_p
\end{aligned}
$$

La matrice $\sqrt{n-1}U$, quindi, contiene i valori delle componenti principali standardizzate (ossia con varianza unitaria).

V. zona W179:X179 del foglio 'X (database originale)' del file benessere(out).xlsx.

.