

Combinazioni lineari

Situazione 25.1. $X \in \mathbb{R}_m^n$ sia la nostra matrice di dati con $n \geq 2$. Come nella nota 24.7 e quando non indicato diversamente, siano $\lambda_1, \dots, \lambda_m$ gli autovalori di X^{COM} con $\lambda_1 \geq \dots \geq \lambda_m$ ed e^1, \dots, e^m una base ortonormale di \mathbb{R}_m tale che

$$\varphi_X e^1 = \lambda_1 e^1, \dots, \varphi_X e^m = \lambda_m e^m$$

Inoltre sia $1 \leq q \leq m$.

Nota 25.2. Sia $g \in \mathbb{R}_m$ con $|g| = 1$. Per definizione

$$X_g^i = \|X^i - \bar{X}, g\|$$

è la lunghezza con segno (calcolata a partire dal baricentro \bar{X}) della proiezione ortogonale del punto X^i sulla retta generata da g . Per l'osservazione 24.9 abbiamo

$$X_g = X^{CE} g^t$$

Questa, per la nota 22.15, è una combinazione lineare delle colonne di X^{CE} con i coefficienti g_j :

$$X_g = \sum_{j=1}^m (X^{CE})_j g_j$$

che, tenendo conto della definizione di X^{CE} , possiamo scrivere in forma ancora più esplicita:

$$X_g = \sum_{j=1}^m (X_j - \bar{X}_j 1^n) g_j$$

In particolare abbiamo

$$X_{e^k} = \sum_{j=1}^m (X^{CE})_j e_j^k = \sum_{j=1}^m (X_j - \bar{X}_j 1^n) e_j^k$$

per ogni $k = 1, \dots, m$.

La k -esima componente principale di X è quindi una combinazione lineare delle colonne di X^{CE} con i coefficienti dell'autovettore e^k .

Osservazione 25.3. $X_{e^j} \perp X_{e^k}$ per ogni $j \neq k$.

Dimostrazione. Siccome per $a, b \in \mathbb{R}^n$ abbiamo $\|a, b\| = \|a^t, b^t\|$, dove il secondo prodotto scalare è calcolato in \mathbb{R}_n , possiamo scrivere

$$\begin{aligned} \|X_{e^j}, X_{e^k}\| &= \|X^{CE} (e^j)^t, X^{CE} (e^k)^t\| \\ &= \|e^j (X^{CE})^t, e^k (X^{CE})^t\| \\ &= \|e^j X^{COM}, e^k\| \\ &= \|\lambda_j e^j, e^k\| = \lambda_j \|e^j, e^k\| = 0 \end{aligned}$$

Autovalori

In R gli autovalori di una matrice reale o complessa si trovano con la funzione `eigen`. Essa calcola, se non si pone l'opzione

```
only.values=T
```

anche un sistema di autovettori; ciò può rallentare il calcolo per matrici molto grandi. Il risultato è una *lista* in R e le componenti si ottengono con la sintassi `$values` e `$vectors`. Creiamo due funzioni per la nostra libreria:

```
Mm.autovalori = function (A,simm=F)
{eigen(A,only.values=T,
  symmetric=simm)$values}
```

```
Mm.autovettori = function (A,simm=F)
{eigen(A,symmetric=simm)$vectors}
```

`Mm.autovalori` restituisce la lista degli autovalori di A , in ordine decrescente (rispetto al modulo se complessi).

`Mm.autovettori` restituisce una *matrice* le cui colonne sono autovettori di A corrispondenti agli autovalori nell'ordine indicato. Con l'opzione `simm=T` possiamo accelerare i calcoli per matrici simmetriche.

Per il calcolo di X^{COM} potremmo usare la definizione 24.2 e la funzione `Smg.cen` definita a pagina 15; possiamo però anche usare la relazione

$$X^{COM} = (n - 1) \text{cov}(X)$$

con

```
Sm.com = function (X)
(nrow(X)-1)*cov(X)
```

Siccome `Mm.autovettori(A)` per $A = X^{COM}$ è una matrice le cui colonne sono gli autovettori di X^{COM} già normati e nell'ordine desiderato (cioè corrispondenti agli autovalori elencati in ordine decrescente) e tenendo conto del fatto che questi autovettori in R sono vettori colonna, possiamo trovare gli autovettori e i componenti principali di X con le seguenti funzioni:

```
Smp.autovettori = function (X)
{Com=Sm.com(X)
Mm.autovettori(Com,simm=T)}
```

```
Smp = function (X)
Smg.cen(X)%*%Smp.autovettori(X)
```

Potremmo anche usare la funzione `princomp` di R:

```
Smp.R = function (X)
{p=princomp(X)$scores
dimnames(p)=NULL; p}
```

L'istruzione `dimnames(p)=NULL` ha lo scopo di ridurre gli attributi della matrice a quelli di una matrice pura.

Per gli autovalori di X^{COM} usiamo

```
Smp.autovalori = function (X)
{Com=Sm.com(X)
Mm.autovalori(Com,simm=T)}
```

In questo numero

- 25 Combinazioni lineari
Autovalori
La matrice $A^t A$
Un metodo con molti nomi
- 26 Componenti principali
Inversione al cerchio unitario
- 27 La lemniscata ellittica
La traccia
- 28 Ortoregressione su iperpiani
Bibliografia

La matrice $A^t A$

Osservazione 25.4. Sia $A \in \mathbb{R}_m^n$. Allora le colonne di A sono linearmente dipendenti se e solo se esiste $x \in \mathbb{R}^m \setminus 0$ tale che $Ax = 0$.

Dimostrazione. $Ax = \sum_{k=1}^m A_k x^k$ è una

combinazione lineare delle colonne di A e ogni tale combinazione lineare può essere scritta in questo modo.

Osservazione 25.5. Sia $A \in \mathbb{R}_m^n$. Allora la matrice $A^t A \in \mathbb{R}_m^m$ è simmetrica. Inoltre:

- (1) $A^t A$ è positivamente semidefinita.
- (2) $A^t A$ è positivamente definita se e solo se le colonne di A sono linearmente indipendenti.

Dimostrazione. (1) Sia $x \in \mathbb{R}_m$. Allora

$$x^t A^t A x = (Ax)^t Ax = \|Ax, Ax\| \geq 0.$$

(2) $A^t A$ sia positivamente definita ed $x \in \mathbb{R}_m^m$ tale che $Ax = 0$. Ciò implica $x^t A^t Ax = 0$ e quindi $x = 0$.

Siano viceversa le colonne di A linearmente indipendenti. Sia $x^t A^t Ax = 0$, cioè $\|Ax, Ax\| = 0$. Allora $Ax = 0$ e quindi $x = 0$.

Corollario 25.6. Sia $A \in \mathbb{R}_m^n$. Allora gli autovalori di $A^t A$ sono ≥ 0 . Essi sono tutti > 0 se e solo se le colonne di A sono linearmente indipendenti.

Un metodo con molti nomi

L'analisi delle componenti principali appare in molti campi della matematica applicata con diversi nomi: *Trasformazione sulle assi principali* in geometria, *trasformazione di Karhunen-Loève* in ingegneria e nella teoria del riconoscimento delle forme e nell'elaborazione delle immagini, *analisi spettrale* in fisica e analisi matematica (ad esempio problemi agli autovalori per equazioni differenziali), *analisi fattoriale* in psicologia (anche se con questo termine spesso si associano obiettivi più ambiziosi della sola riduzione delle dimensioni). Essa è spesso un primo passo preparatorio che permette di applicare altri metodi della statistica multivariata, come l'analisi dei raggruppamenti e la ricerca di funzioni discriminanti.

Componenti principali

Nota 26.1. Consideriamo la matrice di dati

$$X := \begin{pmatrix} 1 & 4 \\ 2 & 1 \\ 5 & 1 \\ 2 & 4 \end{pmatrix}$$

Con `Sm.com(X)` otteniamo

$$X^{COM} = \begin{pmatrix} 9 & -6 \\ -6 & 9 \end{pmatrix}$$

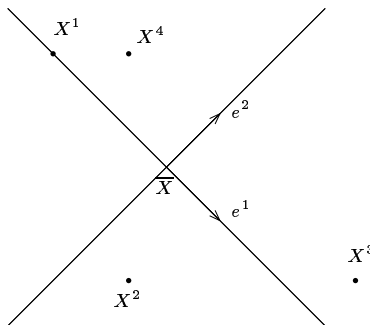
Gli autovalori di X^{COM} sono

$$\lambda_1 = 15 \quad \text{e} \quad \lambda_2 = 3$$

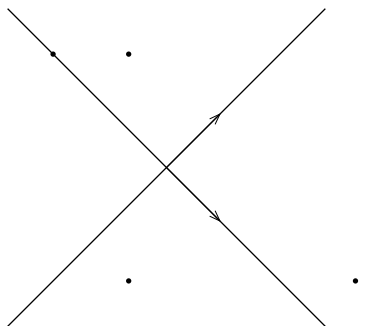
Ad essi corrispondono gli autovettori ortormali

$$e^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad e^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

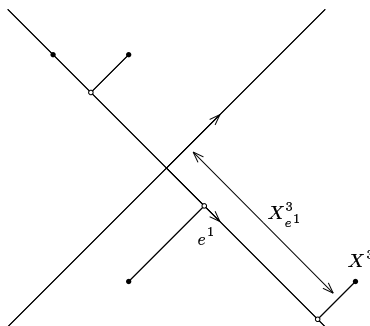
Il baricentro è $\bar{X} = (2.5, 2.5)$; lo troviamo con `colMeans(X)` oppure, in un esempio così semplice, con un calcolo diretto.



Vediamo che effettivamente si ha l'impressione che la variazione maggiore avvenga nella direzione e^1 , quella minore nella direzione e^2 , in accordo con quanto osservato nella nota 24.13. Ciò si vede ancora meglio se togliamo le leggende:



Come osservato, $X_{e^1}^i$ è la lunghezza con segno (calcolata rispetto al centro \bar{X}) della proiezione ortogonale di X^i sulla retta generata da e^1 :



Nota 26.2. Una volta determinato (in una delle possibili scelte) e^1 , otteniamo una proiezione

$$\begin{aligned} \mathbb{R}_m &\rightarrow \mathbb{R}_1 \\ x &\mapsto \bar{X} + \|x - \bar{X}, e^1\| e^1 \end{aligned}$$

nella quale in particolare i punti X^i vengono proiettati secondo

$$X^i \mapsto P_{e^1}^i = \bar{X} + X_{e^1}^i e^1$$

Nell'esempio della nota 26.1 calcoliamo prima le lunghezze delle proiezioni sugli assi principali con

```
p=Smp(X)
print(round(p,2))
```

ottenendo la matrice

-2.12	0.00
0.71	-1.41
2.83	0.71
-1.41	0.71

in cui la prima colonna si riferisce all'asse determinato da e^1 . Usare un righello per misurare (in cm) le lunghezze sull'ultima figura della nota 26.1 per convincersi che il risultato è corretto.

Riportando i valori $X_{e^1}^i$ su un'ascisse otteniamo un'immagine unidimensionale dei nostri dati:



Il cerchietto bianco è qui l'origine di \mathbb{R}_1 .

Sappiamo dalla nota 24.13 che la varianza di questi punti in \mathbb{R}_1 è uguale a $\lambda_1 = 15$, mentre la varianza della proiezione sul secondo asse principale è uguale a $\lambda_2 = 3$ e quindi molto minore. Possiamo perciò sperare che X_{e^1} da solo ci dia sufficienti informazioni. Assumiamo che, in un altro esempio, le proiezioni sull'asse più importante siano distribuite come nella seguente figura:

Allora possiamo considerare i punti come appartenenti a tre gruppi distinti e se gli altri autovalori di X^{COM} sono molto più piccoli di λ_1 , questo raggruppamento potrà, con molta prudenza, essere considerato significativo. Si tenga conto del fatto che proprio l'analisi delle componenti principali è molto sensibile alla scala usata, ad esempio al cambio delle unità di misura nel rilevamento delle variabili.

Inversione al cerchio unitario

Nota 26.3. Chiamiamo *inversione al cerchio unitario* l'applicazione $K : \mathbb{C} \setminus 0 \rightarrow \mathbb{C}$ che manda ogni punto $z \neq 0$ nel punto z' che si trova sulla semiretta che parte dall'origine e passa per z , avendo però un modulo che è il reciproco di quello di z , cioè tale che

$$|z'| = \frac{1}{|z|}$$

Questa applicazione è talvolta anche detta *riflessione al cerchio unitario*. K è univocamente determinata dalla condizione enunciata e

$$K(z) = \frac{z}{|z|^2} = \frac{1}{\bar{z}}$$

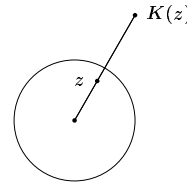
per ogni $z \in \mathbb{C} \setminus 0$.

Dimostrazione. Infatti dobbiamo avere $K(z) = tz$ con $t > 0$ e inoltre deve valere $tz = \frac{1}{|z|}$; essendo $t > 0$ ciò è equivalente a

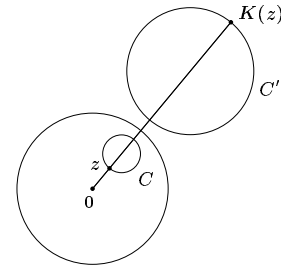
$t|z| = \frac{1}{|z|}$, cioè $t = \frac{1}{|z|^2}$. Quindi l'immagine $K(z) = \frac{z}{|z|^2}$ è univocamente determinata. Osservando che $|z|^2 = z\bar{z}$, vediamo che

$$K(z) = \frac{z}{z\bar{z}} = \frac{1}{\bar{z}}$$

Verificare da soli che $K \circ K = \text{id}$; è invece chiaro direttamente dalla definizione che i punti fissi di K sono esattamente i punti del cerchio unitario e che ogni punto all'interno del cerchio unitario viene trasformato in un punto esterno e viceversa.

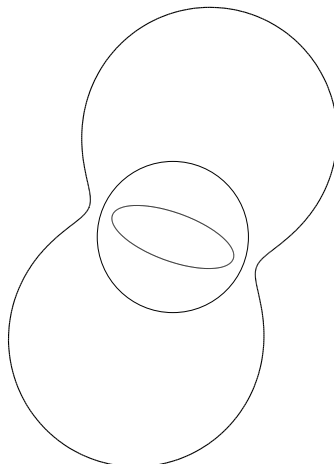


L'inversione al cerchio unitario possiede interessanti proprietà geometriche, molte delle quali sono descritte nel bellissimo libro di Needham e di cui la più importante è quella che attraverso K ogni cerchio C che non passa per l'origine viene trasformato in un cerchio C' (anch'esso non passante per l'origine).



Si noti che il centro di C' non è l'immagine del centro di C !

Vedremo adesso che, mentre l'inversione al cerchio unitario trasforma cerchi in cerchi, un'ellisse (che non sia un cerchio) con centro nell'origine viene invece trasformata in una curva di quarto grado (una lemniscata ellittica); infatti la geometria dell'ellisse è molto più profonda e difficile della geometria del cerchio.



La lemniscata ellittica

Nota 27.1. Consideriamo un'ellisse nel piano con centro nell'origine, descritta da un'equazione

$$f(x, y) = 1$$

dove $f(x, y) = \alpha x^2 + 2\beta xy + \gamma y^2$ è una forma quadratica (reale) positivamente definita. Per $z = x + iy$ scriviamo anche $f(z)$ invece di $f(x, y)$. Per ogni $t \in \mathbb{R}$ allora $f(tz) = t^2 f(z)$. È ovvio inoltre che l'origine non appartiene all'ellisse, perché $f(0, 0) = 0 \neq 1$. Denotiamo di nuovo con K l'inversione al cerchio unitario.

Siano adesso z un punto dell'ellisse e

$$w := K(z) = \frac{z}{|z|^2}$$

Allora

$$f(w) = \frac{1}{|z|^4} f(z) = \frac{1}{|z|^4}$$

perché $f(z) = 1$ essendo z un punto dell'ellisse. D'altra parte abbiamo

$$|z| = \frac{1}{|w|}$$

per definizione di K e quindi $f(w) = |w|^4$. I punti w dell'immagine dell'ellisse sotto K soddisfano quindi l'equazione

$$|w|^4 = f(w)$$

che, se poniamo $w = u + iv$ con $u, v \in \mathbb{R}$, può essere scritta anche nella forma

$$(u^2 + v^2)^2 = f(u, v)$$

oppure, ancora più esplicitamente,

$$(u^2 + v^2)^2 = \alpha u^2 + 2\beta uv + \gamma v^2$$

Si osservi però che oltre ai punti riflessi dell'ellisse anche l'origine soddisfa questa equazione. Curve con questa equazione si chiamano *lemniscate ellittiche* quando, come nella nostra ipotesi, la forma quadratica f è positivamente definita.

Nota 27.2. Vediamo adesso che la lemniscata ellittica può essere utilizzata per rappresentare il quoziente di Rayleigh di una forma quadratica in due dimensioni.

Nelle ipotesi e con la notazione della nota 27.1 sia z_0 il punto sulla circonferenza unitaria determinato da z , cioè $z_0 = \frac{z}{|z|}$.

Siano

$$L := \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

e $\varphi := \bigcirc_p L : \mathbb{R}_2 \rightarrow \mathbb{R}_2$ l'operatore simmetrico associato. Allora $\|p, \varphi p\| = f(p)$ è vediamo che $f(z_0) = \mathcal{R}_\varphi(z_0)$ è proprio il quoziente di Rayleigh di φ in z_0 . D'altra parte

$$f(z_0) = \frac{1}{|z|^2} f(z) = \frac{1}{|z|^2}$$

e quindi $\frac{1}{|z|} = \sqrt{f(z_0)}$ oppure, equivalentemente,

$$|z| = \frac{1}{\sqrt{f(z_0)}}$$

Questa prima equazione mostra che il modulo di un punto z dell'ellisse $f(z) = 1$ è uguale a $\frac{1}{\sqrt{f(z_0)}}$, dove z_0 è il vettore unitario che mostra nella stessa direzione di z .

Oltre a ciò abbiamo però anche

$$w = \frac{z}{|z|^2} = \frac{z_0}{|z|} = z_0 \cdot \sqrt{f(z_0)}$$

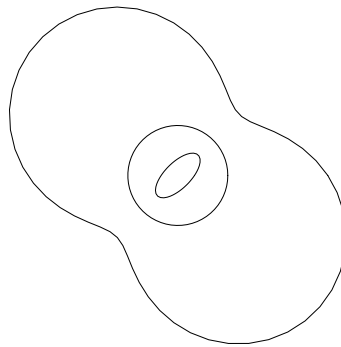
quindi

$$w = z_0 \cdot \sqrt{f(z_0)}$$

Il punto w che si ottiene da un punto z dell'ellisse mediante inversione al cerchio unitario è quindi quel vettore che si ottiene moltiplicando il vettore unitario z_0 con la radice $\sqrt{f(z_0)}$ del quoziente di Rayleigh in quella direzione.

Nel caso statistico di $L = X^{COM}$ con $m = 2$ il fattore $\sqrt{f(z_0)}$ è uguale alla deviazione standard di X_{z_0} .

La realizzazione in \mathbb{R} è semplicissima.



Otteniamo questa figura con il programma

```
# Quadrica della nota 26.1.
f = function(x,y)
9*x^2-12*x*y+9*y^2

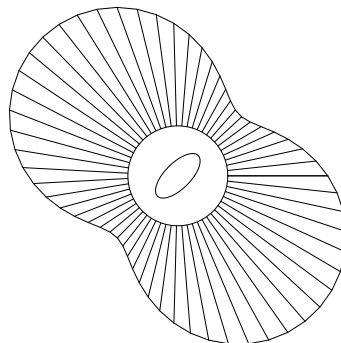
# Radice di f(z).
rfz = function(z)
{x=Re(z); y=Im(z)
sqrt(f(x,y))}

Gr.postscript('27-lem-1.ps',5,5)
larg=7
intervallo=c(-larg/2, larg/2)
Gr(intervallo, intervallo)
alfa=seq(0, 2*pi, length=60)
z=exp(i*alfa)
lines(z)
lines(z/rfz(z))
lines(z*rfz(z))
dev.off()
```

Aggiungendo le righe

```
raggi=Mm(c(z, z*rfz(z)),
rep(NA, length(z))), righe=3)
lines(raggi)
```

possiamo evidenziare i raggi tra i punti z_0 del cerchio unitario e i corrispondenti punti w sulla lemniscata.



La traccia

Definizione 27.3. Sia $A \in \mathbb{R}_m^m$. Definiamo la *traccia* di A , denotata con $\text{tr } A$, come la somma degli elementi della diagonale principale di A , quindi

$$\text{tr } A := \sum_{i=1}^m A_i^i$$

È chiaro che l'applicazione $\text{tr} : \mathbb{R}_m^m \rightarrow \mathbb{R}$ è lineare. La traccia gode però di molte altre proprietà importanti, tra cui i sorprendenti corollari 27.5 e 27.6.

Proposizione 27.4. Siano $A \in \mathbb{R}_m^n$ e $B \in \mathbb{R}_n^m$. Allora $\text{tr } AB = \text{tr } BA$.

Si noti che $AB \in \mathbb{R}_m^m$, mentre $BA \in \mathbb{R}_n^m$.

Dimostrazione. Abbiamo

$$\begin{aligned} \text{tr } AB &= \sum_{i=1}^m (AB)_i^i \\ &= \sum_{i=1}^m \sum_{j=1}^m A_j^i B_i^j = \sum_{j=1}^m \sum_{i=1}^m B_i^j A_j^i \\ &= \sum_{j=1}^m (BA)_j^j = \text{tr } BA \end{aligned}$$

Corollario 27.5. Siano $f \in \mathbb{R}_m$ e $v \in \mathbb{R}_m$. Allora $f v = \text{tr } v f$.

Corollario 27.6. (1) Siano $A \in \mathbb{R}_p^n$ e $v \in \mathbb{R}^n$. Allora

$$\sum_{j=1}^p \|A_j, v\|^2 = v^t A A^t v = \text{tr } A^t v v^t A$$

(2) Siano $f \in \mathbb{R}_m$ e $B \in \mathbb{R}_m^n$. Allora

$$\sum_{i=1}^n \|f, B^i\|^2 = f B^t B f^t = \text{tr } B f^t f B^t$$

Dimostrazione. Per le proposizioni 23.5 e 23.7 abbiamo $\sum_{j=1}^p \|A_j, v\|^2 = v^t A A^t v$ e

$\sum_{i=1}^n \|f, B^i\|^2 = f B^t B f^t$. L'enunciato segue dalla proposizione 27.4.

Nota 27.7. Sia $A \in \mathbb{R}_m^n$. Allora

$$\begin{aligned} \text{tr } A^t A &= \text{tr } A A^t \\ &= \sum_{i=1}^n |A^i|^2 = \sum_{j=1}^m |A_j|^2 \\ &= \sum_{i=1}^n \sum_{j=1}^m (A_j^i)^2 \end{aligned}$$

non è altro che il quadrato della lunghezza di A considerata come vettore di $\mathbb{R}^{m \cdot n}$.

Lemma 27.8. Sia $G \in \mathbb{R}_m^q$. Allora

$$\sum_{i=1}^n \sum_{j=1}^q \|X^i - \bar{X}, G^j\|^2 = \text{tr } G X^{COM} G^t$$

Dimostrazione. Dal corollario 27.6 abbiamo, usando anche la proposizione 22.13,

$$\begin{aligned} &\sum_{i=1}^n \sum_{j=1}^q \|X^i - \bar{X}, G^j\|^2 \\ &= \sum_{i=1}^n \text{tr } G (X^i - \bar{X})^t (X^i - \bar{X}) G^t \\ &= \text{tr } G \left(\sum_{i=1}^n (X^i - \bar{X})^t (X^i - \bar{X}) \right) G^t \\ &= \text{tr } G X^{COM} G^t \end{aligned}$$

Ortoregressione su iperpiani

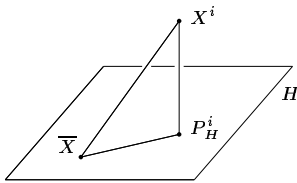
Nota 28.1. Vogliamo adesso dimostrare che, per $1 \leq q \leq m$, l'iperpiano $\bar{X} + \mathbb{R}e^1 + \dots + \mathbb{R}e^q$ minimizza la somma dei quadrati delle distanze dei punti X^i da un iperpiano q -dimensionale H di \mathbb{R}^m . Chiamiamo un iperpiano con questa proprietà un q -iperpiano di regressione ortogonale per X ; si può dimostrare anche in questo caso che esso passa per il baricentro \bar{X} . Possiamo quindi trovare vettori $g^1, \dots, g^q \in \mathbb{R}^m$ con $\|g^j, g^k\| = \delta_{jk}$ per ogni j, k e

$$H = \bar{X} + \mathbb{R}g^1 + \dots + \mathbb{R}g^q$$

Per la proposizione 8.2 la proiezione P_H^i di X^i su H è data da

$$P_H^i = \bar{X} + \sum_{j=1}^q \|X^i - \bar{X}, g^j\| g^j$$

Noi dobbiamo scegliere g^1, \dots, g^q in modo da minimizzare $\sum_{i=1}^n \|X^i - P_H^i\|^2$.



Anche qui abbiamo

$$\|X^i - P_H^i\|^2 = \|X^i - \bar{X}\|^2 - \|P_H^i - \bar{X}\|^2$$

Ma

$$\|P_H^i - \bar{X}\|^2 = \sum_{j=1}^q \|X^i - \bar{X}, g^j\|^2$$

cosicché

$$\sum_{i=1}^n \|X^i - P_H^i\|^2 = \sum_{i=1}^n \|X^i - \bar{X}\|^2 - \sum_{i=1}^n \sum_{j=1}^q \|X^i - \bar{X}, g^j\|^2$$

La somma $\sum_{i=1}^n \|X^i - \bar{X}\|^2$ però dipende solo dalla matrice dei dati X e non dai vettori g^j che vogliamo variare per ottenere il minimo di $\sum_{i=1}^n \|X^i - P_H^i\|^2$ e vediamo che quest'ultima somma è minima se e solo se $\sum_{i=1}^n \sum_{j=1}^q \|X^i - \bar{X}, g^j\|^2$ è massima.

Osserviamo adesso che

$$\sum_{i=1}^n \sum_{j=1}^q \|X^i - \bar{X}, g^j\|^2 = \sum_{j=1}^q \sum_{i=1}^n \|X^i - \bar{X}, g^j\|^2 = \sum_{j=1}^q \mathcal{R}_{\varphi_X}(g^j)$$

usando la nota 24.6. Dobbiamo quindi massimizzare $\sum_{j=1}^q \mathcal{R}_{\varphi_X}(g^j)$, dove g^1, \dots, g^q varia tra i sistemi ortogonali di q vettori di lunghezza 1.

Definizione 28.2. $O(m)$ sia l'insieme delle matrici ortogonali di rango m . È chiaro che una matrice $G \in \mathbb{R}^m$ appartiene ad $O(m)$ se e solo se le righe G^1, \dots, G^m costituiscono una base ortonormale di \mathbb{R}^m .

Osservazione 28.3. Sia $G \in O(m)$. Allora

$$\sum_{i=1}^n \sum_{j=1}^m \|X^i - \bar{X}, G^j\|^2 = \sum_{j=1}^m \mathcal{R}_{\varphi_X}(G^j) = \text{tr } X^{COM}$$

Vediamo così che questa espressione non dipende da G .

Dimostrazione. Infatti in questo caso

$$\text{tr } GX^{COM}G^t = \text{tr } X^{COM}G^tG = \text{tr } X^{COM}$$

Definizione 28.4. Sia $G \in \mathbb{R}_m^q$. Diciamo che G possiede righe ortonormali e scriviamo $G \in O^q(m)$, se le righe di G hanno tutte lunghezza 1 e sono ortogonali tra di loro.

Osservazione 28.5. Sia $G \in \mathbb{R}_m^q$. Allora

$$G \in O^q(m) \iff GG^t = \delta$$

Si noti che $GG^t \in \mathbb{R}_q^q$ e quindi δ è la matrice identica in \mathbb{R}_q^q .

Definizione 28.6. $\text{diag}(\lambda_1, \dots, \lambda_q)$ sia la matrice diagonale i cui coefficienti diagonali sono $\lambda_1, \dots, \lambda_q$.

Lemma 28.7. Siano $L \in \mathbb{R}_m^m$, $F \in O^q(m)$ e $\lambda_1, \dots, \lambda_q \in \mathbb{R}$ tali che $F^tL = \lambda_i F^i$ per ogni i . Allora

$$FLF^t = \text{diag}(\lambda_1, \dots, \lambda_q)$$

e quindi $\text{tr } FLF^t = \sum_{j=1}^q \lambda_j$.

Dimostrazione. Per ogni i, j abbiamo

$$(FLF^t)_j^i = (FL)^i(F^t)_j = \lambda_i F^i(F^t)_j = \lambda_i \delta_{ij}$$

Osservazione 28.8. $D = \text{diag}(\lambda_1, \dots, \lambda_m)$ sia una matrice diagonale e $B \in \mathbb{R}_m^m$. Allora

$$\text{tr } DB = \sum_{j=1}^m \lambda_j B_j^j$$

Dimostrazione. Infatti

$$(DB)_j^i = \sum_{k=1}^m D_k^i B_j^k = \sum_{k=1}^m \delta_{ik} \lambda_i B_j^k = \lambda_i B_j^i$$

Osservazione 28.9. $D = \text{diag}(\lambda_1, \dots, \lambda_q)$ sia una matrice diagonale e $C \in \mathbb{R}_m^q$. Allora

$$\text{tr } CDC^t = \sum_{j=1}^m \lambda_j |C_j|^2$$

Dimostrazione. Per l'osservazione 28.8

$$\text{tr } CDC^t = \text{tr } DC^tC = \sum_{j=1}^m \lambda_j (C^tC)_j^j$$

Ma $(C^tC)_j^j = |C_j|^2$.

Osservazione 28.10. Siano $u \in \mathbb{R}^q$, $v \in \mathbb{R}^s$

$$e w := \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^{q+s}.$$

Allora $|w|^2 = |u|^2 + |v|^2$.
Ciò implica in particolare $|u|^2 \leq |w|^2$.

Lemma 28.11. Sia $C \in O^q(m)$.

Allora $\text{tr } C^tC = q$.

Dimostrazione. $CC^t = \delta \in \mathbb{R}_q^q$, per cui $\text{tr } C^tC = \text{tr } CC^t = \text{tr } \delta = q$.

Lemma 28.12. Siano dati numeri reali x_1, \dots, x_m , $\lambda_1, \dots, \lambda_m$ tali che siano soddisfatte le seguenti condizioni:

- (1) $0 \leq x_j \leq 1$ per ogni j .
- (2) $\sum_{j=1}^m x_j = q$.
- (3) $\lambda_1 \geq \dots \geq \lambda_m$

Allora $\sum_{j=1}^q \lambda_j \geq \sum_{j=1}^m \lambda_j x_j$.

Dimostrazione. L'enunciato può essere interpretato come l'affermazione che il compito di ottimizzazione $Q(x_1, \dots, x_m) = \max Q(x_1, \dots, x_m) := \sum_{j=1}^m \lambda_j x_j$ sotto le condizioni (1) e (2) possiede la soluzione

$$x_1 = \dots = x_q = 1, x_{q+1} = \dots = x_m = 0$$

Ciò è evidente, perché significa che dobbiamo concentrare le „risorse“ x_j nei primi q posti dove la rendita è massima, esaurendo in questo modo però la risorsa totale q .

Teorema 28.13. Siano $L \in \mathbb{R}_m^m$ ed $E \in O(m)$. Sia $F \in \mathbb{R}_m^q$ la matrice con le righe E^1, \dots, E^q . Allora $F \in O^q(m)$ e per ogni $G \in O^q(m)$ vale $\text{tr } FLF^t \geq \text{tr } GLG^t$.

Dimostrazione. Sia $G \in O^q(m)$.

Siccome le righe di E costituiscono una base di \mathbb{R}_m , per ogni i esiste una rappresentazione $G^i = \sum_{j=1}^m C_j^i E^j$. I coefficienti C_j^i formano una matrice $C \in \mathbb{R}_m^q$ per cui $G = CE$. Sia $D := \text{diag}(\lambda_1, \dots, \lambda_m)$.

Per il lemma 28.7 abbiamo $GLG^t = CELE^tC^t = CDC^t$, cosicché dall'osservazione 28.9 segue

$$\text{tr } GLG^t = \sum_{j=1}^m \lambda_j |C_j|^2.$$

Abbiamo inoltre $C = GE^t$ e quindi $CC^t = GE^tEG^t = GG^t = \delta$, per cui $C \in O^q(m)$.

Esiste perciò una matrice $H \in O^{m-q}(m)$ tale che $U := \begin{pmatrix} C \\ H \end{pmatrix} \in O(m)$. Per l'osservazione 28.10 ciò implica $|C_j|^2 \leq |U_j|^2 = 1$ per ogni j .

Per il lemma 28.11 $\text{tr } C^tC = q$ e quindi, per la nota 27.7, $\sum_{j=1}^m |C_j|^2 = q$.

$$\text{Per il lemma 28.7 } \text{tr } FLF^t = \sum_{j=1}^q \lambda_j.$$

L'enunciato segue dal lemma 28.12.

Teorema 28.14. $\bar{X} + \mathbb{R}e^1 + \dots + \mathbb{R}e^q$ è un q -iperpiano di regressione ortogonale per X .

Bibliografia

17123 **A. Coffman/M. Frantz:** Möbius transformations and ellipses. Internet 2004, 9p.
15993 **I. Jolliffe:** Principal component analysis. Springer 2002.
(16783) **T. Needham:** Visual complex analysis. Oldenbourg 2001.