

L'insieme di Rayleigh

Situazione 21.1. V sia uno spazio vettoriale reale di dimensione finita $m \geq 1$ e $\|\cdot\|$ un prodotto scalare reale su V . $\varphi: V \rightarrow V$ sia un'applicazione lineare *simmetrica* rispetto a $\|\cdot\|$, cioè nel senso della definizione 20.2. $\lambda_1, \dots, \lambda_m$ siano gli autovalori (necessariamente reali) di φ e $\lambda_1 \geq \dots \geq \lambda_m$.

Definizione 21.2. Per $v \in V \setminus \{0\}$ sia

$$\mathcal{R}_\varphi(v) := \frac{\|v, \varphi v\|}{\|v, v\|}$$

il *quoziente di Rayleigh* di φ in v . Per un sottoinsieme $X \subset V$ sia inoltre

$$\mathcal{R}_\varphi(X) := \{\mathcal{R}_\varphi(v) \mid v \in X \setminus \{0\}\}$$

$\mathcal{R}_\varphi(X)$ si chiama *l'insieme di Rayleigh* di φ su X . $\mathcal{R}_\varphi(V)$ in alcuni libri di analisi numerica è anche detto semplicemente l'insieme dei valori di φ .

Osservazione 21.3. v sia un autovettore di φ per l'autovalore λ . Allora $\mathcal{R}_\varphi(v) = \lambda$.

Dimostrazione. Un autovettore è $\neq 0$, perciò il quoziente di Rayleigh è definito. Inoltre

$$\frac{\|v, \varphi v\|}{\|v, v\|} = \frac{\|v, \lambda v\|}{\|v, v\|} = \frac{\lambda \|v, v\|}{\|v, v\|} = \lambda$$

Osservazione 21.4. Sia $v \in V \setminus \{0\}$. Allora $\mathcal{R}_\varphi\left(\frac{v}{\|v\|}\right) = \mathcal{R}_\varphi(v)$. Perciò

$$\mathcal{R}_\varphi(V) = \mathcal{R}_\varphi(\{v \in V \mid \|v\| = 1\})$$

Lemma 21.5. e_1, \dots, e_m sia una base ortonormale di V e $v, w \in V$ con

$$\begin{aligned} v &= \alpha_1 e_1 + \dots + \alpha_m e_m \\ w &= \beta_1 e_1 + \dots + \beta_m e_m \end{aligned}$$

con $\alpha_i, \beta_j \in \mathbb{R}$. Allora

$$\|v, w\| = \sum_{k=1}^m \alpha_k \beta_k$$

Dimostrazione. Abbiamo

$$\begin{aligned} \|v, w\| &= \|\alpha_1 e_1 + \dots + \alpha_m e_m, \beta_1 e_1 + \dots + \beta_m e_m\| \\ &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \beta_j \|e_i, e_j\| \\ &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \beta_j \delta_{ij} = \sum_{k=1}^m \alpha_k \beta_k \end{aligned}$$

Corollario 21.6. e_1, \dots, e_m sia una base ortonormale di V e $v \in V$ con

$$v = \alpha_1 e_1 + \dots + \alpha_m e_m$$

Allora:

(1) $\|v, v\| = \sum_{k=1}^m \alpha_k^2$.

(2) Se gli e_k sono autovettori di φ con $\varphi e_k = \lambda_k e_k$ per ogni k , allora

$$\|v, \varphi v\| = \sum_{k=1}^m \lambda_k \alpha_k^2$$

Dimostrazione. (1) segue dal lemma 21.5.

(2) L'ipotesi implica che

$$\begin{aligned} \varphi v &= \varphi(\alpha_1 e_1 + \dots + \alpha_m e_m) \\ &= \alpha_1 \varphi e_1 + \dots + \alpha_m \varphi e_m \\ &= \alpha_1 \lambda_1 e_1 + \dots + \alpha_m \lambda_m e_m \end{aligned}$$

L'enunciato segue ancora dal lemma 21.5.

Lemma 21.7. Siano dati numeri reali a_1, \dots, a_m con $\alpha := \min(a_1, \dots, a_m)$, $\beta := \max(a_1, \dots, a_m)$. Allora l'involuppo convesso dell'insieme $\{a_1, \dots, a_m\}$ è l'intervallo $[\alpha, \beta]$.

Dimostrazione. $I := [\alpha, \beta]$ è un insieme convesso che contiene tutti i punti dati. Dobbiamo dimostrare che I è il più piccolo insieme convesso con questa proprietà. È chiaro che non possiamo togliere le estremità α e β , perché esse appartengono ai punti a_j . Se invece da I togliamo punti interni, otteniamo un insieme che non è più convesso, perché il segmento che congiunge α e β è proprio I .

Teorema 21.8. $\mathcal{R}_\varphi(V) = [\lambda_m, \lambda_1]$.

Dimostrazione. Per la nota 20.16 esiste una base ortonormale e_1, \dots, e_m di V tale che $\varphi e_1 = \lambda_1 e_1, \dots, \varphi e_m = \lambda_m e_m$.

(1) Sia $v \in V \setminus \{0\}$, ad esempio $v = \alpha_1 e_1 + \dots + \alpha_m e_m$. Per il corollario 21.6 allora

$$\frac{\|v, \varphi v\|}{\|v, v\|} = \frac{\lambda_1 \alpha_1^2 + \dots + \lambda_m \alpha_m^2}{\alpha_1^2 + \dots + \alpha_m^2}$$

Per $1 \leq k \leq m$ sia $t_k := \frac{\alpha_k^2}{\alpha_1^2 + \dots + \alpha_m^2}$.

Allora $t_k \geq 0$ e $t_1 + \dots + t_m = 1$, e vediamo che

$$\frac{\|v, \varphi v\|}{\|v, v\|} = t_1 \lambda_1 + \dots + t_m \lambda_m$$

appartiene all'involuppo convesso dei numeri reali $\lambda_1, \dots, \lambda_m$.

(2) Se viceversa sono dati numeri reali $t_1, \dots, t_m \geq 0$ con $t_1 + \dots + t_m = 1$, e se poniamo $\alpha_k := \sqrt{t_k}$ per ogni k , allora

$$\frac{\alpha_k^2}{\alpha_1^2 + \dots + \alpha_m^2} = \frac{t_k}{t_1 + \dots + t_m} = t_k$$

e, ponendo $v := \alpha_1 e_1 + \dots + \alpha_m e_m$, come prima

$$t_1 \lambda_1 + \dots + t_m \lambda_m = \frac{\|v, \varphi v\|}{\|v, v\|}$$

(3) Ciò mostra che $\mathcal{R}_\varphi(V)$ coincide con l'involuppo convesso dei numeri reali $\lambda_1, \dots, \lambda_m$ e quindi, per il lemma 21.7, con $[\lambda_m, \lambda_1]$.

Corollario 21.9.

$$\begin{aligned} \lambda_1 &= \max \mathcal{R}_\varphi(V) \\ \lambda_m &= \min \mathcal{R}_\varphi(V) \end{aligned}$$

In questo numero

- 21 L'insieme di Rayleigh
Matrici normali
Il baricentro
- 22 Spazi ortogonali intermedi
Calcolo matriciale
- 23 Formule per il prodotto scalare
Regressione ortogonale
- 24 La formula di proiezione
La matrice di covarianza
Bibliografia

Matrici normali

Una matrice $A \in \mathbb{R}^m$ si dice *normale*, se $AA^t = A^t A$. Una matrice simmetrica è evidentemente normale, ma anche ogni matrice antisimmetrica (cioè tale che $A^t = -A$) e ogni matrice ortogonale (cioè tale che $A^t = A^{-1}$) è normale. Matrici antisimmetriche o ortogonali però non hanno in genere autovalori reali, si può però dimostrare che, se con $\varphi := \bigcirc Ax$ definiamo $\mathcal{R}_\varphi(V)$

come nella definizione 21.2 (rispetto al prodotto scalare comune), $\mathcal{R}_\varphi(V)$ coincide anche in questo caso con l'involuppo convesso dell'insieme degli autovalori di A . Cfr. Storer/Bulirsch, 85.

Il baricentro

Lemma 21.10. f_1, \dots, f_n siano punti in \mathbb{R}^m . Per $c \in \mathbb{R}^m$ sia

$$F(c) := \sum_{i=1}^n |f_i - c|^2$$

Allora F assume il suo minimo nel baricentro $\frac{f_1 + \dots + f_n}{n}$.

Dimostrazione. Con $c := (c_1, \dots, c_m)$ abbiamo

$$F(c_1, \dots, c_m) = \sum_{i=1}^n \sum_{k=1}^m (f_{ik} - c_k)^2$$

Per la determinazione del minimo consideriamo le equazioni

$$\frac{\partial F}{\partial c_j} = -2 \sum_{i=1}^n (f_{ij} - c_j) = 0$$

cioè $nc_j = \sum_{i=1}^n f_{ij}$ e ciò implica

$$c = \frac{1}{n} (f_1 + \dots + f_n)$$

Si come $\frac{\partial F}{\partial c_j \partial c_k} = 2\delta_{jk}$ per ogni j, k , vediamo che si tratta veramente di un minimo.

Useremo questo lemma per dimostrare che ogni retta di regressione ortogonale di una matrice di dati X passa per il baricentro \bar{X} delle righe di X .

Spazi ortogonali intermedi

Proposizione 22.1. Per ogni $v \in V$ si ha

$$\lambda_m \|v, v\| \leq \|v, \varphi v\| \leq \lambda_1 \|v, v\|$$

Dimostrazione. Ciò per $v = 0$ è banale e segue per $v \neq 0$ dal teorema 21.8.

Lemma 22.2. e_1, \dots, e_m sia una base ortonormale di V tale che

$\varphi e_1 = \lambda_1 e_1, \dots, \varphi e_m = \lambda_m e_m$. Sappiamo dalla nota 20.16 che una tale base esiste. Per $1 \leq r \leq s \leq m$ poniamo

$$E_{r,s} := \mathbb{R}e_r + \dots + \mathbb{R}e_s$$

Allora

$$\mathcal{R}_\varphi(E_{r,s}) = [\lambda_r, \lambda_s]$$

e quindi

$$\begin{aligned} \max \mathcal{R}_\varphi(E_{r,s}) &= \lambda_r = \mathcal{R}_\varphi(e_r) \\ \min \mathcal{R}_\varphi(E_{r,s}) &= \lambda_s = \mathcal{R}_\varphi(e_s) \end{aligned}$$

Dimostrazione. È chiaro che $E_{r,s}$ è φ -invariante e che la matrice di

$$\psi := \bigcirc_{\varphi} v : E_{r,s} \rightarrow E_{r,s}$$

rispetto alla base e_r, \dots, e_s è

$$\begin{pmatrix} \lambda_r & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \lambda_s \end{pmatrix}$$

Ciò mostra che gli autovalori di ψ sono $\lambda_r, \dots, \lambda_s$. ψ è simmetrica per l'osservazione 20.13 e soddisfa le condizioni della situazione 21.1; possiamo quindi applicare il teorema 21.8 a ψ .

Siccome $\lambda_r = \max(\lambda_r, \dots, \lambda_s)$ e $\lambda_s = \min(\lambda_r, \dots, \lambda_s)$, otteniamo l'enunciato, tenendo conto dell'uguaglianza $\lambda_i = \mathcal{R}_\varphi(e_i)$ per ogni i che segue dall'osservazione 21.3.

Corollario 22.3. e_1, \dots, e_m sia una base ortonormale di V tale che

$\varphi e_1 = \lambda_1 e_1, \dots, \varphi e_m = \lambda_m e_m$. Usiamo le notazioni del lemma 22.2.

(1) Sia $1 \leq k \leq m$. Allora

$$\begin{aligned} \max \mathcal{R}_\varphi(E_{k,m}) &= \min \mathcal{R}_\varphi(E_{1,k}) \\ &= \lambda_k = \mathcal{R}_\varphi(e_k) \end{aligned}$$

(2) Sia $1 < k \leq m$. Allora

$$\begin{aligned} \max \mathcal{R}_\varphi(E_{1,k-1}^\perp) &= \max \mathcal{R}_\varphi(E_{k,m}) \\ &= \lambda_k = \mathcal{R}_\varphi(e_k) \end{aligned}$$

(3) Sia $1 \leq k < m$. Allora

$$\begin{aligned} \min \mathcal{R}_\varphi(E_{k+1,m}^\perp) &= \min \mathcal{R}_\varphi(E_{1,k}) \\ &= \lambda_k = \mathcal{R}_\varphi(e_k) \end{aligned}$$

Osservazione 22.4. X ed Y siano sottosinsiemi di V e $v \in X \cap Y$ con $v \neq 0$. Allora

$$\min \mathcal{R}_\varphi(X) \leq \mathcal{R}_\varphi(v) \leq \max \mathcal{R}_\varphi(Y)$$

Lemma 22.5. W_1 e W_2 siano sottospazi vettoriali di V . Allora

$$m + \dim(W_1 \cap W_2) \geq \dim W_1 + \dim W_2$$

Dimostrazione. Per il lemma 19.11

$$\begin{aligned} \dim W_1 + \dim W_2 &= \dim(W_1 + W_2) + \dim(W_1 \cap W_2) \\ &\leq m + \dim(W_1 \cap W_2) \end{aligned}$$

Lemma 22.6. Sia $1 \leq k < m$. W sia un sottospazio vettoriale di V con $\dim W \leq k$. Allora

$$\max \mathcal{R}_\varphi(W^\perp) \geq \lambda_{k+1}$$

Dimostrazione. Scegliamo di nuovo una base ortonormale e_1, \dots, e_m di V tale che $\varphi e_1 = \lambda_1 e_1, \dots, \varphi e_m = \lambda_m e_m$. Usiamo la notazione del lemma 22.2.

Per la proposizione 19.12 abbiamo $V = W \oplus W^\perp$ e quindi

$$\dim W^\perp = m - \dim W$$

Per il lemma 22.5 abbiamo

$$\begin{aligned} m + \dim(W^\perp \cap E_{1,k+1}) &\geq \dim W^\perp + \dim E_{1,k+1} \\ &= m - \dim W + k + 1 \\ &\geq m - k + k + 1 = m + 1 \end{aligned}$$

Ciò implica

$$\dim(W^\perp \cap E_{1,k+1}) \geq 1$$

per cui $W^\perp \cap E_{1,k+1} \neq \emptyset$. Esiste quindi un vettore $v \in W^\perp \cap E_{1,k+1}$ con $v \neq 0$. Con il lemma 22.2 e usando l'osservazione 22.4 segue adesso

$$\begin{aligned} \lambda_{k+1} &= \min \mathcal{R}_\varphi(E_{1,k+1}) \\ &\leq \mathcal{R}_\varphi(v) \leq \max \mathcal{R}_\varphi(W^\perp) \end{aligned}$$

Teorema 22.7 (teorema di Courant). Sia $1 \leq k < m$. Sia \mathcal{U} l'insieme dei sottospazi vettoriali U di V con $\dim U \geq m - k$. Allora

$$\lambda_{k+1} = \min\{\max \mathcal{R}_\varphi(U) \mid U \in \mathcal{U}\}$$

Dimostrazione. Sia \mathcal{U}' l'insieme dei sottospazi vettoriali W di V con $\dim W \leq k$. Per il corollario 22.3 e il lemma 22.6 allora

$$\lambda_{k+1} = \min\{\max \mathcal{R}_\varphi(W^\perp) \mid W \in \mathcal{U}'\}$$

Ma

$$\dim W \leq k \iff \dim W^\perp \geq m - k$$

L'enunciato segue dalla proposizione 19.12.

Calcolo matriciale

Useremo da ora in avanti, come d'uso nell'algebra multilineare e in geometria differenziale, indici in alto per gli indici di riga di una matrice, indici in basso per gli indici di colonna. Quando necessario applichiamo questa notazione anche alla nostra matrice dei dati, cosicché a pagina 15 abbiamo

$$\begin{aligned} X^i &= x_i \\ X_j^i &= x_{ij} \end{aligned}$$

Raccogliamo nei prossimi punti alcune formule del calcolo matriciale nella nuova notazione.

Nota 22.8. Per $A \in \mathbb{R}_m^n$ e $B \in \mathbb{R}_s^m$ abbiamo $AB \in \mathbb{R}_s^n$ con

$$(AB)_j^i = \sum_{\alpha=1}^m A_\alpha^i B_j^\alpha = A^i B_j$$

per ogni i, j . Si noti che il prodotto matriciale fornisce un'applicazione

$$\mathbb{R}_m^n \times \mathbb{R}_s^m \rightarrow \mathbb{R}_s^n$$

Corollario 22.9. Per $A \in \mathbb{R}_n^n$ e $v \in \mathbb{R}^n$ si ha $Av \in \mathbb{R}^n$ con

$$(Av)^i = \sum_{\alpha=1}^n A_\alpha^i v^\alpha = A^i v$$

per ogni i .

Corollario 22.10. Per $f \in \mathbb{R}_m$ e $B \in \mathbb{R}_s^m$ si ha $fB \in \mathbb{R}_s$ con

$$(fB)_j = \sum_{\alpha=1}^m f_\alpha B_j^\alpha = f B_j$$

per ogni j .

Corollario 22.11. Per $A \in \mathbb{A}_m^n$ e $B \in \mathbb{R}_s^m$ si hanno

$$\begin{aligned} (AB)^i &= A^i B = \sum_{\alpha=1}^m A_\alpha^i B^\alpha \\ (AB)_j &= AB_j = \sum_{\alpha=1}^m A_\alpha B_j^\alpha \end{aligned}$$

Dimostrazione. Per il corollario 22.10 abbiamo

$$(A^i B)_j = A^i B_j = (AB)_j^i$$

e per il corollario 22.9

$$(AB_j)^i = A^i B_j = (AB)_j^i$$

per ogni i, j .

Corollario 22.12. Siano $v \in \mathbb{R}^n$ ed $f \in \mathbb{R}_s$. Allora $vf \in \mathbb{R}_s^m$ con

$$(vf)_j^i = v^i f_j \quad \text{per ogni } i, j.$$

Dimostrazione. Ciò è un caso speciale della nota 22.8.

Proposizione 22.13. Siano $A \in \mathbb{R}_m^n$ e $B \in \mathbb{R}_s^m$. Allora

$$AB = \sum_{\alpha=1}^m A_\alpha B^\alpha$$

Dimostrazione. Per il corollario 22.12 abbiamo

$$\begin{aligned} \left(\sum_{\alpha=1}^m A_\alpha B^\alpha\right)_j^i &= \sum_{\alpha=1}^m (A_\alpha B^\alpha)_j^i \\ &= \sum_{\alpha=1}^m A_\alpha^i B_j^\alpha = (AB)_j^i \end{aligned}$$

Corollario 22.14. Siano $A \in \mathbb{R}_m^n$ e $v \in \mathbb{R}^m$. Allora

$$Av = \sum_{\alpha=1}^m A_\alpha v^\alpha$$

Nota 22.15. Siano $A \in \mathbb{R}_m^n$ ed $f \in \mathbb{R}_m$. Allora

$$Af^t = \sum_{\alpha=1}^m A_\alpha f_\alpha$$

Af^t è quindi una combinazione lineare di A_1, \dots, A_m con i coefficienti f_1, \dots, f_m . Useremo questa osservazione fra poco per $A = X^{CE}$.

Osservazione 22.16. Sia $A \in \mathbb{R}_m^n$. Allora

$$\begin{aligned} (A^i)^t &= (A^t)_i \\ (A_j)^t &= (A^t)^j \end{aligned}$$

per ogni i, j .

Formule per il prodotto scalare

Nota 23.1. Con $\| \cdot \|$ denotiamo il prodotto scalare comune sia in \mathbb{R}^n che in \mathbb{R}_m .

Per $u, v \in \mathbb{R}^n$ abbiamo quindi

$$\|u, v\| = u^t v = \sum_{\alpha=1}^m u^\alpha v^\alpha$$

per $f, g \in \mathbb{R}_m$ abbiamo

$$\|f, g\| = f g^t = \sum_{\alpha=1}^m f_\alpha g_\alpha$$

Osservazione 23.2. Per $A \in \mathbb{R}_n^n$ e $v, w \in \mathbb{R}^n$ vale

$$v^t A w = \|v, A w\| = \|A^t v, w\|$$

Per $B \in \mathbb{R}_m^m$ ed $f, g \in \mathbb{R}_m$ vale

$$f B g^t = \|f B, g\| = \|f, g B^t\|$$

Dimostrazione.

$$\begin{aligned} \|v, A w\| &= v^t A w \\ &= (A^t v)^t w = \|A^t v, w\| \end{aligned}$$

$$\begin{aligned} \|f B, g\| &= f B g^t \\ &= f (g B^t)^t = \|f, g B^t\| \end{aligned}$$

Corollario 23.3. Siano $A \in \mathbb{R}_p^n, B \in \mathbb{R}_s^n$. Allora

$$(A^t B)_j^i = \|A_i, B_j\|$$

per ogni i, j .

Dimostrazione. Usando la nota 22.8 e l'osservazione 22.16 abbiamo

$$\begin{aligned} (A^t B)_j^i &= (A^t)^i B_j \\ &= (A_i)^t B_j = \|A_i, B_j\| \end{aligned}$$

Lemma 23.4. Per $A \in \mathbb{R}_p^n$ e $v \in \mathbb{R}^n$ abbiamo

$$\|A_j, v\| = (A^t v)^j$$

per ogni j .

Dimostrazione. Dal corollario 22.11 e dall'osservazione 22.16 segue

$$\begin{aligned} \|A_j, v\| &= (A_j)^t v \\ &= (A^t)^j v = (A^t v)^j \end{aligned}$$

Proposizione 23.5. Siano $A \in \mathbb{R}_p^n$ e $v \in \mathbb{R}^n$. Allora

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

Dimostrazione. Usando il lemma 23.4 e l'osservazione 23.2 abbiamo

$$\begin{aligned} \sum_{j=1}^p \|A_j, v\|^2 &= \sum_{j=1}^p ((A^t v)^j)^2 \\ &= \|A^t v, A^t v\| = \|v, A A^t v\| \end{aligned}$$

Lemma 23.6. Per $f \in \mathbb{R}_m$ e $B \in \mathbb{R}_m^m$ abbiamo

$$\|f, B^i\| = (f B^t)_i$$

per ogni i .

Dimostrazione. Dal corollario 22.11 e dall'osservazione 22.16 segue

$$\|f, B^i\| = f (B^i)^t = f (B^t)_i = (f B^t)_i$$

Proposizione 23.7. Siano $f \in \mathbb{R}_m$ e $B \in \mathbb{R}_m^m$. Allora

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

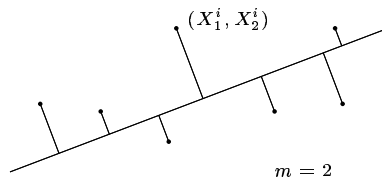
Dimostrazione. Usando il lemma 23.6 e l'osservazione 23.2 abbiamo

$$\begin{aligned} \sum_{i=1}^n \|f, B^i\|^2 &= \sum_{i=1}^n ((f B^t)_i)^2 \\ &= \|f B^t, f B^t\| = \|f B^t B, f\| \end{aligned}$$

Regressione ortogonale

Situazione 23.8. Sia di nuovo $X \in \mathbb{R}_m^n$ la nostra matrice di dati con $n \geq 2$.

Nota 23.9. Nella regressione ortogonale cerchiamo una retta in \mathbb{R}_m tale da minimizzare la somma dei quadrati delle distanze dei punti X^i da questa retta; cfr. pagina 9. Ogni retta con questa proprietà si chiama una retta di regressione ortogonale di X .



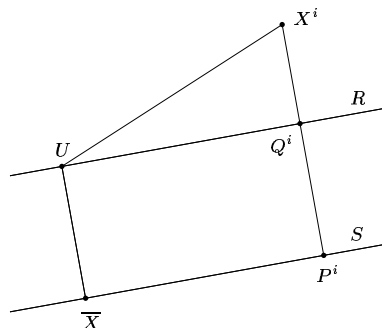
Osservazione 23.10. Esiste almeno una retta di regressione ortogonale di X .

Dimostrazione. Infatti la funzione da minimizzare è ovviamente continua ed è evidente che ci si può limitare a un insieme compatto di parametri da variare.

Lemma 23.11. Ogni retta di regressione ortogonale nella situazione 23.8 passa per il baricentro $\bar{X} = (X_1, \dots, X_m)$ dei punti X^i .

Dimostrazione. R sia una retta di regressione ortogonale. U sia la proiezione ortogonale di \bar{X} su R . Per ogni i sia Q^i la proiezione ortogonale di X^i su R .

Siano S la retta parallela ad R passante per \bar{X} e P^i la proiezione ortogonale di X^i su S .



Allora $Q^i - U = P^i - \bar{X}$ per ogni i . Dal teorema di Pitagora abbiamo

$$|X^i - Q^i|^2 = |X^i - U|^2 - |Q^i - U|^2$$

e

$$\begin{aligned} |X^i - P^i|^2 &= |X^i - \bar{X}|^2 - |P^i - \bar{X}|^2 \\ &= |X^i - \bar{X}|^2 - |Q^i - U|^2 \end{aligned}$$

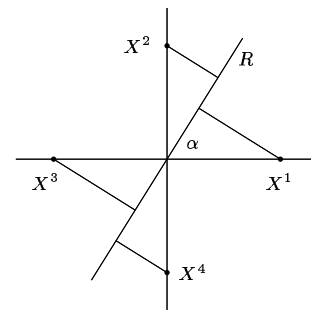
Dal lemma 21.10 segue adesso

$$\begin{aligned} \sum_{i=1}^n |X^i - P^i|^2 &= \sum_{i=1}^n |X^i - \bar{X}|^2 - \sum_{i=1}^n |Q^i - U|^2 \\ &\leq \sum_{i=1}^n |X^i - U|^2 - \sum_{i=1}^n |Q^i - U|^2 \\ &= \sum_{i=1}^n |X^i - Q^i|^2 \end{aligned}$$

Nota 23.12. La retta di regressione ortogonale in genere non è univocamente determinata. Consideriamo ad esempio i quattro punti

- $X^1 = (1, 0)$
- $X^2 = (0, 1)$
- $X^3 = (-1, 0)$
- $X^4 = (0, -1)$

nel piano \mathbb{R}_2 . Per il lemma 23.11 ogni retta di regressione ortogonale R passa per il baricentro che in questo caso coincide con l'origine. Come nella figura sia α l'angolo tra R e l'ascisse.



Dal disegno si vede che la somma dei quadrati delle distanze dei punti X^i dalla retta R è uguale a $2(\sin^2 \alpha + \cos^2 \alpha) = 2$, indipendentemente da α .

Nell'esempio della nota 23.12 entrambe le colonne di X hanno media 0 e ciò implica $MX = 0$. La matrice X^{COM} definita a pagina 24 è quindi, per il lemma 24.3, uguale a

$$X^t X = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2\delta$$

Ciò significa che ogni vettore $\neq 0$ di \mathbb{R}_2 è un autovettore di X^{COM} per l'unico autovalore 2 e quindi possiamo scegliere una qualsiasi base ortonormale e^1, e^2 di \mathbb{R}_2 come assi principali e quindi anche le componenti principali non sono univocamente determinate.

La formula di proiezione

Nota 24.1. Per il lemma 23.11 ogni retta di regressione ortogonale di X passa per il baricentro \bar{X} dei punti X^i e può quindi essere scritta nella forma

$$R_g := \bar{X} + \mathbb{R}g$$

con $g \in \mathbb{R}_m$ e $|g| = 1$. Per la proposizione 8.2 la proiezione ortogonale P_g^i di X^i su R_g è uguale a

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

Come nel lemma 23.11 dal teorema di Pitagora abbiamo

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

Ma

$$|P_g^i - \bar{X}|^2 = \|X^i - \bar{X}, g\|^2$$

cosicché

$$\begin{aligned} \sum_{i=1}^n |X^i - P_g^i|^2 &= \sum_{i=1}^n |X^i - \bar{X}|^2 - \sum_{i=1}^n \|X^i - \bar{X}, g\|^2 \end{aligned}$$

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

La matrice di covarianza

Definizione 24.2. La matrice

$$X^{COM} := (X^{CE})^t X^{CE}$$

si chiama la *matrice di covarianza maggiore* di X (termine non standard), la matrice

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

la *matrice di covarianza* di X . Entrambe le matrici appartengono ad \mathbb{R}_m^m e sono simmetriche.

Possiamo definire un operatore

$$\begin{aligned} \varphi_X : \mathbb{R}_m &\rightarrow \mathbb{R}_m \\ f &\mapsto fX^{COM} \end{aligned}$$

evidentemente simmetrico.

Lemma 24.3. $X^{COM} = X^t X - X^t M X$.

Dimostrazione. Sappiamo dal corollario 15.11 che $M^2 = M$. Inoltre chiaramente $M^t = M$, per cui anche $M^t M = M$. Perciò

$$\begin{aligned} X^{COM} &= (X - M X)^t (X - M X) \\ &= (X^t - X^t M^t)(X - M X) \\ &= X^t X - X^t M X - X^t M^t X + X^t M^t M X \\ &= X^t X - X^t M X - X^t M X + X^t M X \\ &= X^t X - X^t M X \end{aligned}$$

Osservazione 24.4.

$$\begin{aligned} (X^{COM})_j^i &= \|X_i^{CE}, X_j^{CE}\| \\ \text{cov}(X)_j^i &= s_{X_i X_j} \end{aligned}$$

Dimostrazione. Per il corollario 23.3 abbiamo

$$\begin{aligned} (X^{COM})_j^i &= ((X^{CE})^t X^{CE})_j^i \\ &= \|X_i^{CE}, X_j^{CE}\| \end{aligned}$$

Dividendo per $n - 1$ otteniamo la seconda equazione.

Osservazione 24.5. $(X^{CE})^i = X^i - \bar{X}$ per ogni i .

Nota 24.6. Sia $g \in \mathbb{R}_m$ con $|g| = 1$. Per la proposizione 23.7 abbiamo

$$\begin{aligned} \sum_{i=1}^n \|X^i - \bar{X}, g\|^2 &= \|g X^{COM}, g\| \\ &= \mathcal{R}_{\varphi_X}(g) \end{aligned}$$

Come nella nota 24.1 sia P_g^i la proiezione ortogonale di X^i sulla retta $\bar{X} + \mathbb{R}g$. Da quella stessa nota segue allora

$$\begin{aligned} \min\{\sum_{i=1}^n |X^i - P_g^i|^2 \mid g \in \mathbb{R}_m \text{ con } |g| = 1\} \\ = \max\{\mathcal{R}_{\varphi_X}(g) \mid g \in \mathbb{R}_m \text{ con } |g| = 1\} \end{aligned}$$

Nota 24.7. $\lambda_1, \dots, \lambda_m$ siano gli autovalori (necessariamente reali) di X^{COM} e $\lambda_1 \geq \dots \geq \lambda_m$. Per la nota 20.16 esiste una base ortonormale e^1, \dots, e^m di \mathbb{R}_m tale che

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

Per la nota 24.6 $\bar{X} + \mathbb{R}e^1$ è una retta di regressione ortogonale di X .

Definizione 24.8. Per $g \in \mathbb{R}_m$ sia

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

per ogni $i = 1, \dots, n$. Otteniamo in questo modo un vettore colonna $X_g \in \mathbb{R}^n$.

Osservazione 24.9. Sia $g \in \mathbb{R}_m$. Allora

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

Dimostrazione. Per ogni i abbiamo

$$\begin{aligned} (X^{CE} g^t)^i &= (X^{CE})^i g^t = (X^i - \bar{X})g^t \\ &= \|X^i - \bar{X}, g\| = X_g^i \end{aligned}$$

Osservazione 24.10. $M X^{CE} = 0$.

Dimostrazione.

$$\begin{aligned} M X^{CE} &= M(X - M X) = M X - M^2 X \\ &= M X - M X = 0 \end{aligned}$$

per il corollario 15.11.

Corollario 24.11. Sia $g \in \mathbb{R}_m$. Allora

$$\bar{X}_g = 0$$

Dimostrazione. Applicando il corollario 15.6 ad X_g abbiamo $M X_g = \bar{X}_g 1^n$. È quindi sufficiente dimostrare che $M X_g = 0$.

Ma $M X_g = M X^{CE} g^t = 0$ per l'osservazione 24.10.

Una dimostrazione diretta è altrettanto facile:

$$\begin{aligned} n \bar{X}_g &= \sum_{i=1}^n \|X^i - \bar{X}, g\| \\ &= \sum_{i=1}^n (X^i - \bar{X}), g \end{aligned}$$

Ma $\sum_{i=1}^n (X^i - \bar{X}) = 0$.

Corollario 24.12. Sia $g \in \mathbb{R}_m$ con $|g| = 1$. Allora

$$\begin{aligned} |X_g|^2 &= \mathcal{R}_{\varphi_X}(g) \\ s_{X_g}^2 &= \frac{1}{n-1} \mathcal{R}_{\varphi_X}(g) \end{aligned}$$

Dimostrazione.

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

come sappiamo dalla nota 24.6.

La seconda equazione segue adesso dal corollario 24.11.

Nota 24.13. Nelle ipotesi della nota 24.7 per $1 \leq k \leq m$ sia \mathcal{G}_k l'insieme dei vettori $g \in \mathbb{R}_m$ con $|g| = 1$ e, per $k > 1$, $g \perp e_1, \dots, e_{k-1}$.

Dal corollario 24.12 segue che

$$s_{X_{e^1}}^2 = \max\{s_{X_g}^2 \mid g \in \mathbb{R}_m \text{ e } |g| = 1\} = \lambda_1$$

In questo senso con e^1 abbiamo scoperto una direzione di massima varianza (non univocamente determinata come sappiamo dall'osservazione 23.12).

Dal teorema 22.3 vediamo anche che $\bar{X} + \mathbb{R}e^2$ è una retta che minimizza la somma $\sum_{i=1}^n |X^i - P_g^i|^2$ per $g \in \mathcal{G}_2$, e che

$$s_{X_{e^2}}^2 = \max\{s_{X_g}^2 \mid g \in \mathcal{G}_2\} = \lambda_2$$

In generale, per $1 \leq k \leq m$, possiamo concludere che $\bar{X} + \mathbb{R}e^k$ è una retta che minimizza $\sum_{i=1}^n |X^i - P_g^i|^2$ per $g \in \mathcal{G}_k$ e che

$$s_{X_{e^k}}^2 = \max\{s_{X_g}^2 \mid g \in \mathcal{G}_k\} = \lambda_k$$

I vettori X_{e^1}, \dots, X_{e^m} sono detti *componenti principali* di X ; non sono comunque univocamente determinati, come si vede dalla nota 23.12.

Bibliografia

6023 **D. Bini/M. Capovani/O. Menchi:** Metodi numerici per l'algebra lineare. Zanichelli 1988.

12323 **T. Lehmann/W. Oberschelp/E. Pelikan/R. Reppes:** Bildverarbeitung für die Medizin. Springer 1997.

16226 **F. Paset:** Regressione, correlazione e analisi delle componenti principali. Tesi Univ. Ferrara 2003.

1003 **J. Stoer/R. Bulirsch:** Einführung in die numerische Mathematik II. Springer 1978.