

Spazio delle variabili e delle osservazioni

Analisi Esplorativa

Aldo Solari



① Spazio delle variabili

② Spazio delle osservazioni

③ Appendice: vettori



Interpretazione geometrica

Spazio delle variabili

- del vettore delle medie (trasposto) $\bar{x}' = [\bar{x}_1, \dots, \bar{x}_j, \dots, \bar{x}_p]$ come
baricentro delle n unità statistiche $x'_i = u'_i, i = 1, \dots, n,$
interpretate come n punti p -dimensionali

Spazio delle osservazioni

- della devianza $ns_{jj}, j = 1, \dots, p$ come quadrato della lunghezza
del vettore \tilde{x}_j scarto dalla media, ovvero $\|\tilde{x}_j\|^2$
 $n \times 1$ $n \times 1$
- della covarianza $ns_{jk}, j \neq k = 1, \dots, p$ come prodotto $\tilde{x}'_j \tilde{x}_k$
 $1 \times n$ $n \times 1$
- della correlazione $r_{jk}, j \neq k = 1, \dots, p$ come coseno dell'angolo
formato dai vettori \tilde{x}_j e \tilde{x}_k
 $n \times 1$ $n \times 1$

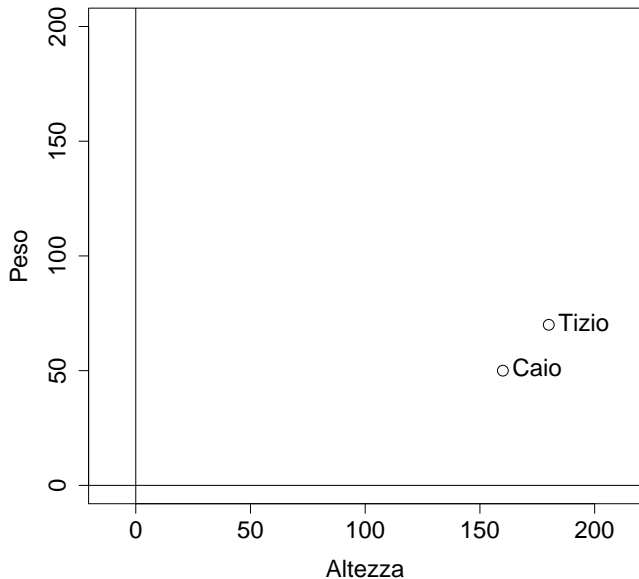


Esempio

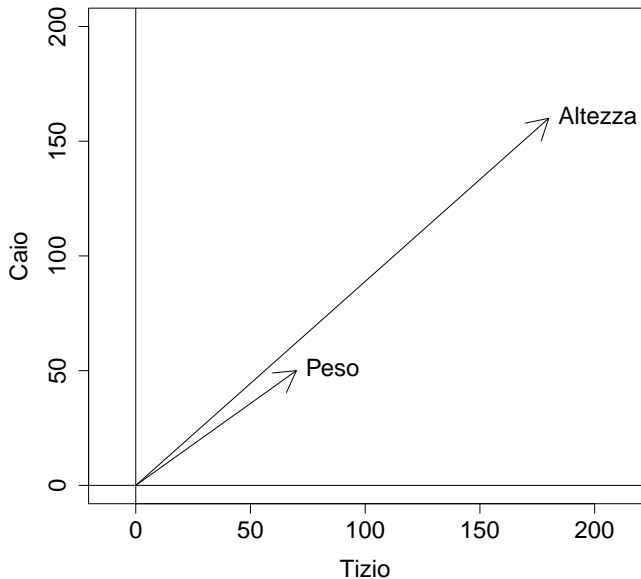
	Altezza	Peso
Tizio	180	70
Caio	160	50



Spazio delle variabili



Spazio delle osservazioni



Matrice X

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}$$



Outline

- ① Spazio delle variabili
- ② Spazio delle osservazioni
- ③ Appendice: vettori



Spazio delle variabili: n punti p -dimensionali

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} x'_1 \\ x'_2 \\ \cdots \\ x'_i \\ \cdots \\ x'_n \end{bmatrix} = \begin{bmatrix} u'_1 \\ u'_2 \\ \cdots \\ u'_i \\ \cdots \\ u'_n \end{bmatrix}$$

dove $x'_i = u'_i = [x_{i1} \cdots x_{ij} \cdots x_{in}]$ è l' i -simo vettore riga
 $1 \times p \quad 1 \times p$

L' i -sima riga di X contiene il profilo dell' i -sima unità statistica



Vettore delle medie $\bar{x}_{p \times 1}$

Vettore delle medie:

$$\bar{x}_{p \times 1} = \begin{bmatrix} \bar{x}_1 \\ \dots \\ \bar{x}_j \\ \dots \\ \bar{x}_p \end{bmatrix}$$

Il vettore delle medie trasposto

$$\bar{x}'_{1 \times p} = [\bar{x}_1 \cdots \bar{x}_j \cdots \bar{x}_p]$$

può essere interpretato come il baricentro di n punti p -dimensionali



3 punti 2-dimensionali e baricentro

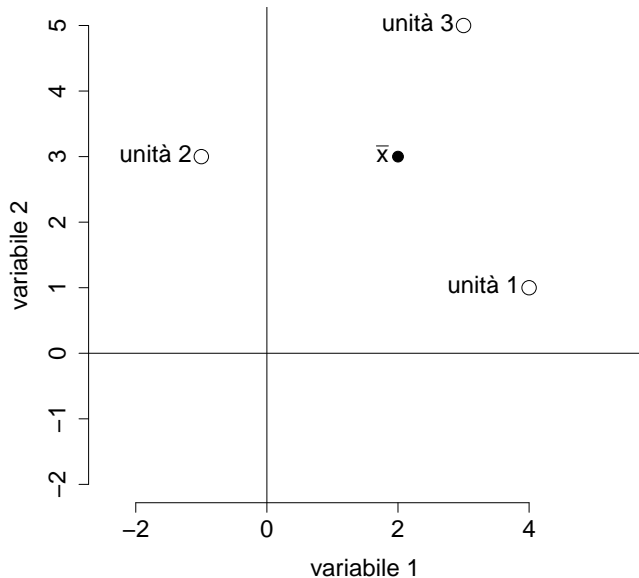
	variabile 1	variabile 2
unità 1	4	1
unità 2	-1	3
unità 3	3	5

$$X_{3 \times 2} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\bar{x}'_{1 \times 2} = \begin{bmatrix} 2 & 3 \end{bmatrix}$$



3 punti 2-dimensionali e baricentro



Outline

- ① Spazio delle variabili
- ② Spazio delle osservazioni
- ③ Appendice: vettori



Spazio delle osservazioni: p vettori n -dimensionali

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_j & \cdots & x_p \\ n \times 1 & n \times 1 & & n \times 1 & & n \times 1 \end{bmatrix}$$

dove $x_j = \begin{bmatrix} x_{1j} \\ \cdots \\ x_{ij} \\ \cdots \\ x_{nj} \end{bmatrix}_{n \times 1}$ è il j -simo vettore colonna



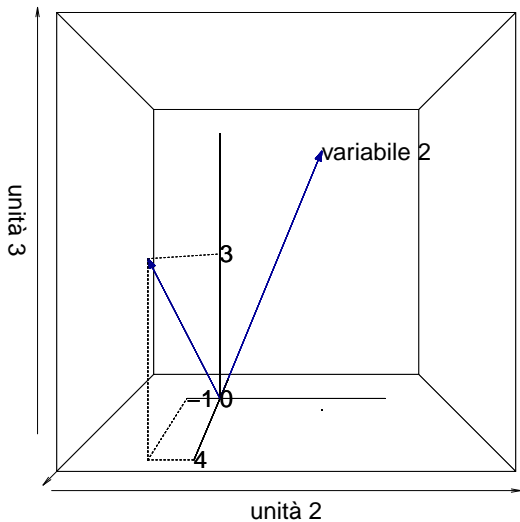
2 vettori 3-dimensionali

	variabile 1	variabile 2
unità 1	4	1
unità 2	-1	3
unità 3	3	5

$$X_{3 \times 2} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$



2 vettori 3-dimensionali



Vettore scarto dalla media

Consideriamo il vettore scarto dalla media \tilde{x}_j :

$$\tilde{x}_j = \begin{matrix} n \times 1 \end{matrix} \begin{bmatrix} x_{1j} - \bar{x}_j \\ \dots \\ x_{ij} - \bar{x}_j \\ \dots \\ x_{nj} - \bar{x}_j \end{bmatrix} = \begin{bmatrix} x_{1j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{bmatrix} - \begin{bmatrix} \bar{x}_j \\ \dots \\ \bar{x}_j \\ \dots \\ \bar{x}_j \end{bmatrix} = \begin{matrix} n \times 1 \end{matrix} x_j - \bar{x}_j \begin{matrix} 1 \\ n \times 1 \end{matrix}$$

dove $\begin{matrix} 1 \\ n \times 1 \end{matrix}$ è il vettore unitario



Vettore scarto dalla media

I vettori \tilde{x}_j e $\bar{x}_j \mathbf{1}$ sono perpendicolari

$$\bar{x}_j \mathbf{1}' \tilde{x}_j = \bar{x}_j \sum_{i=1}^n (x_{ij} - \bar{x}_j) = 0$$

Due vettori a e b sono perpendicolari se $a' b = 0$; vedi Appendice



$$X_{2 \times 2} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \end{bmatrix}$$

$$\tilde{x}_1 = x_1 - \bar{x}_1 \mathbf{1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix}$$

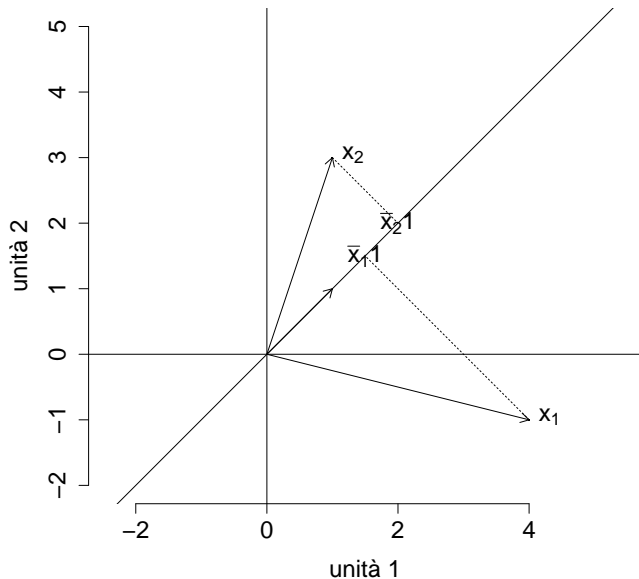
$$\tilde{x}_2 = x_2 - \bar{x}_2 \mathbf{1} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

\tilde{x}_1 e $\bar{x}_1 \mathbf{1}$ sono perpendicolari

$$\bar{x}_1 \mathbf{1}' \tilde{x}_1 = \begin{bmatrix} 1.5 & 1.5 \end{bmatrix} \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix} = 3.75 - 3.75 = 0$$



2 vettori 2-dimensionali



$$X_{3 \times 2} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\tilde{x}_1 = x_1 - \bar{x}_1 \mathbf{1}_{3 \times 1} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$$

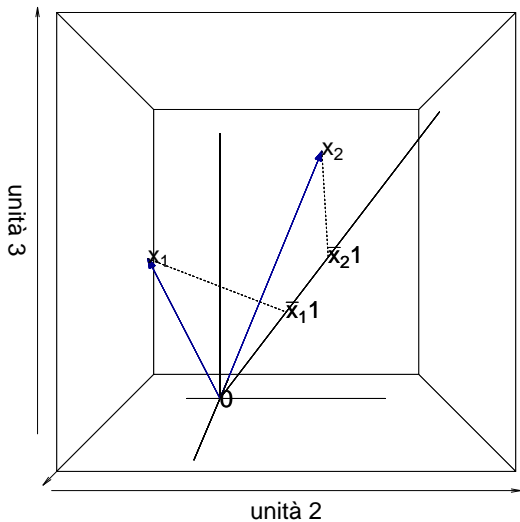
$$\tilde{x}_2 = x_2 - \bar{x}_2 \mathbf{1}_{3 \times 1} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

\tilde{x}_1 e $\bar{x}_1 \mathbf{1}_{3 \times 1}$ sono perpendicolari

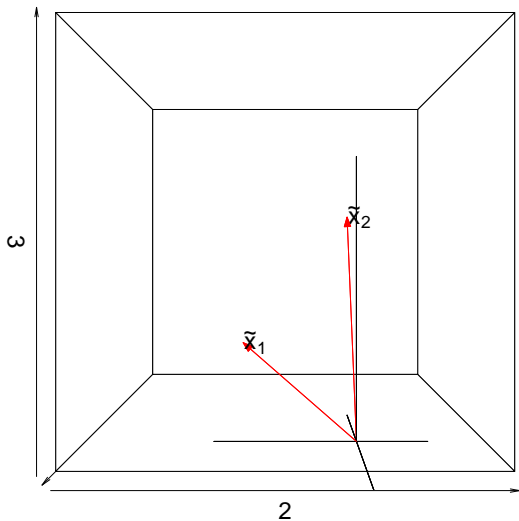
$$\bar{x}_1 \mathbf{1}'_{1 \times 3} \tilde{x}_1 = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} = 4 - 6 + 2 = 0$$



2 vettori 3-dimensionali



Vettori scarto dalla media



Devianza e codevianza

Il quadrato della lunghezza (norma o modulo) di \tilde{x}_j è la **devianza**

$$\| \tilde{x}_j \|_{n \times 1}^2 = \tilde{x}_j' \tilde{x}_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = ns_{jj}$$

Il prodotto (interno o scalare) di \tilde{x}_j e \tilde{x}_k è la **codevianza**

$$\tilde{x}_j' \tilde{x}_k = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = ns_{jk}$$

La lunghezza del vettore a è definita come $\| a \|_{n \times 1} = \sqrt{a' a}_{1 \times n n \times 1}$; vedi Appendice

Il prodotto dei vettori a e b è definito come $a' b_{1 \times n n \times 1} = \sum_{i=1}^n a_i b_i$; vedi Appendice



Correlazione

Abbiamo

$$\tilde{x}'_j \tilde{x}_k = \sqrt{\tilde{x}'_j \tilde{x}_j}_{1 \times n} \sqrt{\tilde{x}'_k \tilde{x}_k}_{n \times 1} \cos(\theta_{jk})$$

dove θ_{jk} è l'angolo formato dai due vettori \tilde{x}_j e \tilde{x}_k , quindi risulta che

$$\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2} \cos(\theta_{jk})$$

Il coseno dell'angolo θ_{ab} formato da due vettori a e b è definito da

$$\cos(\theta_{ab}) = \left(\begin{matrix} a' & b \end{matrix} \right)_{1 \times n} / \left(\sqrt{\begin{matrix} a' & a \end{matrix}}_{1 \times n} \sqrt{\begin{matrix} b' & b \end{matrix}}_{n \times 1} \right); \text{ vedi Appendice}$$



Correlazione

La correlazione r_{jk} è il coseno dell'angolo θ_{jk} formato dai due vettori \tilde{x}_j e \tilde{x}_k :

$\tilde{x}_k :$
 $n \times 1$

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} = \cos(\theta_{jk})$$



Esempio

$$X_{3 \times 2} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}, \quad \tilde{x}_1 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}_{3 \times 1}, \quad \tilde{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}_{3 \times 1} \quad \text{quindi}$$

$$\tilde{x}_1' \tilde{x}_1 = 14 = 3s_{11}$$

$$\tilde{x}_2' \tilde{x}_2 = 8 = 3s_{22}$$

$$\tilde{x}_1' \tilde{x}_2 = -2 = 3s_{12}$$

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} = -.189$$

$$\theta^\circ = \arccos(-.189) \frac{180^\circ}{\pi} = 100.89^\circ$$



Outline

- ① Spazio delle variabili
- ② Spazio delle osservazioni
- ③ Appendice: vettori



Prodotto di due vettori

Siano dati due vettori colonna $a_{n \times 1} = \begin{bmatrix} a_1 \\ \dots \\ a_i \\ \dots \\ a_n \end{bmatrix}$ e $b_{n \times 1} = \begin{bmatrix} b_1 \\ \dots \\ b_i \\ \dots \\ b_n \end{bmatrix}$.

Il prodotto (*inner product*) di $a_{n \times 1}$ e $b_{n \times 1}$ è definito come

$$a'_{1 \times n} b_{n \times 1} = \begin{bmatrix} a_1 & \dots & a_i & \dots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \dots \\ b_i \\ \dots \\ b_n \end{bmatrix} = a_1 b_1 + \dots + a_n b_n = \sum_{i=1}^n a_i b_i$$

Si noti che vale $a'_{1 \times n} b_{n \times 1} = b'_{1 \times n} a_{n \times 1}$.



Lunghezza di un vettore

La lunghezza (norma, modulo) di un vettore $a_{n \times 1} = \begin{bmatrix} a_1 \\ \dots \\ a_i \\ \dots \\ a_n \end{bmatrix}$ è definita

come

$$\|a_{n \times 1}\| = \sqrt{a_1^2 + \dots + a_n^2} = \sqrt{\sum_{i=1}^n a_i^2} = \sqrt{1 \times n \begin{matrix} a' & a \\ n \times 1 & 1 \times n \end{matrix}}$$



Lunghezza di un vettore

La lunghezza di $a_{2 \times 1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ è

$$\|a_{2 \times 1}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17} = 4.123106$$

La lunghezza di $a_{3 \times 1} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix}$ è

$$\|a_{3 \times 1}\| = \sqrt{4^2 + (-1)^2 + 3^2} = \sqrt{26} = 5.09902$$



Vettore unitario

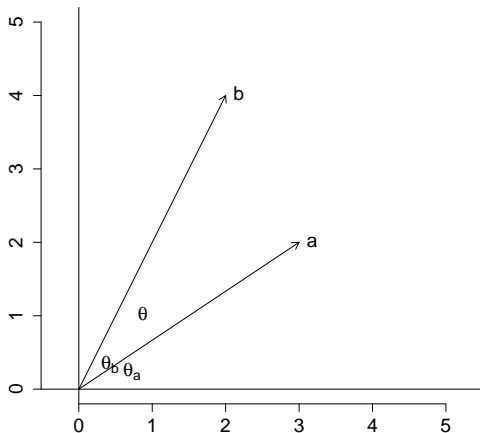
Il vettore unitario $\mathbf{1}_{n \times 1}$ è definito come

$$\mathbf{1}_{n \times 1} = \begin{bmatrix} 1 \\ \dots \\ 1 \\ \dots \\ 1 \end{bmatrix}$$



Angolo fra due vettori

Siano dati due vettori $a_{2 \times 1} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ e $b_{2 \times 1} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$



Angolo fra due vettori

Dalla figura, l'angolo θ può essere rappresentato come differenza tra gli angoli θ_b e θ_a . Per definizione

$$\cos(\theta_a) = \frac{a_1}{\|a\|_{2 \times 1}}, \quad \cos(\theta_b) = \frac{b_1}{\|b\|_{2 \times 1}}$$

$$\sin(\theta_a) = \frac{a_2}{\|a\|_{2 \times 1}}, \quad \sin(\theta_b) = \frac{b_2}{\|b\|_{2 \times 1}}$$

e ricordiamo la formula di sottrazione di archi:

$$\cos(\theta) = \cos(\theta_b - \theta_a) = \cos(\theta_b) \cos(\theta_a) + \sin(\theta_b) \sin(\theta_a)$$



Angolo fra due vettori

Di conseguenza l'angolo θ tra due vettori $\underset{2 \times 1}{a}$ e $\underset{2 \times 1}{b}$ è dato da

$$\begin{aligned}\cos(\theta) &= \cos(\theta_2 - \theta_1) \\ &= \frac{b_1}{\|\underset{2 \times 1}{b}\|} \frac{a_1}{\|\underset{2 \times 1}{a}\|} + \frac{b_2}{\|\underset{2 \times 1}{b}\|} \frac{a_2}{\|\underset{2 \times 1}{a}\|} \\ &= \frac{a_1 b_1 + a_2 b_2}{\|\underset{2 \times 1}{a}\| \|\underset{2 \times 1}{b}\|} \\ &= \frac{\underset{1 \times 2}{a'} \underset{2 \times 1}{b}}{\|\underset{2 \times 1}{a}\| \|\underset{2 \times 1}{b}\|} \\ &= \frac{\underset{1 \times 2}{a'} \underset{2 \times 1}{b}}{\sqrt{\underset{1 \times 2}{a'} \underset{2 \times 1}{a}} \sqrt{\underset{1 \times 2}{b'} \underset{2 \times 1}{b}}}\end{aligned}$$



Angolo fra due vettori

In generale, dati due vettori $a_{n \times 1}$ e $b_{n \times 1}$,

$$\cos(\theta) = \frac{a'_{1 \times n} b_{n \times 1}}{\|a_{n \times 1}\| \|b_{n \times 1}\|} = \frac{a'_{1 \times n} b_{n \times 1}}{\sqrt{a'_{1 \times n} a_{n \times 1}} \sqrt{b'_{1 \times n} b_{n \times 1}}}$$

Si noti che poichè

- $\cos(90^\circ) = \cos(270^\circ) = 0$
- $\cos(\theta) = 0$ solo se $a'_{1 \times n} b_{n \times 1} = 0$

i vettori $a_{n \times 1}$ e $b_{n \times 1}$ sono perpendicolari quando $a'_{1 \times n} b_{n \times 1} = 0$



Angolo fra due vettori

Il coseno dell'angolo tra $a_{2 \times 1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ e $b_{2 \times 1} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ è dato da

$$\cos \theta = \frac{4 \cdot 1 + (-1) \cdot 3}{\sqrt{4^2 + (-1)^2} \sqrt{1^2 + 3^2}} = \frac{1}{\sqrt{17} \sqrt{10}} = 0.0766965$$

quindi $\theta = \arccos(0.0766965) = 1.494024$ misurato in radianti.

In gradi, abbiamo $\theta = 1.494024 \cdot \frac{180^\circ}{\pi} = 85.60129^\circ$, ricordando che l'angolo in gradi è uguale all'angolo in radianti per $180^\circ/\pi$

