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) $\underset{1\times p}{\bar{x}'} = [\bar{x}_1, \ldots, \bar{x}_j, \ldots, \bar{x}_p]$ come baricentro delle n unità statistiche $\underset{1\times p}{x_i'} = \underset{1\times p}{u_i'}$, $i = 1, \ldots, n$, interpretate come n punti p-dimensionali

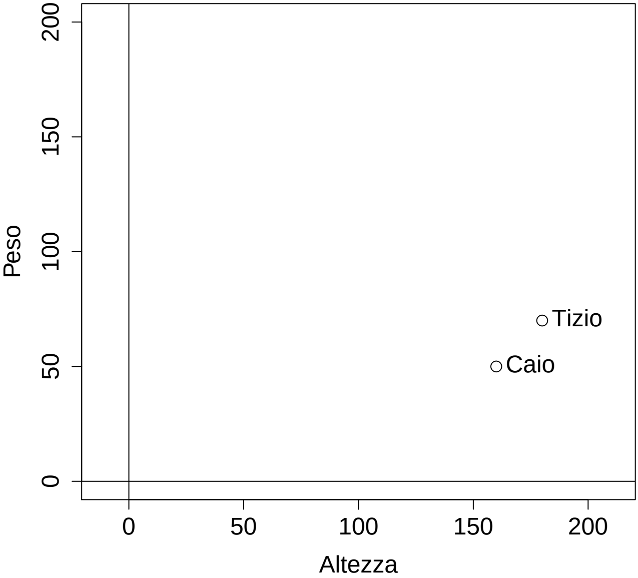
Spazio delle osservazioni

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

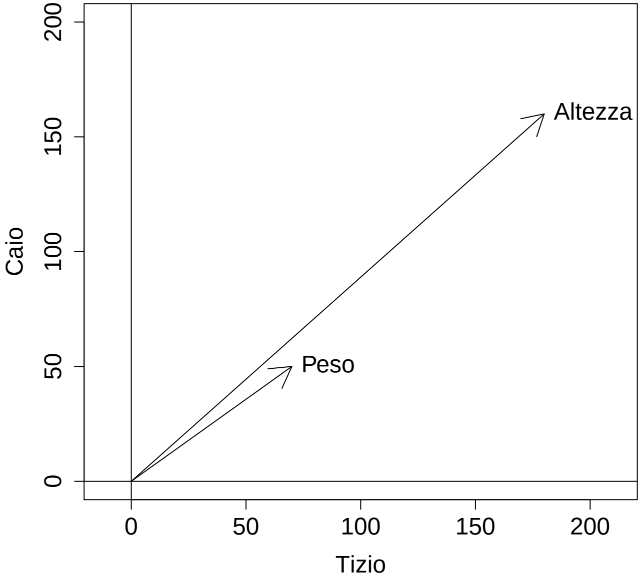
Esempio

	Altezza	Peso
Tizio	180	70
Caio	160	50

Spazio delle variabili



Spazio delle osservazioni



Matrice X

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

Outline

Spazio delle variabili: n punti p-dimensionali

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

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

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

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

Vettore delle medie:

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

Il vettore delle medie trasposto

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

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

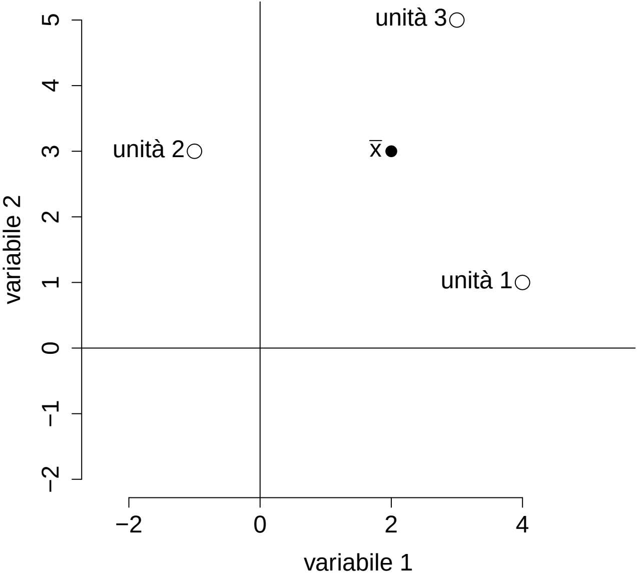
3 punti 2-dimensionali e baricentro

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

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

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

3 punti 2-dimensionali e baricentro



Outline

Spazio delle osservazioni: p vettori n-dimensionali

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

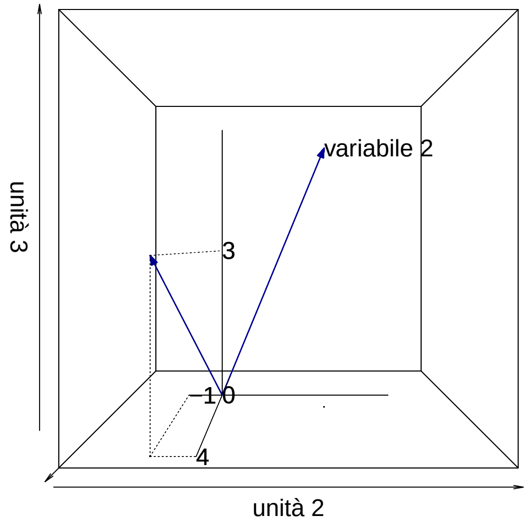
dove $\underset{n\times 1}{x_j} = \begin{bmatrix} x_{1j} \\ \cdots \\ x_{ij} \\ \cdots \\ x_{nj} \end{bmatrix}$ è il j-simo vettore colonna

2 vettori 3-dimensionali

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

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

2 vettori 3-dimensionali



Vettore scarto dalla media

Consideriamo il vettore scarto dalla media $\underset{n\times 1}{\tilde{x}_j}$:

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

dove $\underset{n\times 1}{1}$ è il vettore unitario

Vettore scarto dalla media

I vettori $\underset{n\times 1}{\tilde{x}_j}$ e $\bar{x}_j \underset{n\times 1}{1}$ sono perpendicolari

$$\langle \bar{x}_j \underset{n\times 1}{1}, \underset{n\times 1}{\tilde{x}_j} \rangle = \bar{x}_j \underset{1\times n}{1'} \underset{n\times 1}{\tilde{x}_j} = \bar{x}_j \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) = 0$$

Due vettori $\underset{n\times 1}{a}$ e $\underset{n\times 1}{b}$ sono perpendicolari se $\langle a, b\rangle = \underset{1\times n}{a'} \underset{n\times 1}{b} = 0$; vedi Appendice

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

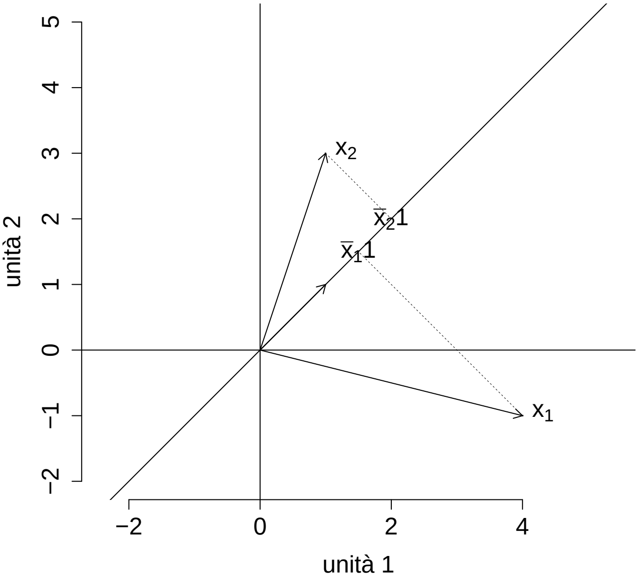
$$\underset{2\times 1}{\tilde{x}_1} = \underset{2\times 1}{x_1} - \bar{x}_1 \underset{2\times 1}{1} = \begin{bmatrix} 4 \\ -1 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix}$$

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

$\underset{2\times 1}{\tilde{x}_1}$ e $\bar{x}_1 \underset{2\times 1}{1}$ sono perpendicolari

$$\langle \bar{x}_1 \underset{1\times 2}{1}, \underset{2\times 1}{\tilde{x}_1} \rangle = \bar{x}_1 \underset{1\times 2}{1'} \underset{2\times 1}{\tilde{x}_1} = \begin{bmatrix} 1.5 & 1.5 \end{bmatrix} \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix} = 3.75 - 3.75 = 0$$

2 vettori 2-dimensionali



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

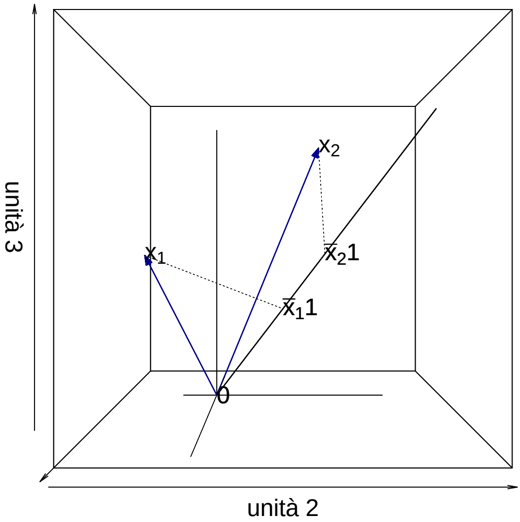
$$\underset{3\times 1}{\tilde{x}_1} = \underset{3\times 1}{x_1} - \bar{x}_1 \underset{3\times 1}{1} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$$

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

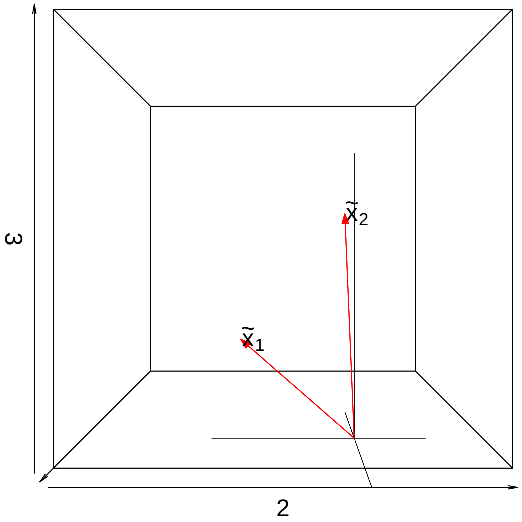
$\underset{3\times 1}{\tilde{x}_1}$ e $\bar{x}_1 \underset{3\times 1}{1}$ sono perpendicolari

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

2 vettori 3-dimensionali



Vettori scarto dalla media



Devianza e codevianza

Il quadrato della lunghezza di $\underset{n\times 1}{\tilde{x}_j}$ è la **devianza**

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

Il prodotto di $\underset{n\times 1}{\tilde{x}_j}$ e $\underset{n\times 1}{\tilde{x}_k}$ è la **codevianza**

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

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

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

Correlazione

Abbiamo

$$\langle \tilde{x}_j, \tilde{x}_k \rangle = \underset{1\times n}{\tilde{x}'_j} \underset{n\times 1}{\tilde{x}_k} = \sqrt{\underset{1\times n}{\tilde{x}'_j} \underset{n\times 1}{\tilde{x}_j}} \sqrt{\underset{1\times n}{\tilde{x}'_k} \underset{n\times 1}{\tilde{x}_k}} \cos(\theta_{jk})$$

dove θ_{jk} è l'angolo formato dai due vettori $\underset{n\times 1}{\tilde{x}_j}$ e $\underset{n\times 1}{\tilde{x}_k}$, 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 $\underset{n\times 1}{a}$ e $\underset{n\times 1}{b}$ è definito da

$\cos(\theta_{ab}) = \left(\underset{1\times n}{a'} \underset{n\times 1}{b} \right) / \left(\sqrt{\underset{1\times n}{a'} \underset{n\times 1}{a}} \sqrt{\underset{1\times n}{b'} \underset{n\times 1}{b}} \right)$; vedi Appendice

Correlazione

La correlazione r_{jk} è il coseno dell'angolo θ_{jk} formato dai due vettori $\underset{n\times 1}{\tilde{x}_j}$ e $\underset{n\times 1}{\tilde{x}_k}$:

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

Esempio

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

$$\|\tilde{x}_1\|^2 = \underset{1\times 3}{\tilde{x}_1'}\underset{3\times 1}{\tilde{x}_1} = 14 = 3s_{11}$$

$$\|\tilde{x}_2\|^2 = \underset{1\times 3}{\tilde{x}_2'}\underset{3\times 1}{\tilde{x}_2} = 8 = 3s_{22}$$

$$\langle \tilde{x}_1, \tilde{x}_2 \rangle = \underset{1\times 3}{\tilde{x}_1'}\underset{3\times 1}{\tilde{x}_2} = -2 = 3s_{12}$$

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

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

Outline

Prodotto di due vettori

Siano dati due vettori colonna $\underset{n\times 1}{a} = \left[\begin{array}{c} a_1 \\ \dots \\ a_i \\ \dots \\ a_n \end{array}\right]$ e $\underset{n\times 1}{b} = \left[\begin{array}{c} b_1 \\ \dots \\ b_i \\ \dots \\ b_n \end{array}\right]$.

Il prodotto (*inner product*) di $\underset{n\times 1}{a}$ e $\underset{n\times 1}{b}$ è definito come

$$\underset{1\times n}{a'}\underset{n\times 1}{b} = \left[\begin{array}{ccccc} a_1 & \cdots & a_i & \cdots & a_n \end{array}\right] \left[\begin{array}{c} b_1 \\ \dots \\ b_i \\ \dots \\ b_n \end{array}\right] = a_1 b_1 + \ldots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

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

Lunghezza di un vettore

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

$$\|\underset{n\times 1}{a}\| = \sqrt{a_1^2 + \ldots + a_n^2} = \sqrt{\sum_{i=1}^{n} a_i^2} = \sqrt{\underset{1\times n}{a'}\underset{n\times 1}{a}}$$

Lunghezza di un vettore

La lunghezza di $\underset{2\times 1}{a} = \left[\begin{array}{c} 4 \\ -1 \end{array}\right]$ è
$\|\underset{2\times 1}{a}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17} = 4.123106$

La lunghezza di $\underset{3\times 1}{a} = \left[\begin{array}{c} 4 \\ -1 \\ 3 \end{array}\right]$ è
$\|\underset{3\times 1}{a}\| = \sqrt{4^2 + (-1)^2 + 3^2} = \sqrt{26} = 5.09902$

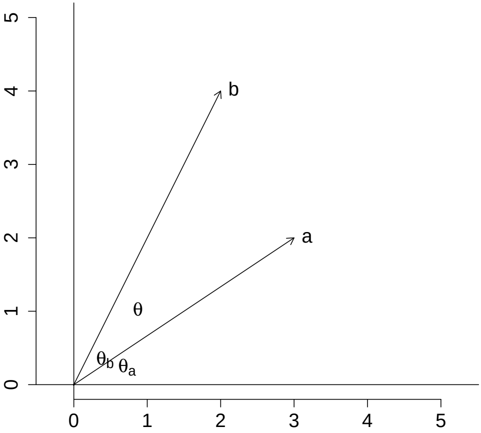
Vettore unitario

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

$$\underset{n\times 1}{1} = \left[\begin{array}{c} 1 \\ \dots \\ 1 \\ \dots \\ 1 \end{array} \right]$$

Angolo fra due vettori

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



Angolo fra due vettori

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

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

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

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 $\underset{n\times 1}{a}$ e $\underset{n\times 1}{b}$,

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

Si noti che poichè

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

i vettori $\underset{n\times 1}{a}$ e $\underset{n\times 1}{b}$ sono perpendicolari quando $\underset{1\times n}{a'}\underset{n\times 1}{b} = 0$

Angolo fra due vettori

Il coseno dell'angolo tra $\underset{2\times 1}{a} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ e $\underset{2\times 1}{b} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ 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 $\theta = 1.494024 \cdot \dfrac{180^\circ}{\pi} = 85.60129^\circ$, ricordando che l'angolo in gradi è uguale all'angolo in radianti per $180^\circ/\pi$