Rappresentazione geometrica dei dati multidimensionali

Il vettore è una m-tupla ordinata di numeri reali che esprime un blocco di informazioni: $(x_{i1}, x_{i2}, \ldots, x_{im})$ che include le rilevazioni sulle variabili effettuate sulla i-esima unità. Di tali vettori se ne usano n (numero di unità rilevate). L'insieme delle rilevazioni forma la matrice dei dati **X**.

Esempio

Membri del Mercosur. Popolazione residente, prodotto interno lordo por-capite e saldo della bilancia dei pagamenti. Dati al 1995

$$
\begin{array}{cc}
 & \begin{array}{ccc} \quad Pop. & Pil.pc & Sbp \end{array} \\
X = & \begin{bmatrix} 34.6 & 8.121 & 2.165 \\ 155.9 & 4.335 & -3.200 \\ 4.8 & 1.854 & -1.675 \\ 3.2 & 5.563 & -0.761 \\ 8.1 & 8.640 & -0.222 \\ 14.2 & 4.739 & 1.374 \end{bmatrix}
\begin{array}{l} \text{Argentina} \\ \text{Brasile} \\ \text{Paraguay} \\ \text{Uruguay} \\ \text{Bolivia} \\ \text{Cile} \end{array}
\end{array}
$$

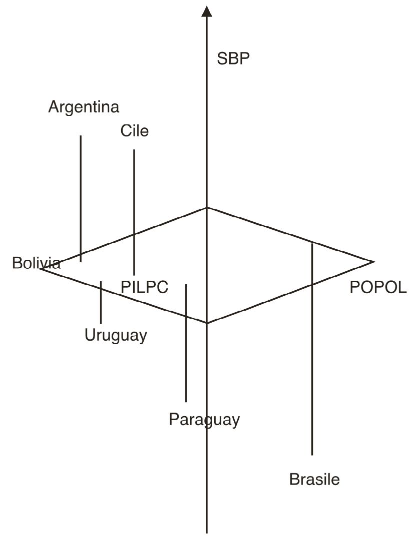
Ogni elemento della matrice è un numero (non necessariamente usato come tale)

$$X_{ij} \qquad i = 1, \ldots, n; \qquad j = 1, 2, \ldots, m$$

è il valore riscontrato per la j-esima variabile sulla i-esima unità. La matrice dei dati è la base empirica delle nostre elaborazioni.

Le relazioni statistiche sono in genere multi-dimensionali ovvero su ogni unità di indagine si acquisisce una pluralità di informazioni, di natura qualitativa e quantitativa che servono per dare un quadro completo dell'unità rispetto alle caratteristiche ritenute opportune da studiare.

Per varie ragioni le informazioni qualitative sono espresse con numeri anche se sono effettuabili solo comparazioni del tipo uguali/diverse e, al più, degli ordinamenti del tipo: basso, medio, altro. Nel caso in esempio però le informazioni sono tutte su scala metrica cioè si possono assimilare a dei punti nello spazio e calcolare una distanza tra di essi.
Tutte le informazioni acquisite in una rilevazione vengono organizzate in una matrice o tabella che prevede, per ogni riga, tutti valori rilevati su di una stessa unità e, in ogni colonna tutti i valori rilevati nel campione per una stessa variabile.



La rappresentazione in 3D avviene a mezzo dei vettori coordinate formati da zero e da un valore "uno" in una posizione specifica

L'Argentina è così rappresentata

$$\begin{bmatrix} 34.6 \\ 8.121 \\ 2.165 \end{bmatrix} = 34.6 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 8.121 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2.165 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \quad e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \text{ l'uno è in posizione "i"}$$

Spazio delle variabili e delle unità.

Le analisi statistiche della matrice dei dati sono talvolta dirette alle relazioni tra le variabili ed in questo caso la si considera un aggregato di m colonne

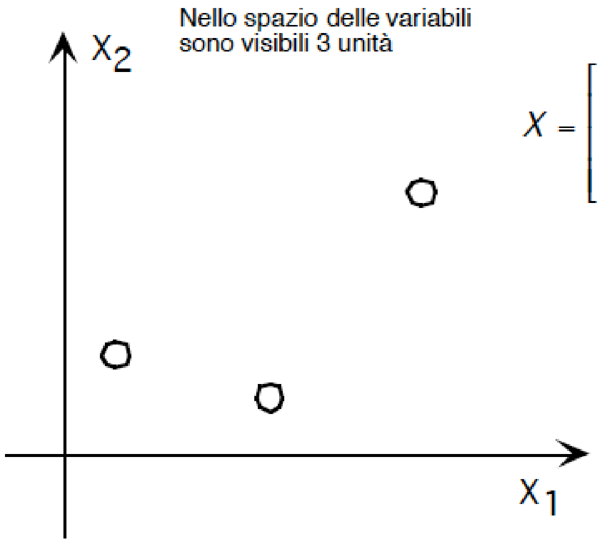
$$X = [x_1, x_2, \ldots, x_m] \qquad x_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix}$$

Ogni $\mathbf{x}_i$ è il vettore colonna formato dalle osservazioni per la i-esima variabile su tutte le n unità incluse nella rilevazione. Si dice che in questa analisi ci si muove nello spazio delle variabili che è un sottoinsieme di R^m. In alternativa l'angolatura di studio potrebbe puntare di più sulle relazioni tra le unità ed in questo caso lamatrice dei dati è considerata un aggregato di n righe

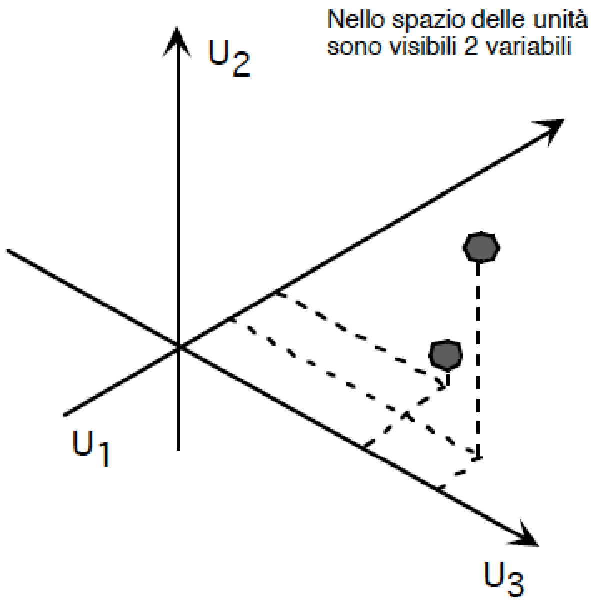
$$X = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad u_i = [u_{i1}, u_{i2}, \ldots, u_{in}]$$

Ogni $\mathbf{u}_i$ è è il vettore riga formato dalle rilevazioni effettuate su tutte le m variabili rispetto alla i-ma unità. Si dice che in queste analisi ci si muove nello spazio delle unità, sottinsieme di R^n.

Esempio



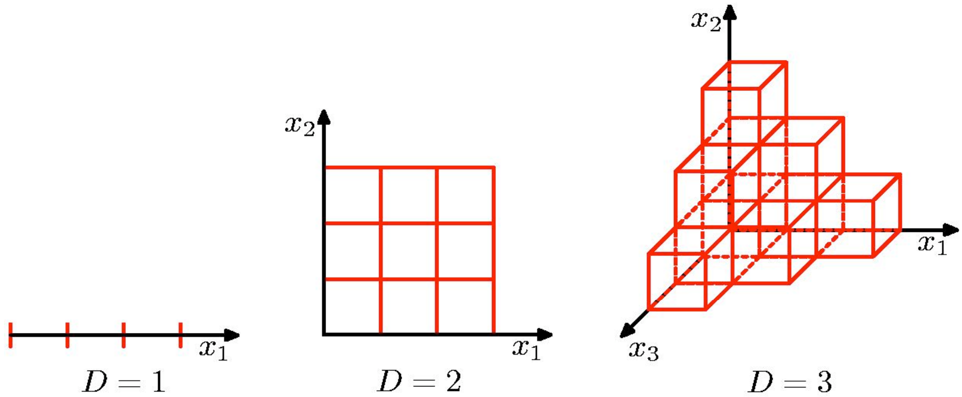
$$X = \begin{bmatrix} 1 & 2 \\ 4 & 1 \\ 7 & 5 \end{bmatrix};$$

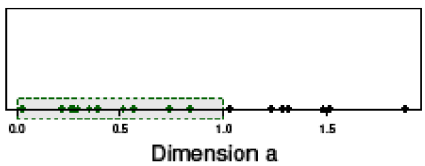


La disposizione dei punti dallo spazio (dalle variabili o delle unità) è la struttura che deve essere spiegata e semplificaa. L'analisi multivariata può fare riferimento sia ad al numero di unità n che al numero di variabili m ed a seconda della scelta si richiamano certe tecniche e non altre. Un punto in comune è che, sia nel caso delle unità che nel caso delle variabili, le dimensioni effettivamente in gioco non sono n o m, ma dei valori inferiori.

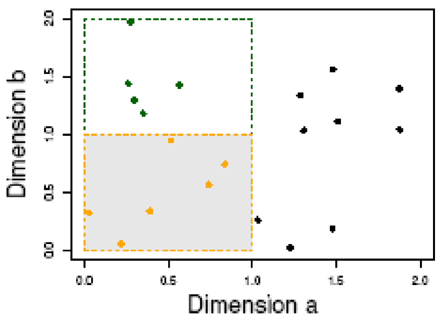
La maledizione della dimensionalità

La qualità di una tecnica di analisi multivariata è in genere misurata sul rapporto: dimensioni necessarie/dimensioni totali e sul rapporto variabilità nello spazio effettivo/variabilità nello spazio totale. Il problema della iperdimesionalità (o della maledizione della dimnensionalità) è che man mano che aumentano le dimensioni i punti diventano sempre più sparsi ed è più difficile individuare relazioni significative.

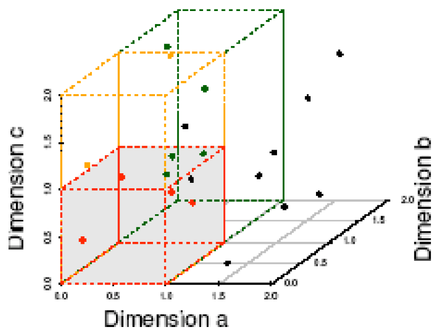




(a) 11 Objects in One Unit Bin



(b) 6 Objects in One Unit Bin



(c) 4 Objects in One Unit Bin

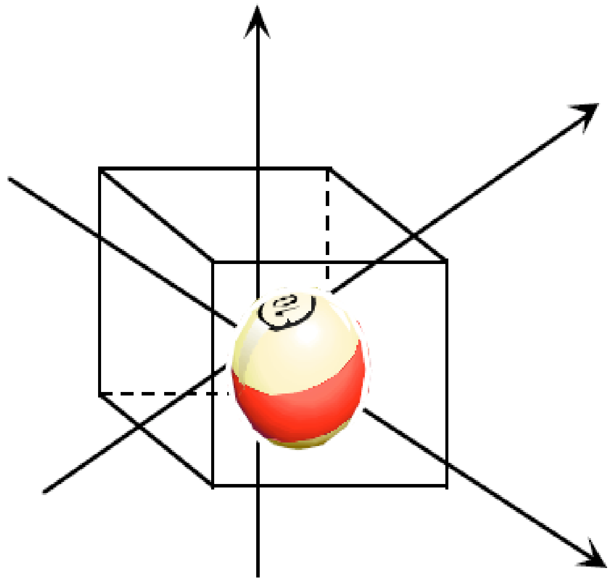
Consideriamo l'ipercubo $[-a,a]^m$ che abbia inscritta una ipersfera di raggio r=a. La quota parte di ipervolume contenuta nella sfera è

$$q_m = \frac{\text{ipervolume sfera}}{\text{ipervolume cubo}} = \frac{\pi^{\frac{m}{2}}}{m2^{m-1}\Gamma(m/2)}$$

dove $\Gamma(.)$ è la funzione gamma che è simile al fattoriale cioè se m/2 è intero vale $\Gamma(m/2) = (m-1)!$ Se m è piccolo il centro dell'ipercubo è impor-tante come le altre zone. All'aumentare di m si ha però

$$\lim_{m \to \infty} q_m = 0$$

Cioè il volume dell'ipercubo tende sempre più a concentrarsi sugli angoli con conseguenze molto serie per le nostre analisi: se i nostri punti sono analizzati con una griglia regolare, la maggior parte delle ipercelle risulterà vuota.



Dimesioni dello spazio	m	1	2	3	4	5	6	7
quota parte di ipervolume	q_m	1	0.785	0.524	0.308	0.164	0.081	0.037

Norme di vettori

La norma di un vettore è una funzione definita in R^m ed a valori in R^+

$$f : R^m \rightarrow R^+$$

Tale che

$$\begin{array}{ll} f(x) = 0 \quad \Rightarrow & x = 0 \\ f(x) \geq 0 & \forall\, x \in R^m \\ f(x+y) \leq f(x) + f(y) & \forall\, x,y \in R^m \\ f(\alpha x) = |\alpha| f(x) & \alpha \in R, x \in R^m \\ |f(x) - f(y)| \leq f(x-y) & \forall\, x,y \in R^m \end{array}$$

Un'importante classe di norme è quella di Minkowsky:

$$\|x\|_p = \left[\sum_{i=1}^{m} |x_i|^p\right]^{\frac{1}{p}} \qquad p \geq 1$$

Tale classe include diversi casi speciali

$$\begin{array}{ll} \text{City block :} & \|x\|_1 = \sum |x_i| \\ \text{Euclidea} & \|x\|_2 = \sqrt{\sum x_i^2} = \left(x^t x\right)^{\frac{1}{2}} \\ \text{Tchebycheff :} & \|x\|_\infty = \max_{1 \leq i \leq m} \{|x_i|\} \end{array}$$

Se nella norma non è indicato l'indice si intende quella euclidea. Tutte le norme sono equivalenti dato che rispondono ad un preciso ordinamento numerico:

$$\|x\|_p \leq \|x\|_{p+1}$$

e quindi se

$$\|x\|_p \leq \|y\|_p \leq \|z\|_p \quad \Rightarrow \quad \|x\|_q \leq \|y\|_q \leq \|z\|_q$$

indipendentemente dal valore assegnato ad p e da q (purché p,q≥1 altrimenti non si tratta di norme)

a) Le norme di Minkowsky verificano la disuguaglianza di Holder

$$|x^t y| \leq \|x\|_p \|y\|_q \qquad \text{per} \quad \frac{1}{p} + \frac{1}{q} = 1$$

che ha come caso speciale la diseguaglianza Cauchy-Schwartz per p=2

$$|x^t y| \leq \|x\|_2 \|y\|_2$$

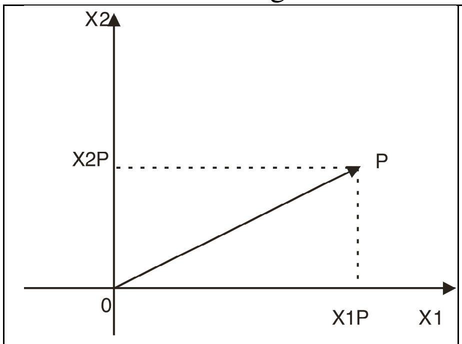
b) Ogni vettore non nullo può essere trasformato (riscalato) in modo da avere norma unitaria.

$$y = \frac{1}{\|x\|}x \Rightarrow \|y\| = \left\| \frac{1}{\|x\|}x \right\| = \frac{1}{\|x\|}\|x\| = 1$$

Esempio

$$x = \begin{bmatrix} 6 \\ 3 \\ -2 \end{bmatrix} \quad \|x\| = \sqrt{x^t x} = \sqrt{36+9+4} = 7 \Rightarrow y = \begin{bmatrix} 6/7 \\ 3/7 \\ -4/7 \end{bmatrix}$$

c) Rappresentazione geometrica del vettore.
I vettori possono essere trattati come entità geometriche. Come mostra la figura, un vettore può essere rapprresentato come una freccia che si diparte dall'origine per raggiungere un punto con coordinate fornite dagli elementi del vettore.



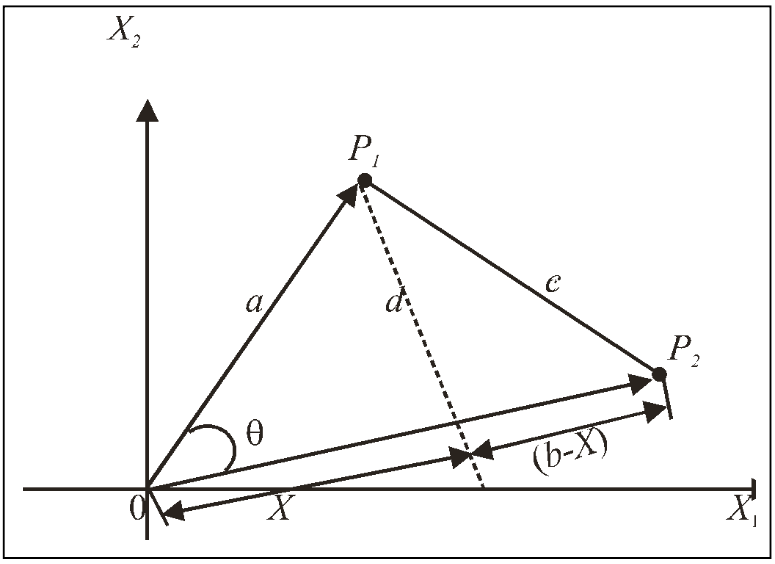
$\overline{OP}$ *è il vettore* $P \equiv \begin{bmatrix} x_{1P} \\ x_{2P} \end{bmatrix}$

con norma $\|P\| = \sqrt{x_{1P}^2 + x_{2P}^2}$

(*in base al teorema di Pitagora*)

$$P = x_{1P}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_{2P}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_{1P}e_1 + x_{2P}e_2$$

d) La legge dei coseni (teorema di Carnot)
I vettori posso differire per lunghezza, per l'angolo formato con gli assi o per entrambi.
Rispetto all'angolo θ tra i vettori P1 e P2 possiamo stabilire un utile teorema.



$d = asen\theta$

$X = a\cos\theta$

$\overline{OX}$ è la proiezione di P_1 su P_2

d può essere visto come la distanza perpendicolare tra P_1 e P_2

Ne consegue che:

$$\begin{aligned} c^2 &= d^2 + (b-X)^2 = d^2 + b^2 - 2bX + X^2 \\ &= a^2 sen\theta + b^2 - 2ab\cos\theta + a^2\cos^2\theta \\ &= a^2\left[sen^2\theta + \cos^2\theta\right] + b^2 - 2ab\cos\theta \end{aligned}$$

poiché

$$sen^2\theta + \cos^2\theta = 1$$

otteniamo:

$$c^2 = a^2 + b^2 - 2ab\cos\theta \;\; (la \;\; legge \;\; di \;\; Carnot)$$

Rispetto ai vettori si ha:

$$c = \|P_1 - P_2\|; \quad a = \|P_1\| \; ; \, b = \|P_2\|$$

e quindi

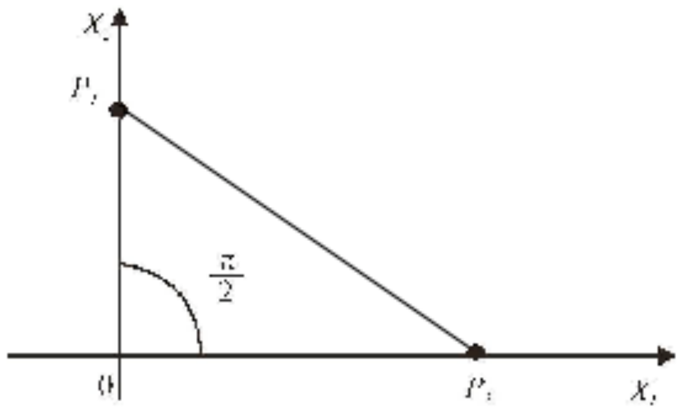
$$\|P_1 - P_2\|^2 = \|P_1\|^2 + \|P_2\|^2 - 2\|P_1\| \, \|P_2\|\cos\theta$$

Se i due vettori fossero ortogonali e cioè se

$$\theta = \frac{\pi}{2} \Rightarrow 90°$$

allora cos(θ)=0 e la distanza tra i due punti vettori sarebbe

$$\|P_1 - P_2\|^2 = \|P_1\|^2 + \|P_2\|^2$$



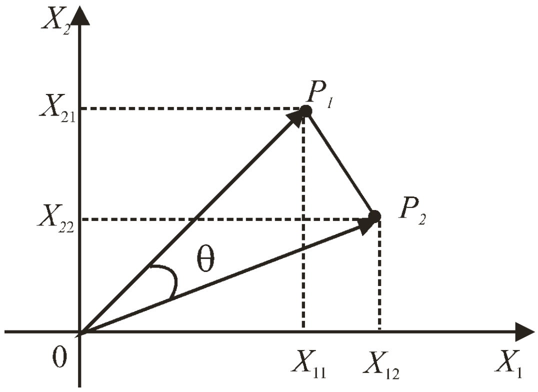
Questo succede quando i punti si trovano sugli assi di un sistema ed ogni punto su di un diverso asse

e) Interpretazione geometrica del prodotto scalare

Il prodotto scalare (o interno) di due vettori

$$P_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \; ; \; P_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$$

dipende dalla lunghezza dei vettori (cioè dalla loro norma) e dall'angolo che essi formano.



Dal teorema di Carnot sappiamo che

$$\overline{P_1P_2}^2 = \overline{OP_1}^2 + \overline{OP_2}^2 - 2\overline{OP_1}\cdot\overline{OP_2}\cos\theta$$

I segmenti possono essere espressi come distanze euclidee (al quadrato)

$$\overline{P_1P_2}^2 = \left(x_{11}^2 - x_{21}^2\right)^2 + \left(x_{12}^2 - x_{22}^2\right)^2 = \left(x_{11}^2 + x_{21}^2\right) + \left(x_{12}^2 + x_{22}^2\right) - 2\left(x_{11}x_{21} + x_{12}x_{22}\right)$$

Poiché

$$\overline{OP_1}^2 = \left(x_{11}^2 + x_{21}^2\right)\ ;\ \ \overline{OP_2}^2 = \left(x_{12}^2 + x_{22}^2\right)$$

Si ha:

$$-2\left(x_{11} - x_{21} + x_{12} - x_{22}\right) = -2\overline{OP_1}\cdot\overline{OP_2}\cos\theta$$

$$\left(x_{11} - x_{21} + x_{12} - x_{22}\right) = \left[\sqrt{\left(x_{11}^2 + x_{21}^2\right)}\cdot\sqrt{\left(x_{11}^2 + x_{21}^2\right)}\right]\cdot\cos\theta$$

E quindi:

$$\cos\theta = \frac{x_{11}x_{21} + x_{12}x_{22}}{\sqrt{\left(x_{11}^2 + x_{21}^2\right)\cdot\left(x_{12}^2 + x_{22}^2\right)}} = \frac{P_1^tP_2}{\|P_1\|\cdot\|P_2\|} = \frac{P_1^t}{\|P_1\|}\cdot\frac{P_2}{\|P_2\|}$$

Ne consegue che il prodotto scalare è pari al prodotto delle lunghezze dei due vettori per il coseno dell'angolo da essi formato.

$$P_1^tP_2 = \|P_1\|\cdot\|P_2\|\cos\theta$$

Dati due vettori noti, ovunque essi siano, possiamo determinare l'angolo tra essi compreso.
Se i vettori sono normalizzati

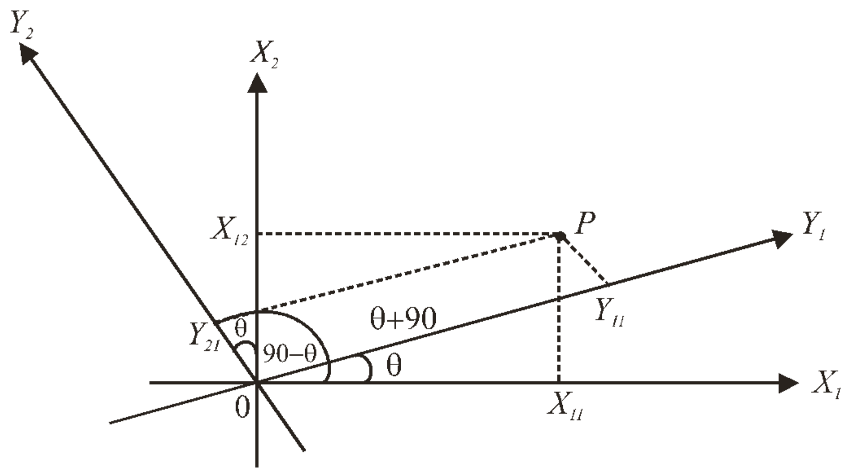
$$P_1^+ = \frac{P_1}{\|P_1\|},\ P_2^+ = \frac{P_2}{\|P_2\|}$$

il prodotto scalare è pari al coseno:

$$P_1^{+^t}P_2^+ = \cos\theta \Rightarrow\ -1 \le P_1^{+^t}P_2^+ \le 1$$

f) Trasformazioni ortogonali (rotazione degli assi)
È' spesso necessaria o opportuna la trasformazione del piano cartesiano: rotazioni, contrazioni, dilatazioni, deformazioni, etc. allo scopo di semplificare il problema trattato. Ogni trasformazione del piano può essere descritta algebricamente con delle equazioni che collegano il piano cartesiano originale con il nuovo sistema di coordinate.

Consideriamo un sistema di assi ortogonali x_1 e x_2 ed un punto $P=(x_1,x_2)$. Guardiamo questo punto anche con un altro sistema di assi ortogonali y_1 e y_2 che è ruotato di un angolo θ rispetto al piano (x_1,x_2) con riferimento all'origine. Questa è una rotazione ortogonale cioè il movimento del piano è rigido e solidale e ogni nuovo asse forma con il corrispondente vecchio asse lo stesso angolo (rotazione in senso antiorario. Moto rigido con perno sull'origine).



Il vettore P può essere scritto in entrambi i sistemi come:

$$\text{sistema } (x_1 x_2) \qquad\qquad \text{sistema } (y_1 y_2)$$
$$P = x_{11}e_1 + x_{12}e_2 \qquad\qquad P = y_{11}e_1^* + y_{12}e_2^*$$

dove e_1, e_2, e_1^*, e_2^* sono dei vettori normalizzati ed ortogonali (ovvero vettori coordinate) tali che

$$e_i^T e_j = d_{ij} \qquad e_i^{*t} e_j^* = d_{ij} \qquad d_{ij} = \text{delta di Kronecker } \Rightarrow d_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Nel sistema (x_1, x_2) effettuiamo il prodotto scalare:

$$e_1^{*t}\left(x_{11}\cdot e_1 + x_{12}\cdot e_2\right) = e_1^{*t}P = x_{11}e_1^{*t}e_2 + x_{12}e_1^{*t}e_2$$

Lo ripetiamo nel sistema (y_1, y_2) per ottenere:

$$e_1^{*t}\left(y_{11}\cdot e_1^* + y_{12}\cdot e_2^*\right) = e_1^{*t}P = y_{11}e_1^{*t}e_1^* + y_{12}e_1^{*t}e_2^* = y_{11}\cdot 1 + y_{12}\cdot 0 = y_{11}$$

e quindi, eguagliando i due risultati, si avrà

$$x_{11}e_1^{*t}e_1 + x_{12}e_1^{*t}e_2 = y_{11}$$

Allo stesso modo, effettuando il prodotto scalare per il secondo vettore per P, si ottiene

$$e_2^{*t} P = x_{11} e_2^{*t} e_1 + x_{12} e_2^{*t} e_2 \;;$$

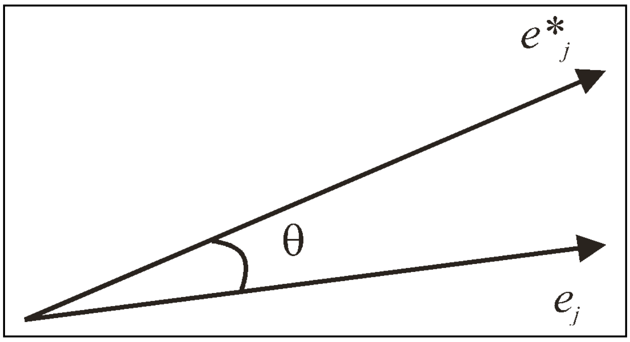
Si arriva pertanto al sistema:

$$\begin{aligned} y_{11} &= x_{11} e_1^{*t} e_1 + x_{12} e_1^{*t} e_2 \\ y_{12} &= x_{11} e_2^{*t} e_1 + x_{12} e_2^{*t} e_2 \end{aligned}$$

Cerchiamo di scoprire la natura geometrica del particolare prodotto scalare $e_i^{*t} e_j$. Innanzitutto:

$$e_i^{*t} e_j = \frac{e_i^{*t} e_j}{\left\| e_i^* \right\| \cdot \left\| e_j \right\|} \quad \text{dato che } \left\| e_i^* \right\| = \left\| e_j \right\| = 1$$

Il rapporto tra il prodotto scalare ed il prodotto delle norme dei vettori, come si è visto, è pari al coseno dell'angolo tra i due vettori



$$\frac{e_i^{*t} e_j}{\left\| e_i^* \right\| \cdot \left\| e_j \right\|} = \cos\theta$$

$$\theta = arcos\left(\frac{e_i^{*t} e_j}{\left\| e_i^* \right\| \cdot \left\| e_j \right\|} \right)$$

Ad ogni prodotto scalare del sistema di trasformazione può essere applicata la relazione con il coseno (tenuto conto del giusto angolo)

$$\begin{aligned} & e_1^{*t} e_1 = \cos\theta \;\; ; \;\; e_2^{*t} e_1 = \cos(\theta + 90) = -sen(\theta) \\ & e_1^{*t} e_2 = \cos(90 - \theta) = \text{sen}\theta \;\; ; \;\; e_2^{*t} e_2 = \cos\theta \end{aligned}$$

Il sistema di trasformazione può quindi essere riespresso come

$$\begin{aligned} y_{11} &= x_{11}\cos\theta - x_{12} sen\theta \\ y_{12} &= x_{11} sen\theta + x_{12}\cos\theta \end{aligned}$$

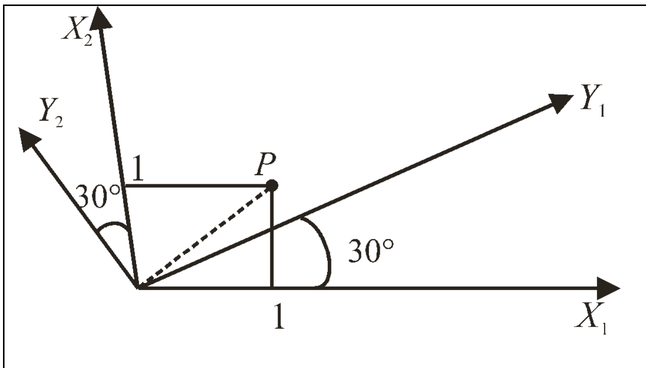
$$\begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} \cos\theta & +sen\theta \\ -sen\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = x_{11} \begin{bmatrix} \cos\theta \\ -sen\theta \end{bmatrix} + x_{12} \begin{bmatrix} sen\theta \\ \cos\theta \end{bmatrix}$$

In notazione matriciale compatta si ha: $\mathbf{Y=XQ^t}$ dove **Q** è una matrice ortogonale detta di rotazione. Tale matrice è anche ortonormale cioè ortogonale e con norma unitaria di righe e di colonne.

$$Q^tQ = QQ^t = \begin{bmatrix} \cos\theta & -sen\theta \\ +sen\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & sen\theta \\ -sen\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2\theta + sen^2\theta & 0 \\ 0 & \cos^2\theta + sen^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Esempio
Supponiamo che θ=30° e determiniamo la corrispondente matrice di rotazione.



$$\cos(\theta) = \sqrt{3}/2 \ ; \ sen(\theta) = 1/2 \Rightarrow$$

$$Q = \begin{bmatrix} \sqrt{3}/2 & +\frac{1}{2} \\ -\frac{1}{2} & \sqrt{3}/2 \end{bmatrix}$$

Il punto P rimane fermo. È il sistema degli assi che ruota. Se P=(1,1) nel sistema (x_1,x_2) nel nuovo sistema (y_1,y_2) diventerà

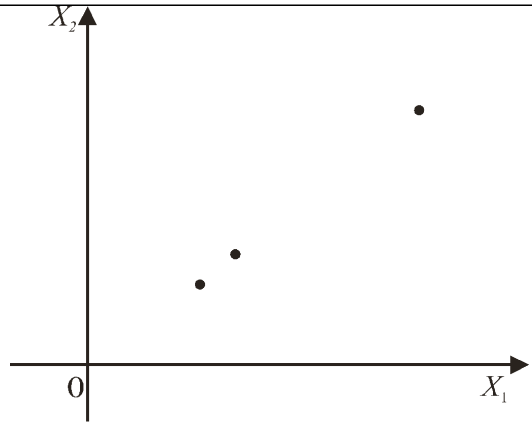
$$\begin{bmatrix} \sqrt{3}/2 & +\frac{1}{2} \\ -\frac{1}{2} & \sqrt{3}/2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (\sqrt{3}+1)/2 \\ (\sqrt{3}-1)/2 \end{bmatrix}$$

Laddove nel vecchio sistema di assi, le coordinate del punto avevano lo stesso valore, nel nuovo sistema i rapporti cambiano e una delle coordinate diventa maggiore dell'altra. Questo, come potremo accertare, non è privo di significato.

g)**Riduzione delle dimensioni**
Consideriamo i punti:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} ; \begin{bmatrix} 2 \\ 2 \end{bmatrix} ; \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

I punti sono perfettamente allineati lungo la retta $x_2=x_1$. Nota una coordinata l'altra si determina in modo automatico. Si vuole ora spostare l'asse x_1 in modo da farlo coincidere con la retta di equazione $x_2=x_1$.



La matrice di rotazione necessaria per tale operazione è

$$\theta = \frac{\pi}{4} \Rightarrow \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -sen\left(\frac{\pi}{4}\right) \\ sen\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & +\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

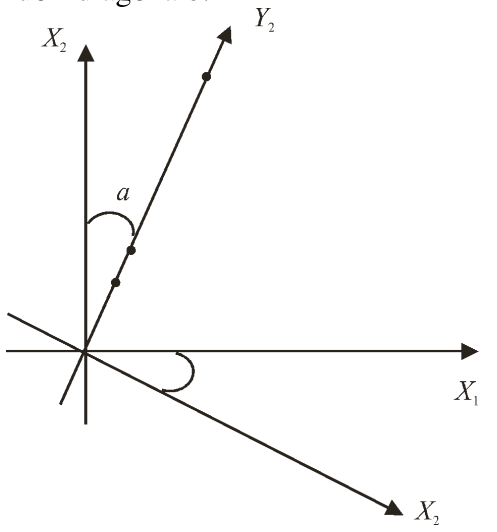
I punti trasformati sono perciò

$$\frac{1}{2}\begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

$$\frac{1}{2}\begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

$$\frac{1}{2}\begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 6\sqrt{2} \end{bmatrix}$$

La rotazione avviene in senso orario per cui si inverte il segno degli elementi fuori diagonale.



L'aggregato dei punti nelle coordinate originali era:

$$\begin{bmatrix} 1 & 2 & 6 \\ 1 & 2 & 6 \end{bmatrix}$$

Nelle nuove coordinate diventa invece:

$$\begin{bmatrix} 0 & 0 & 0 \\ \sqrt{2} & 2\sqrt{2} & 6\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 6 \end{bmatrix}$$

e quindi, di fatto, è stata eliminata una dimensione che era solo un duplicato dell'altra: basta un solo asse per rappresentare i tre punti dati. È chiaro che nella realtà in punti non sono mai perfettamente allineati, pure una opportuna rotazione degli assi consente di trascurare dimensioni poco informative.

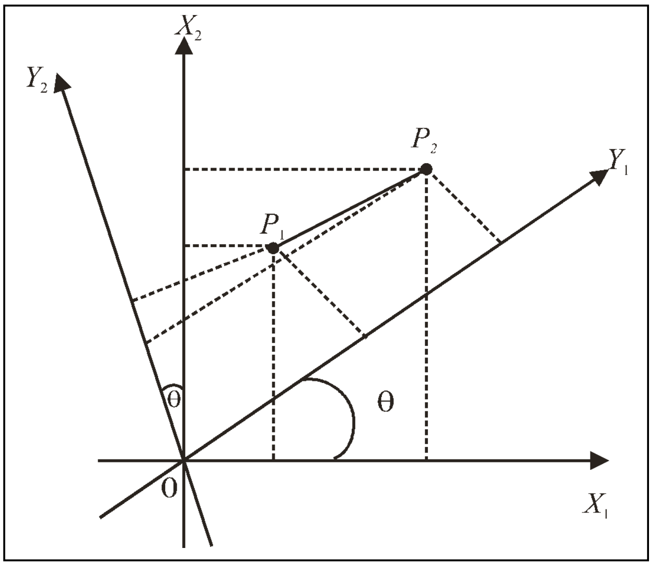
h)Rotazione degli assi e distanze.

Le rotazioni ortogonali mantengono inalterata la distanza euclidea tra i punti. Consideriamo la trasformazione **Y=RX** dove

$$R = \begin{bmatrix} \cos(\theta) & +sen(\theta) \\ -sen(\theta) & \cos(\theta) \end{bmatrix}$$

N.B. La rotazione ora àè in senso antiorario. Indichiamo i nuovi punti nel sistema ruotato con

$$Q_1 = RP_1 \quad \text{e} \quad Q_2 = RP_2$$



La distanza euclidea tra i duepunti nelle nuove coordinate è data da

$$\begin{aligned}\|Q_1 - Q_2\|^2 &= (Q_1 - Q_2)^t (Q_1 - Q_2) = Q_1^t Q_1 - Q_1^t Q_2 - Q_2^t Q_1 + Q_2^t Q_2 \\ &= P_1^t (R^t R) P_1 - P_1^t (R^t R) P_2 - P_2^t (R^t R) P_1 + Q_2^t (R^t R) Q_2 \\ &= P_1^t P_1 - P_1^t P_2 - P_2^t P_1 + P_2^t P_2 = (P_1 - P_2)^t (P_1 - P_2) \\ &= \|P_1 - P_2\|^2\end{aligned}$$

Quindi la distanza tra i punti P_1 e P_2 è la stessa nei due sistemi di assi purché la rotazione sia ortogonale.

Le proprietà delle rotazioni ortogonali si applicano oltre che al piano anche allo spazio m dimensionale: anche se qui sono meno immediate le relazioni tra gli elementi della matrice di rotazione e le funzioni trigonometriche degli angoli che i vettori formano con gli assi. In generale, ogni trasformazione ortogonale corrisponde ad una rotazione degli assi (intorno all'origine) di un dato angolo. Le distanze tra i punti restano comunque invariate.

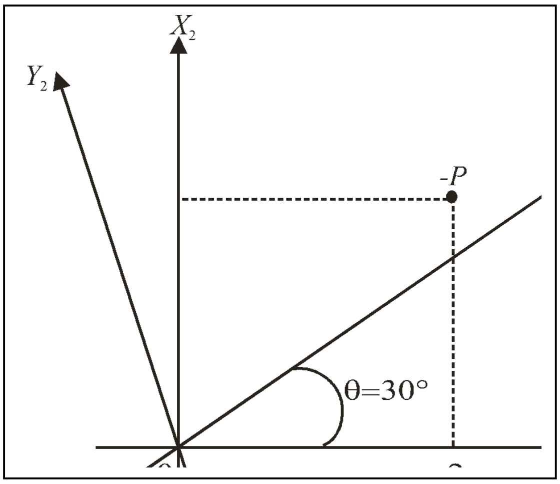
L'uguaglianza delle norme – sotto rotazione ortogonale – non implica che gli elementi dei vettori abbiano lo stesso peso all'interno dei vettori in relazione agli altri elementi.

Esempio

Consideriamo il punto P=(2,√3) ed effettuiamo una rotazione degli assi di 30° (π/6)

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 2 \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{5}{2} \end{bmatrix}$$

Dal sistema $y_1 y_2$ si ruota a $x_1 x_2$



Il ruolo delle dimensioni è diverso nei due sistemi.

Nei due sistemi, la norma al quadrato nonché i pesi degli elementi (nell'ambito del vettore normalizzato) sono

$$\begin{array}{cc} Sistema\ (x_1, x_2) & Sistema\ (y_1, y_2) \\ \dfrac{\left(\frac{\sqrt{3}}{2}\right)^2 + \left(\frac{5}{2}\right)^2}{7} = 1 & \dfrac{(2)^2 + (\sqrt{3})^2}{7} = 1 \\ \dfrac{3}{28} + \dfrac{25}{28} = 1 & \dfrac{16}{28} + \dfrac{12}{28} = 1 \end{array}$$

Nel secondo sistema il peso delle dimensioni si inverte e l'asse più rilevante o preferenziale è y_1 e non più x_2. È' chiaro che a seconda della posizione del punto e dell'angolo di rotazione del sistema si può fare in modo da precostituire l'ordinamento dei pesi desiderato.

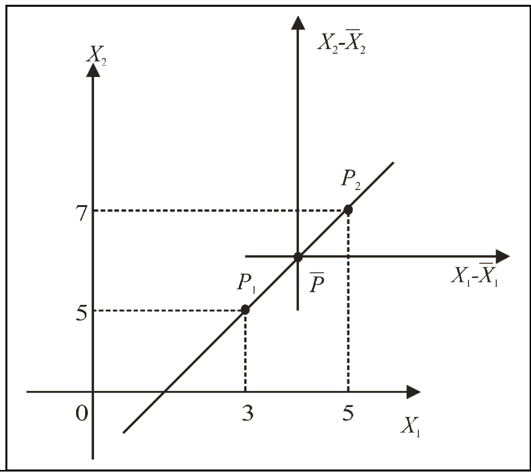
i) Scelta del perno (pivot) di rotazione

Tutte le rotazioni fin qui proposte sono state riferite all'origine. Tale scelta è corretta se si vuole stabilire il peso più rilevante per un solo punto, ma se i punti sono più di uno ci si deve dare un criterio per individuare il perno di rotazione più appropriato.

La sola rotazione non basta ad alterare significativamente
il sistema dei pesi degli assi.
Infatti, per i punti P_1=(3,5) e P_2=(5,7) Si ha

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 4\sqrt{2} & 6\sqrt{2} \end{bmatrix}$$

I punti sono allineanti, ma la semplice rotazione intorno all'origine non lo evidenzia.



Per superare questo problema, gli assi debbono essere spostati sul punto centrale tra quelli dati: $\overline{P}$ =(4,6). tale punto è ottenuto come media aritmetica dei valori.

Le nuove coordinate si ottengono come scarti dalla media: $x_1 - \bar{x}_1$ e $x_2 - \bar{x}_2$
La matrice degli scarti è

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

che, ruotata di 45°, produce

$$\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 0 & 0 \end{bmatrix}$$

che finalmente chiarisce il numero vero di dimensioni necessarie per rappresentare i punti dati nell'esempio (una e non due come sembrava all'inizio). Il centramento dei punti (misurazione come scarto dalle medie) consente di individuare le relazioni lineari perfette azzerando una o più dimensioni.

j) Centramento della matrice dei dati

Per varie ragioni è opportuno centrare la matricedei dati e cioè è più pratico operare con una matrice di scarti piuttosto che con la matrice dei valori originali.

$$\hat{X} \begin{bmatrix} x_{11} - \hat{\mu}_1 \;, & x_{12} - \hat{\mu}_2 & \dots & x_{1m} - \hat{\mu}_m \\ x_{21} - \hat{\mu}_1 & x_{22} - \hat{\mu}_2 & & x_{2m} - \hat{\mu}_m \\ \vdots & & & \vdots \\ x_{n1} - \hat{\mu}_1 & x_{n2} - \hat{\mu}_2 & & x_{nm} - \hat{\mu}_m \end{bmatrix} \; dove \;\; \hat{\mu}_1 = \frac{1}{n} \sum_{j=1}^{n} x_{j1}$$

Conviene inoltre esprimere la matrice degli scarti in esplicita notazione matriciale:

$$\hat{X} = CX \qquad \text{dove} \quad C = \left(I_n - \frac{1}{n} \nu_n \nu_n^t \right)$$

$$\nu_n^t = [\overbrace{1,1,\ldots,1}^{\text{n-volte}}]$$

e **C** è la cosiddetta matrice di centramento. Ecco un esempio

$$\begin{bmatrix} 1-3 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} ; \quad \hat{\mu}^t = (2, \quad 1, \quad 3) \Rightarrow \hat{x} = \begin{bmatrix} -1 & 0 & -6 \\ 2 & -1 & 7 \\ 0 & 1 & 2 \\ -1 & 0 & -3 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \right\} \begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 8 & 4 & 12 \\ 8 & 4 & 12 \\ 8 & 4 & 12 \\ 8 & 4 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 1 & 3 \\ 2 & 1 & 3 \\ 2 & 1 & 3 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -6 \\ 2 & -1 & 7 \\ 0 & 1 & 2 \\ -1 & 0 & -3 \end{bmatrix}$$

La matrice di centramento **C** è idempotente e simmetrica

$$C = C^t \quad \text{e} \quad C \cdot C = C^2 = C$$

$$C^t = \left[I_n - \frac{1}{n} \nu_n \nu_n^t \right]^t = \left[I_n^t - \frac{1}{n} \left(\nu_n \nu_n^t \right) \right] = I_n - \frac{1}{n} \nu_n \nu_n^t = C$$

$$C^t \cdot C = C \cdot C = \left[I_n - \frac{1}{n} \nu_n \nu_n^t \right] \cdot \left[I_n - \frac{1}{n} \nu_n \nu_n^t \right] = I_n - \frac{1}{n} \nu_n \nu_n^t - \frac{1}{n} \nu_n \nu_n^t + \frac{1}{n^2} \nu_n \nu_n^t \nu_n \nu_n^t$$

$$= I_n - \frac{1}{n} \nu_n \nu_n^t - \frac{1}{n} \nu_n \nu_n^t + \frac{1}{n^2} n \nu_n \nu_n^t = I_n - \frac{1}{n} \nu_n \nu_n^t = C$$

La matrice di devianze-codevianze èfacilmente ottenuta dal prodotto matriciale

$$\hat{X}^t \hat{X} = X^t C^t C X = X^t C X = V$$

La matrice di varianze-covarianze

$$W = \frac{1}{n}V = \frac{1}{n}X^t CX = \left[\frac{1}{\sqrt{n}} \quad X^t \quad C^t\right]\left[\frac{1}{\sqrt{n}} \quad C \quad X\right]$$

Si ottiene dal prodotto di matrici trasformate

$$Y = \frac{1}{\sqrt{n}}CX \quad \Rightarrow \quad Y^t Y = W$$

Esempio

$$Y = \frac{1}{2}\begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix}\begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 & 0 & -6 \\ 2 & -1 & 7 \\ 0 & 1 & 2 \\ -1 & 0 & -3 \end{bmatrix}$$

$$Y^t Y = \begin{bmatrix} -1/2 & 1 & 0 & -1/2 \\ 0 & -1/2 & 1/2 & 0 \\ -3 & 7/2 & 1 & -3/2 \end{bmatrix}$$

$$= \begin{bmatrix} 6/4 & -2/4 & 23/4 \\ -2/4 & 2/4 & -5/4 \\ 23/4 & -5/4 & 67/4 \end{bmatrix} = \mathbf{W}$$

Questi risultati giustificano l'adozione della trasformazione lineare

$$\hat{X} = CX$$

Il cui effetto è di spostare l'origine degli assi sul centroide della matrice dei dati

$$\hat{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)$$

La dispersione dei punti rimane tuttavia invariata.

E' importante non perdere di vista il risultato ottenuto: il contenuto informativo della matrice dei dati può essere fatto confluire nella matrice di varianze-covarianze. Almeno quello essenziale, visto che comunque un'informazione si perde: la differenza tra le medie delle variabili originarie che è pari a zero in tutte le variabili cenbtrate.