Dall'espressione ⊛ si ha che:

$$\frac{\partial a^T S a}{\partial a_1} = 2 a_1 S_{11} + 2 \sum_{j=2}^{q} a_j S_{1j}$$

$$= 2 \sum_{j=1}^{q} a_j S_{1j}$$

Analogamente

$$\frac{\partial a^T S a}{\partial a_K} = 2 a_K S_K + 2 \sum_{j=K+1}^{q} a_j S_{Kj} + 2 \sum_{i=1}^{K-1} a_i S_{Ki}$$

$$= 2 \sum_{j=1}^{q} a_j S_{Kj}$$

$$\Downarrow$$

$$\frac{\partial a^T S a}{\partial a} = 2 S a$$

Concetti preliminari

δ) Interpretazione geometrica del prodotto scalare tra due vettori

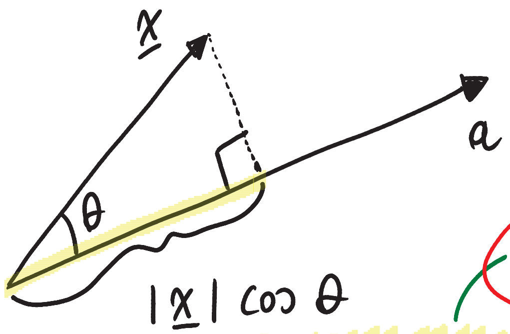
$\underset{p\times 1}{a}$; $\underset{p\times 1}{\underline{x}}$

$$|a| = \sqrt{a^T a} = \sqrt{\sum a_i^2}$$

Identità fondamentale

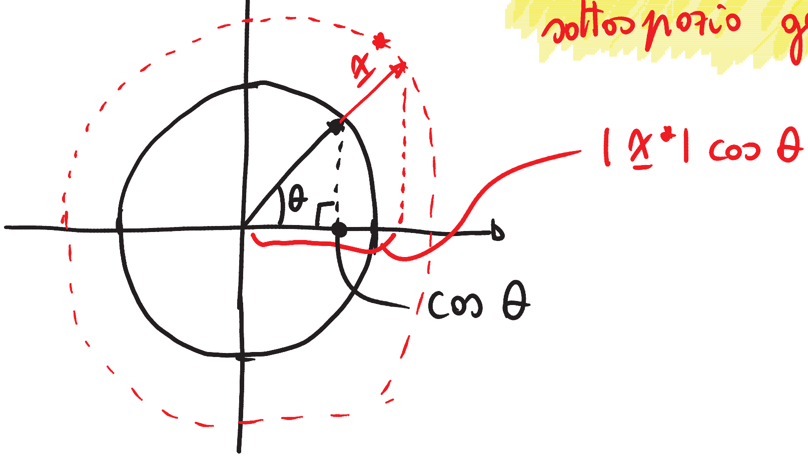
$$a^T \underline{x} = \underline{x}^T a = |a|\,|\underline{x}| \cos\theta \qquad (**)$$

dove θ è l'angolo formato tra i due vettori a e $\underline{x}$.



con segno

↳ lunghezza della proiezione di $\underline{x}$ sul sottospazio generato da a.



Dall'identità fondamentale si ha dunque che:

$$\frac{\underline{x}^T a}{|a|} = |\underline{x}^T| \cos \theta$$

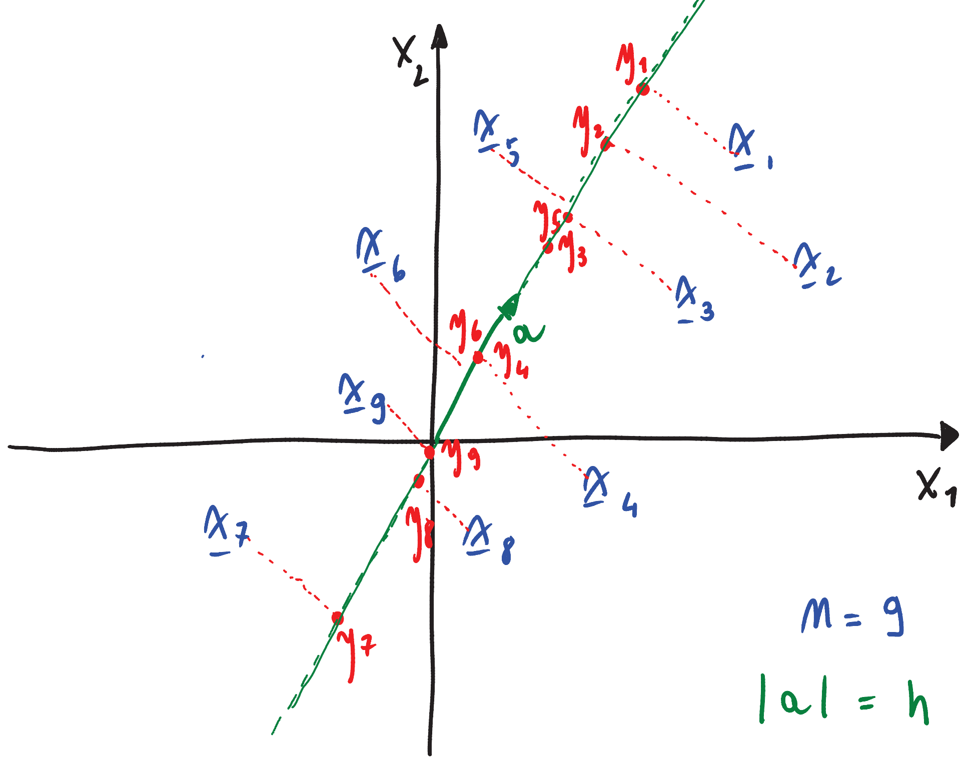
$$\Downarrow$$

L'identità fondamentale (**) dice dunque che il risultato del prodotto scalare $\underline{x}^T a$ coincide con la lunghezza → con segno della proiezione di $\underline{x}$ sulla retta generata da a, espressa in una unità di misura coincidente con la lunghezza $|a|$ del vettore a.

— ∘ — ∘ — ∘ —

OSSERVAZIONE: Interpretazione geometrica di $y = \underset{n \times q}{X} a$

Semplifichiamo: $\underset{n \times 2}{X}$ ⇒ matrice di dati relativi ad n individui su cui si rilevano 2 caratteri



ANALISI DISCRIMINANTE

Dati analizzati

Si ha una popolazione suddivisa in 2 sottopopolazioni (gruppi).
Dalla 1° sottopopolazione si estrae un campione di dimensione n_1. Dalla 2° sottopopolazione si estrae un campione di dimensione n_2.
Su ciascuna delle $n = n_1 + n_2$ unità statistiche si rilevano q caratteri quantitativi.
Dalla rilevazione si ottiene la seguente matrice di dati:

$$\underset{n \times q}{X} = \left[\begin{array}{c} \underset{n_1 \times q}{X_1} \\ \hdashline \underset{n_2 \times q}{X_2} \end{array} \right] = \{ x_{igj} \}$$

gruppo

unità

variabile

Obiettivi:

1) Costruire una regola di classificazione

2) Individuare le variabili che maggiormente contribuiscono a spiegare l'appartenenza di una unità statistica alla prima o alla seconda sottopopolazione.

Metodologia:

Si definisce una funzione lineare delle q

variabili (funzione discriminante lineare)

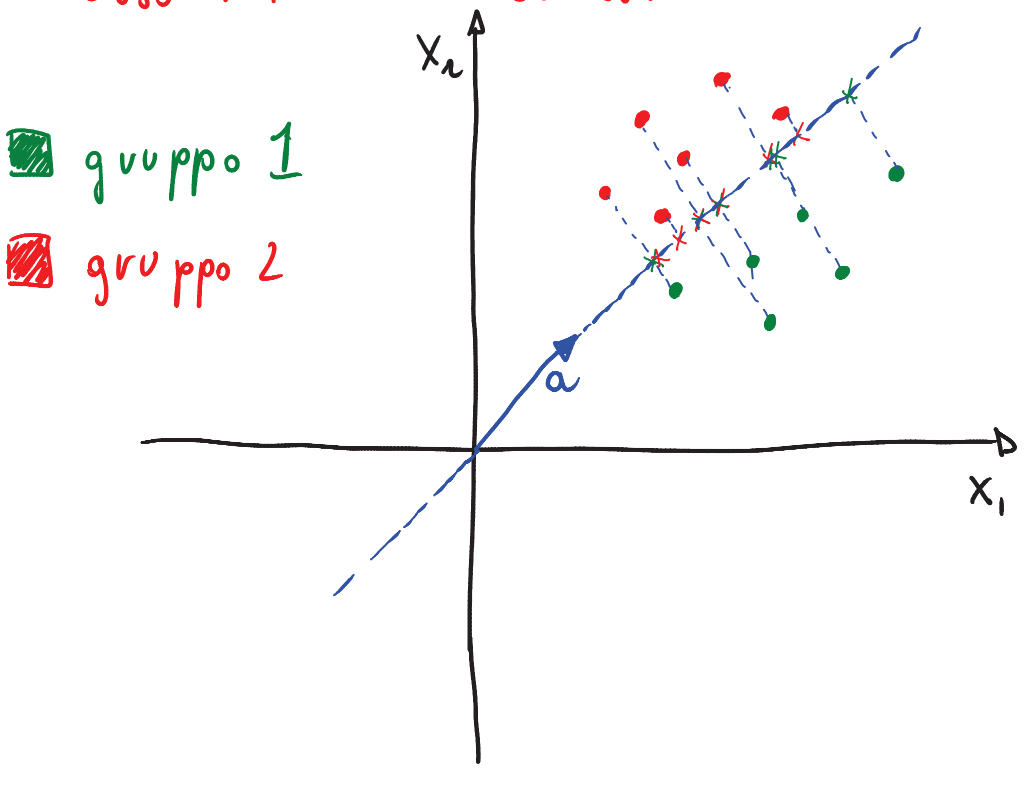
$$y = X a \longrightarrow \text{punteggi discriminanti (scores)}$$

determinando a in modo tale da rendere più evidenti possibili le differenze tra le unità statistiche appartenenti alle due sottopopolazioni.

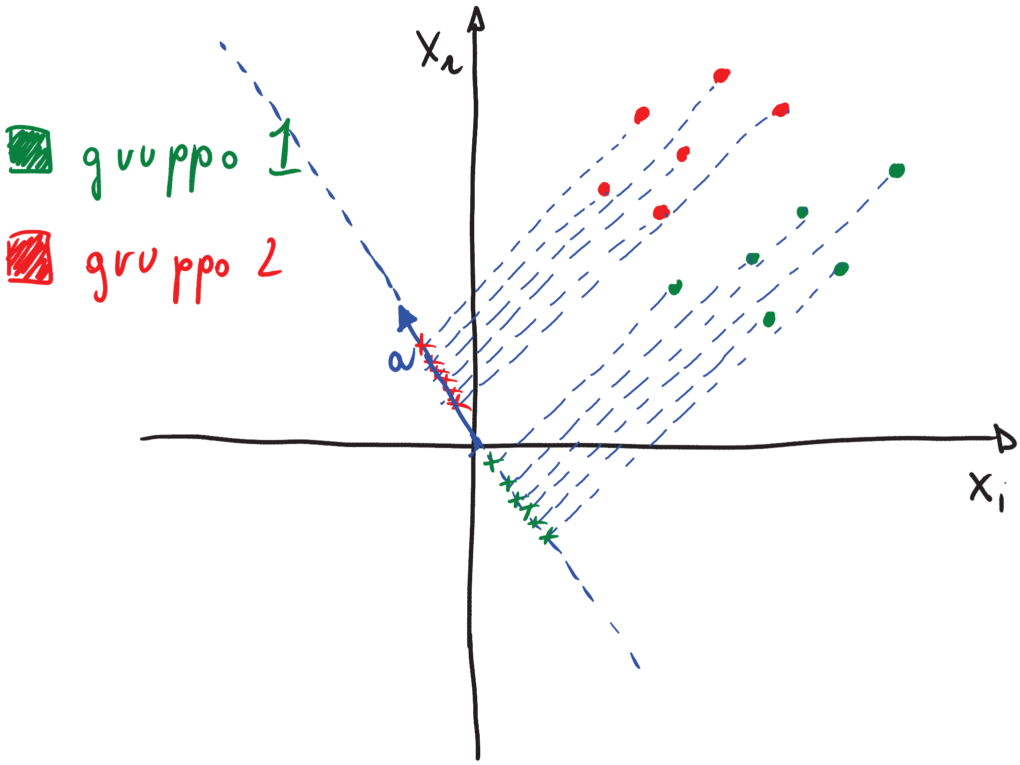
Esemplificando:

- 2 variabili
- $n_1 = n_2 = 6$
- le 12 righe di X sono dei punti in $\mathbb{R}^2$. Considerando $y = Xa$ sto in sostanza proiettando i 12 punti sulla retta generata da a.

Caso 1: a scarsamente discriminante



Caso 2 : a molto discriminante



$\Downarrow$

Il vettore a deve essere determinato in modo tale che le unità statistiche appartenenti al medesimo gruppo abbiano punteggi discriminanti simili mentre le unità appartenenti a gruppi diversi abbiano punteggi diversi.

In termini più tecnici desidero che gli scores presentino una bassa varianza nei gruppi ed un'alta varianza fra i gruppi.

- $\underset{q\times q}{S}$ = matrice di varianza e covarianza ottenuta da $\underline{X}$

- $S = S_w + S_b$

- $a^T S a = Var(y)$ (vedi β)

$$Var(y) = a^T (S_w + S_b) a$$

$$= \underbrace{a^T S_w a}_{} + \underbrace{a^T S_b a}_{}$$

↳ Varianza tra i gruppi degli scores

Varianza nei gruppi degli scores

A questo punto, sembra ragionevole scegliere a in modo tale da massimizzare $a^T S_b a$.

Si giunge così al problema di ottimo vincolato

$$\begin{cases} \max\limits_{a} & a^T S_b a \\ \text{sub} & a^T S a = 1 \end{cases}$$

→ perché si introduce il vincolo?

che viene risolto ricorrendo alla tecnica dei moltiplicatori di Lagrange.

- Si definisce la Funzione Lagrangiana

$$L = a^T S_b a - \lambda (a^T S a - 1)$$

↳ moltiplicatore di Lagrange

- Si risolve il sistema $\begin{cases} \dfrac{\partial L}{\partial a} = 0_{q \times 1} \\ \dfrac{\partial L}{\quad} = 0 \end{cases}$

- Si risolve il sistema $\begin{cases} \dfrac{\partial L}{\partial a} = \underset{q\times 1}{0} \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases}$

$$\frac{\partial L}{\partial \lambda} = a^T S a - 1$$

$$\frac{\partial L}{\partial a} = \frac{\partial\, a^T S_b a}{\partial a} - \lambda \frac{\partial\, a^T S a}{\partial a}$$

(vedi γ) $= 2 S_b a - 2\lambda S a$

$$2 S_b a - 2\lambda S a = \mathbf{0}$$

$$S_b a = \lambda S a$$

$$\Downarrow$$

$$\begin{cases} S_b a = \lambda S a \\ a^T S a = 1 \end{cases}$$

Prendiamo in considerazione la prima equazione e moltiplichiamo entrambi i membri per S^{-1}:

membri per S^{-1}:

S^{-1} esiste se Rango$(X) = q$ e $n-1 > q$

$$S^{-1} S_b \, a = \lambda a$$

$$\Downarrow$$

a è un autovettore di $S^{-1}S_b$ e λ è il corrispondente autovalore.

Inoltre, dalla ⊛, si ottiene premoltiplicando per a^T:

$$a^T S_b a = \lambda \underbrace{a^T S a}_{=1} = \lambda \qquad \text{(2° equazione sistema)}$$

$$\lambda = a^T S_b a$$

a è l'autovettore associato all'autovalore più grande di $S^{-1}S_b$.

Quanti autovalori non nulli ha la matrice $S^{-1}S_b$?

Essendo Rango$(S_b) = 1$ (vedi α), allora anche Rango$(S^{-1}S_b) = 1 \Rightarrow S^{-1}S_b$ ha un unico autovalore non nullo.

Ricordando che

$$S_b = \frac{n_1 n_2}{n^2} (\underline{\mu}_1 - \underline{\mu}_2)(\underline{\mu}_1 - \underline{\mu}_2)^T$$

risulta semplice calcolare analiticamente il valore di λ e l'espressione di a.
In particolare, sia

$$h = \frac{\sqrt{n_1 n_2}}{n} (\underline{\mu}_1 - \underline{\mu}_2) .$$

Si ha che:

$$S_b = h\, h^T$$

$$S^{-1} S_b\, a = \lambda a \Rightarrow S^{-1} h\, h^T a = \lambda a$$

$$\underbrace{h^T S^{-1} h}_{\text{scalare}}\, h^T a = \lambda h^T a \Rightarrow \lambda = h^T S^{-1} h$$

$$\lambda = \frac{n_1 n_2}{n} (\underline{\mu}_1 - \underline{\mu}_2)^T S^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

Per quanto riguarda la determinazione del vettore a:

$$S_b\, a = \lambda S a \Rightarrow S a = \frac{1}{\lambda} S_b\, a$$

$$S a = \frac{1}{\lambda} h\, h^T a$$

scalare

$$S a = h \frac{h^T a}{\lambda}$$

$$S a = \frac{h^T a}{\lambda} h$$

$\rightarrow$ non influisce

$$S a = \frac{n\, a}{\lambda}\, h$$

$$a = \frac{h^T a}{\lambda}\, S^{-1} h$$

non influisce sulla direzione di a ma solo sul suo modulo.

scalare

$$a = \frac{h^T a}{\lambda} \cdot \sqrt{\frac{n_1 n_2}{n}}\; S^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

$$a = \gamma \sqrt{\frac{n_1 n_2}{n}}\; S^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \qquad \gamma = \frac{h^T a}{\lambda}$$

γ viene determinato in modo tale da soddisfare il vincolo $a^T S a = 1$

$$a = \gamma S^{-1} h \Rightarrow a^T S a = \gamma^2 h^T S^{-1} S S^{-1} h$$

$$\gamma^2 h^T S^{-1} \underbrace{S S^{-1}}_{I_q} h = 1$$

$$\gamma^2 \underbrace{h^T S^{-1} h}_{\lambda} = 1 \Rightarrow \gamma^2 \lambda = 1$$

$$\gamma = \frac{1}{\sqrt{\lambda}}$$

Concludendo:

$$a = \frac{1}{n} \sqrt{\frac{n_1 n_2}{\lambda}}\; S^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

↳ operativamente si è soliti porre

$$a = S^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

OSSERVAZIONE: Molto spesso la f. d. è definita come segue:

$$a^* = S_w^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

Ciò deriva dal fatto che i problemi di ottimizzazione

$$\begin{cases} \max\limits_a & a^T S_b a \\ \text{sub} & a^T S a = 1 \end{cases}$$

e

$$\begin{cases} \max\limits_a & \dfrac{a^T S_b a}{a^T S_w a} \\ \text{sub} & a^T S a = 1 \end{cases}$$

risultano equivalenti in virtù del fatto che

$$S_b = S - S_w$$

Di conseguenza a e a^* differiscono solo nel modulo ma non nella direzione.

UTILIZZO dei PUNTEGGI DISCRIM.

- $y = \underset{\sim}{X} a$

Definizione di un criterio di classificazione.

Analizzando una NUOVA unità statistica si è ottenuto il seguente vettore di dati:

$$\underline{x}^* = [x_1^*, \dots, x_q^*]$$

È più verosimile ritenere che $\underline{x}^*$ provenga dal gruppo 1 o dal gruppo 2?

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1m_1} \\ y_{21} \\ \vdots \\ y_{2m_2} \end{bmatrix}$$

Scores delle unità del gruppo 1

scores delle unità del gruppo 2

CENTROIDI:

Centroide del 1° gruppo:

$$\bar{y}_1 \quad 1 \quad \sum^{m_1} y_{1i}$$

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_{1i}$$

Centroide del 1° gruppo:

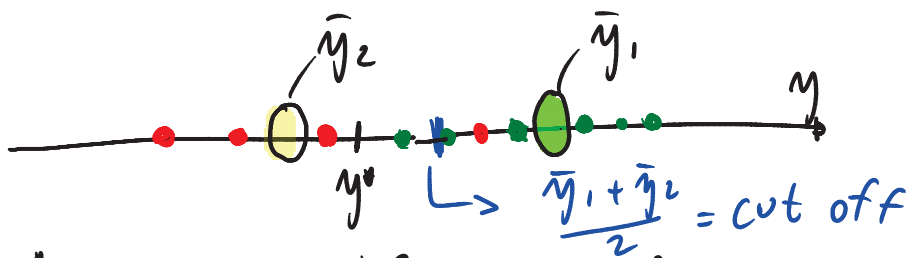
$$\bar{y}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} y_{2i}$$

Score della nuova unità statistica

$$y^* = \underline{x}^* a$$

Regola di classificazione: se y^* è più vicino a $\bar{y}_1$ di quanto non lo sia a $\bar{y}_2$ allora classifico $\underline{x}^*$ come appartenente al gruppo 1. Al contrario, se y^* è più vicino a $\bar{y}_2$ allora classifico $\underline{x}^*$ come appartenente al gruppo 2.

ES:



$\underline{x}^*$ proviene dal gruppo 2.

In formule:

2) Sia $\bar{y}_1 < \bar{y}_2$, allora

- se $y^* < \dfrac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow \underline{x}^*$ proviene