

1) Cosa si intende per approccio parametrico e non-parametrico nell'ambito della classificazione? Fare un esempio di classificatore parametrico e non parametrico.

Dispense "Classificazione (1)"

2) Qual è il principio su cui si basa il classificatore SVM? Cosa si intende per margine?

Dispense "Classificazione (2)"

**3)** Come può essere matematicamente definita la loss function (su un singolo pattern) per l'addestramento di una rete neurale?

Dispense "Reti Neurali"

**4)** Quali sono i più noti algoritmi di clustering?

Dispense "Clustering"

5) Dato un training set composto dai seguenti pattern:

$$\mathbf{x}_1 = \begin{bmatrix} 1.9 \\ -2.5 \\ -5.2 \\ 7.6 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 9.9 \\ -1.6 \\ -4.9 \\ 3.2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 6.3 \\ -0.5 \\ 1.9 \\ -2.0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 2.5 \\ -8.8 \\ 1.6 \\ 9.4 \end{bmatrix}$$

a cui sono associate le seguenti osservazioni (variabile dipendente):

$$y_1 = 3.4, \quad y_2 = -2.9, \quad y_3 = -7.2, \quad y_4 = 7.5$$

formulare il problema di *multiple linear regression* definendo la matrice  $\mathbf{X}$  e il vettore  $\mathbf{y}$ .

### Svolgimento

La matrice  $\mathbf{X}$  di dimensione  $n \times (d + 1)$  si ottiene disponendo su ciascuna riga gli  $n$  pattern del training set e inserendo a destra una colonna con tutti valori 1:

$$\mathbf{X} = \begin{bmatrix} 1.9 & -2.5 & -5.2 & 7.6 & 1 \\ 9.9 & -1.6 & -4.9 & 3.2 & 1 \\ 6.3 & -0.5 & 1.9 & -2.0 & 1 \\ 2.5 & -8.8 & 1.6 & 9.4 & 1 \end{bmatrix}$$

Il vettore  $\mathbf{y}$  di dimensione  $n$  si ottiene disponendo in colonna i valori  $y_i$ :

$$\mathbf{y} = \begin{bmatrix} 3.4 \\ -2.9 \\ -7.2 \\ 7.5 \end{bmatrix}$$

Anche se non richiesto dall'esercizio, definita la matrice  $\mathbf{X}$  e il vettore  $\mathbf{y}$  le relazioni tra variabili indipendenti e dipendente possono essere scritte in forma matriciale come:

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta}$$

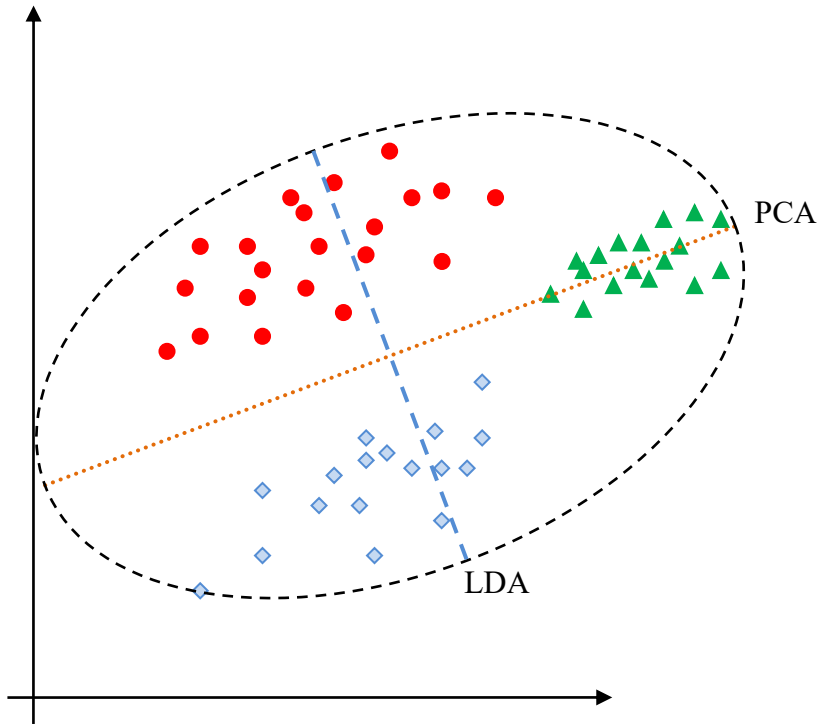
Dove  $\boldsymbol{\beta}$  rappresenta il vettore di dimensione  $d + 1$  dei parametri da determinare:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

In formato matriciale la funzione obiettivo da minimizzare (*Least Square*) può essere scritta come:

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

6) Date le distribuzioni riportate nel grafico sottostante, indicare graficamente le soluzioni ottenute (iperpiani) con gli algoritmi PCA e LDA per ridurre la dimensionalità dei pattern (da  $d = 2$  a  $k = 1$ ). Motivare la risposta.



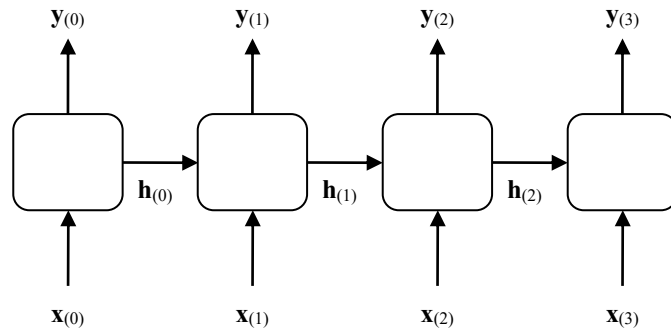
### Svolgimento

La tecnica PCA identifica l'iperpiano sul quale proiettando i pattern (indipendentemente dalla loro classe) si conserva al massimo l'informazione (linea puntinata arancione). In questo caso corrisponde all'asse principale di un'ellisse che comprende tutti i pattern.

La tecnica LDA identifica l'iperpiano sul quale proiettando i pattern si distinguono al meglio le tre classi (linea tratteggiata blu). Nella linea disegnata in figura le proiezioni dei pattern delle classi sono completamente separate.

7) Data una cella di rete ricorrente (RNN) con 100 neuroni e 500 pesi addestrabili, disegnare lo schema grafico del suo *unfolding in time* su 4 stati. Quanti neuroni e pesi addestrabili ha la rete *unfolded*? Giustificare la risposta.

**Svolgimento**



Il numero totale di neuroni è pari a 400 (100 neuroni per ogni stato) mentre il numero totale di pesi addestrabili è pari a 500 in quanto i pesi della cella sono comuni a tutte le sue istanze.