

1) Per l'addestramento di una rete neurale che cosa si intende con vettore di output desiderato? Come può essere definito? Come si può calcolare l'errore da retro-propagare a partire dal vettore desiderato e dal valore calcolato dalla rete per un pattern? Oltre alla spiegazione riportare un esempio.

Dispense "Reti Neurali"

2) Come è definita la funzione di attivazione Relu? Perché consente di addestrare reti neurali profonde limitando il problema del vanishing gradient?

Dispense "Deep Learning (parte 1)"

3) Descrivere a grandi linee l'approccio di classificazione AdaBoost.

Dispense "Classificazione (parte 2)"

4) Nell'ambito dei multi-classificatori quali sono le più comuni tecniche di fusione a livello di decisione e di confidenza?

Dispense "Classificazione (parte 2)"

---

5) Un classificatore *Nearest Neighbor* (NN), con un *training set* (TS) composto da  $n = 8000$  pattern di dimensionalità  $d = 7$ , utilizza come metrica la *distanza euclidea*. Calcolare il numero di somme, sottrazioni e moltiplicazioni necessarie (trascurando la radice quadrata) per effettuare la classificazione di un pattern  $x$  supponendo che non vengano utilizzate strutture spaziali specifiche per indicizzare il TS.

### **Svolgimento**

Per classificare un pattern è necessario calcolare le distanze da tutti gli  $n$  pattern del TS e selezionare la classe del pattern con distanza minima. Sapendo che la distanza euclidea tra due punti  $d$ -dimensionali è definita come  $\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_d - q_d)^2}$ , per calcolare ognuna delle  $n$  distanze saranno necessarie  $(d - 1)$  somme,  $d$  sottrazioni e  $d$  moltiplicazioni (elevamento al quadrato).

Con un TS composto da 8000 pattern 7-dimensionali saranno necessarie:

- $(7 - 1) \cdot 8000 = 48000$  somme;
- $7 \cdot 8000 = 56000$  sottrazioni;
- $7 \cdot 8000 = 56000$  moltiplicazioni.

6) Dato un training set composto dai seguenti pattern:

$$\mathbf{x}_1 = \begin{bmatrix} 3.4 \\ -7.2 \\ -7.5 \\ 0.5 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -6.6 \\ -4.7 \\ 4.5 \\ 0.3 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -6.5 \\ 5.2 \\ -5.3 \\ -4.9 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 0.1 \\ -3.6 \\ -2.4 \\ -4.8 \end{bmatrix}$$

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

$$y_1 = 0.3, \quad y_2 = -9.3, \quad y_3 = 6.3, \quad y_4 = 1.5$$

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:

$$\mathbf{X} = \begin{bmatrix} 3.4 & -7.2 & -7.5 & 0.5 & 1 \\ -6.6 & -4.7 & 4.5 & 0.3 & 1 \\ -6.5 & 5.2 & -5.3 & -4.9 & 1 \\ 0.1 & -3.6 & -2.4 & -4.8 & 1 \end{bmatrix}$$

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

$$\mathbf{y} = \begin{bmatrix} 0.3 \\ -9.3 \\ 6.3 \\ 1.5 \end{bmatrix}$$

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:

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

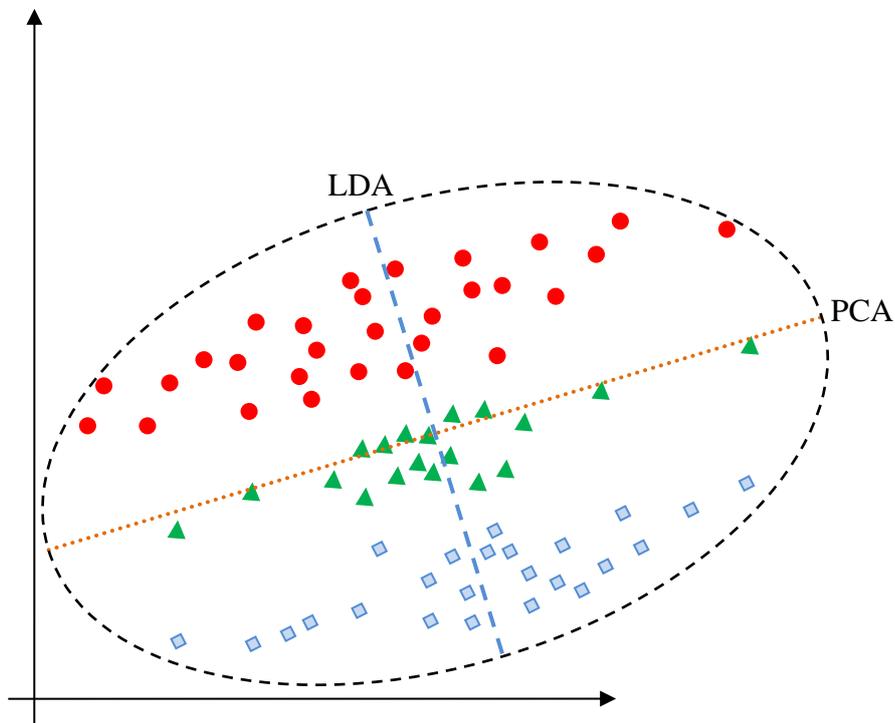
Dove  $\boldsymbol{\beta}$  rappresenta il vettore di dimensione  $d + 1$  dei parametri da determinare:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}$$

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

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

7) 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.