

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.

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

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

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

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 (trascuando la radice quadrata) per effettuare la classificazione di un pattern x supponendo che non vengano utilizzate strutture spaziali specifiche per indicizzare il TS.

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} .

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.

