

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

Dispense “Deep Learning (1)”

2) Che cosa si intende per clustering? Fare esempi di applicazioni.

Dispense “Clustering”

3) Nel caso di pattern non-linearmente separabili, nella formulazione di SVM lineare come si approccia il problema?

Dispense “Classificazione (2)”

4) Nell’ambito dell’apprendimento automatico quali sono le principali cause di overfitting?

Dispense “Fondamenti”

5) Un classificatore *Nearest Neighbor* (NN), con un *training set* (TS) composto da $n = 5000$ pattern di dimensionalità $d = 6$, utilizza come metrica la *distanza euclidea*. Calcolare il numero di somme, sottrazioni e moltiplicazioni necessarie 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 5000 pattern 6-dimensionali saranno necessarie:

- $(6 - 1) \cdot 5000 = 25000$ somme;
- $6 \cdot 5000 = 30000$ sottrazioni;
- $6 \cdot 5000 = 30000$ moltiplicazioni.

N.B.: essendo la radice quadrata una funzione monotona crescente, individuare la minima distanza euclidea equivale a individuare la minima distanza euclidea al quadrato pertanto il calcolo della radice può essere omesso.

6) Data una rete neurale MLP a 3 livelli senza bias composta da:

- 32 neuroni di Input
- 64 neuroni Intermedi
- 16 neuroni di Output

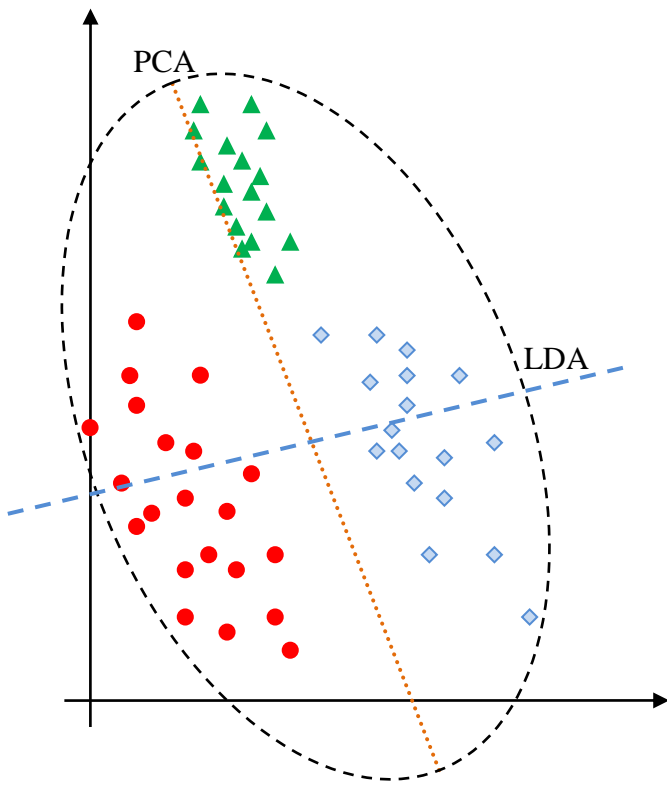
Calcolare, motivandone la risposta, il numero di pesi totale.

Svolgimento

Nel caso di una rete neurale MLP il numero di pesi è pari al numero di connessioni presenti. Il numero di connessioni (e quindi di pesi) presenti tra due livelli consecutivi (i e $i + 1$) si può calcolare come il prodotto del numero di neuroni del livello i per il numero di neuroni del livello $i + 1$. Nel caso dell'utilizzo del bias, il numero di neuroni di ogni livello i dovrà essere incrementato di uno.

Pertanto il numero totale di pesi sarà pari a: $32 \cdot 64 + 64 \cdot 16 = 3072$.

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.