


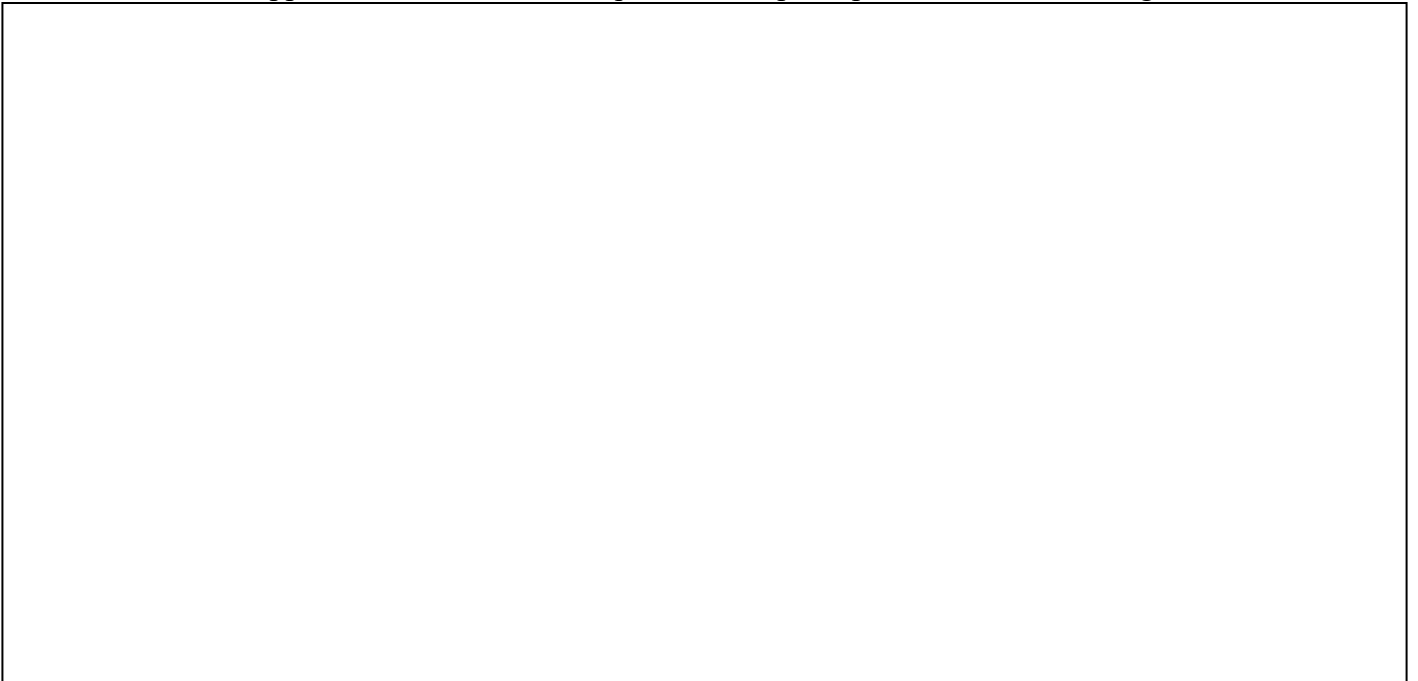
1) Come è definita la funzione di attivazione Relu? Perché consente di addestrare reti neurali profonde limitando il problema del vanishing gradient?

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

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



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



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.

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.

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.

