

First Name and Last Name:

University ID(Matricola):

Esercizio 1 (punti 5 su 30)

Disegnare un percettrone, eventualmente multilivello, con indicazione dei pesi e delle step function, che calcoli la funzione booleana $(A \text{ AND } B) \text{ XOR } (C \text{ OR } D)$.

Esercizio 2 (punti 6 su 30)

Data la seguente espressione:

$$2X^2Y + 3XY + 4$$

- Disegnare il grafo TensorFlow dell'espressione
- Scrivere il codice Python relativo al grafo del punto a), inizializzando la variabile X a 5 ed Y a 7, creando una sessione, valutando l'espressione ed infine chiudendo la sessione.

Esercizio 3 (punti 5 su 30)

Disegnare la struttura a lattice per l'estrazione di dipendenze funzionali da un dataset con 6 attributi, indicando il numero totale di dipendenze candidate.

Esercizio 4 (punti 7 su 30)

Data la seguente signature matrix:

Shingle	S ₁	S ₂	S ₃	S ₄	S ₅
0	1	1	0	0	0
1	0	1	1	1	1
2	0	0	0	1	0
3	0	1	0	0	1
4	1	1	0	1	1
5	0	0	1	1	0
6	1	0	1	0	1
7	0	0	1	1	0

- a) (4 punti) Calcolare la signature di ogni colonna usando le seguenti 6 funzioni hash:

$$h_1(x) = (4x^2 + 3) \bmod 8; h_2(x) = (2x^3 + 7) \bmod 8; h_3(x) = (3x^4 + 5) \bmod 8;$$

$$h_4(x) = (5x^2 - 2) \bmod 8; h_5(x) = (7x - 3) \bmod 8; h_6(x) = (6x^3 + 8) \bmod 8;$$

Mostrare l'evoluzione della matrice delle signature di minhash simulando l'esecuzione dell'algoritmo per il loro calcolo. Inoltre, calcolare le similarità di Jaccard tra tutte le coppie di signature di minhash.

- b) (3 punti) Sulla base delle similarità di Jaccard calcolate al punto b), suddividere la matrice delle signature in bande (2 da 3 o 3 da 2) in modo tale da ridurre la probabilità di falsi positivi nella selezione dei documenti da confrontare.

Esercizio 5 (punti 7 su 30)

Dati i seguenti punti in uno spazio bidimensionale:

(1, 2)(3, 7)(5, 14)(7, 5)(2, 10)(4, 3)(6, 2)

Usare l'algoritmo *K-means* per suddividere tali punti in 2 cluster, utilizzando la L3 norm e svolgendo 3 iterazioni dell'algoritmo. Per la scelta dei 2 centroidi, alla prima iterazione utilizzare il metodo che prevede la selezione del punto più vicino al centroide di tutti i punti forniti e quello più lontano da quest'ultimo.