

Cognome e Nome:

Matricola:

(Scrivere solo nello spazio bianco. Se necessario, usare il retro del foglio.

Non sono ammessi elaborati su fogli diversi.)

Esercizio 1 (punti 6 su 30)

Costruire un percettrone multilivello che calcoli l'espressione $(A \text{ XOR } B) \text{ OR } (C \text{ XOR } D)$.

Esercizio 2 (punti 7 su 30)

Dati i seguenti punti in uno spazio bidimensionale: $(1, 2)(1, 3)(3, 4)(3, 1)(1, 1)(4, 5)(2, 2)(2, 1)(4, 3)$

Mostrare i passi di un algoritmo di clustering gerarchico bottom-up (mostrando ad ogni passo cluster e centroidi) per raggruppare i suddetti punti in 3 cluster, usando la L3-norm.

Esercizio 3 (punti 6 su 30)

Dati i seguenti due schemi di database, di cui DB1 gestito dalla Federazione Italiana Gioco Calcio (FIGC) e DB2 gestito dalla Lega Calcio di serie A:

// DB1 FIGC

SQUADRE2021(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

CALCIATORE(N.Tessera, Cognome, Nome, Nazionalità, Data-Nascita, Statura, Ruolo)

// DB2 Lega Calcio

CALCIATORE(Num-Tessera, Codice Fiscale, Cognome, Nome, Email, Goal in carriera)

SQUADRE2022(Nome, Calciatore, Data-Inizio, Data-Fine, Ingaggio)

Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.

Esercizio 4 (punti 7 su 30)

Scrivere le funzioni Map e Reduce di un programma distribuito che legga in input il database integrato dell'esercizio 3 e dica per ogni squadra il numero di calciatori che hanno militato nella stessa.

Esercizio 5 (punti 4 su 30)

Supponiamo che un gestore di telefonia mobile abbia addestrato 3 classificatori per prevedere se i propri clienti stiano per cambiare gestore. I classificatori sono stati addestrati su un training dataset avente come attributi predittivi l'Età del cliente, gli anni di anzianità con il gestore e la cifra spesa per ricariche nell'anno precedente, e producono sia un output binario di tipo SI/NO per dire se si prevede o meno che il cliente cambi gestore, che una probabilità che cambi. In tabella sono riportati i risultati prodotti dai 3 classificatori su un campione di 3 clienti, indicare il risultato finale che verrebbe prodotto da due classificatori di tipo *Random Forest*, di cui il primo utilizzi uno schema *Hard Voting* ed il secondo uno schema *Soft Voting*.

Cliente	Cl1	Prob. CL1	Cl2	Prob. CL2	Cl3	Prob. CL3	Cambiato Gestore
1	NO	0,4	SI	0,6	NO	0,45	NO
2	SI	0,65	NO	0,35	SI	0,55	SI
3	NO	0,3	NO	0,35	NO	0,25	NO
4	SI	0,7	SI	0,75	SI	0,65	SI
5	NO	0,2	NO	0,25	SI	0,98	NO
6	SI	0,85	SI	0,66	NO	0,33	SI
7	SI	0,82	NO	0,48	NO	0,44	NO
8	NO	0,46	SI	0,64	NO	0,45	NO
9	NO	0,37	SI	0,72	SI	0,55	SI
10	SI	0,51	NO	0,11	SI	0,52	SI