

Cognome e Nome:

Matricola:

**Esercizio 1 (punti 7 su 30)**

Dato un database con 6 attributi A, B, C, D, E, F:

- a) (3 punti) Disegnare la struttura a lattice di un algoritmo di FD discovery column based;
- b) (3 punti) Indicare il numero di FD candidate di ciascun livello del lattice;
- c) (1 punto) Indicare come cambiano le risposte di cui ai punti a) e b) in caso di RFD discovery.

**Esercizio 2 (punti 6 su 30)**

Data la seguente confusion matrix:

TP = 501;      TN = 102;      FP = 108;      FN = 139;

Indicare la formula dei seguenti parametri e calcolarne il valore sulla confusion matrix di cui sopra:

*Accuratezza, True positive rate, True negative rate, Precision, Recall ed F1-measure*

**Esercizio 3 (punti 6 su 30)**

Data la seguente signature matrix:

Shingle	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	S <sub>4</sub>
0	1	1	0	1
1	0	1	1	0
2	1	0	0	1
3	0	0	1	0
4	0	0	1	1
5	1	0	0	0
6	1	1	0	1
7	1	1	1	1
8	0	1	1	0
9	0	1	0	0
10	1	0	1	0
11	0	1	0	0

- a) (3 punti) Calcolare la similarità di Jaccard tra ogni coppia di colonne;
- b) (3 punti) Se si suddivide la matrice in 3 bande di 4 righe ciascuna, calcolare per ogni coppia di signature la probabilità che esse vengano selezionate per il confronto.

**Esercizio 4 (punti 6 su 30)**

Dato il seguente frammento di dataset:

Istanza	$x_1$	$x_2$	$x_3$	$x_4$
i1	-1	2	-5	4
i2	3	2	5	-2
i3	7	-3	0	-6
i4	9	-1	-2	3
i5	4	0	-2	-1

Verificare se esistono valori dei 5 pesi  $w_i$  e del termine di bias che facciano in modo che un classificatore SVM lineare restituisca la classe positiva sulle prime 3 istanze e quella negativa sulle ultime 2.

**Esercizio 5 (punti 5 su 30)**

Dati i seguenti punti in uno spazio bidimensionale:

(2,1)(4, 2)(5,2)(4,3)(5,3)(2,5)(3,6)(8,6)(3,9)(2,11)(4,10)(3,12)

Mostrare i passi di un algoritmo di clustering gerarchico (mostrando ad ogni passo cluster e centroidi) per raggruppare i suddetti punti in 3 cluster, usando la funzione di distanza  $L_4$ -norm.