Fondamenti di Data Science e Machine Learning - Prof. G. Polese - Anno Accademico 2020/21 Prova Scritta (ore 9:00) 7/2/2022

First Name and Last Name:

University ID(Matricola):

Esercizio 1 (punti 6 su 30)

Data la seguente espressione:

$$3X^2Y + 2XY + 5$$

- a) Disegnare il grafo TensorFlow dell'espressione
- b) Scrivere il codice Python relativo al grafo del punto a), inizializzando la variabile X a 6 ed Y a 3, creando una sessione, valutando l'espressione ed infine chiudendo la sessione.

Esercizio 2 (punti 7 su 30)

Dati i seguenti punti in uno spazio bidimensionale: (2, 7)(5, 0)(3, 7)(9, 2)(4, 3)(6, 10)(8, 9)(1,8)

Usare l'algoritmo *K-means* per suddividere tali punti in 2 cluster, utilizzando la L2 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 poi quello più lontano dal punto così selezionato.

Esercizio 3 (punti 6 su 30)

Date la seguente tabella di transazioni di un negozio di ferramenta

Trans.ID	Data	Prodotto	Quantità
11	11/5/2021	Lampada	1
11	11/5/2021	Cacciavite	1
11	11/5/2021	Martello	1
12	12/5/2021	Trapano	1
12	12/5/2021	Lima	1
12	12/5/2021	Livella	1
13	15/5/2021	Lampada	1
13	15/5/2021	Cacciavite	1
14	18/5/2021	Trapano	1
14	18/5/2021	Lampada	1
14	18/5/2021	Martello	1
14	18/5/2021	Cacciavite	1

- a) Elencare i frequent item sets con l'algoritmo Apriori supporto di almeno il 50%
- b) Trovare tutte le regole di associazione con supporto e confidenza di almeno il 50%

Esercizio 4 (punti 6 su 30)

Dato il seguente Dataset:

Item	\mathbf{X}_{1}	\mathbf{X}_2	X_3	X_4	Y
1	1	-6	7	-14	0
2	-1	6	-13	11	1
3	2	-2	-6	12	0
4	-6	3	8	-9	1
5	10	18	-11	9	1

Dove X₁, X₂, X₃, X₄ sono gli attributi predittivi ed Y quello dipendente. Verificare l'esistenza ed eventualmente fornire la funzione di decisione di un classificatore SVM lineare che sul dataset in oggetto produca gli stessi valori per l'attributo dipendente.

Esercizio 5 (punti 5 su 30)

Si considerino 2 dataset con attributi A, B e C, D rispettivamente. Disegnare la struttura a lattice per l'estrazione di inclusion dependency ed indicare il numero totale di inclusion dependency candidate.