	National University of Computer and Emerging Sciences, *Lahore*
Course	Data Mining
Homework No.	HW 2
Total Marks	30
Start Date	
Due Date/Time	23 March 2024, 11:59 PM
Submission	Google Classroom

Note : You must submit it in handwritten form
Late submissions will not be accepted

Q1: 10 marks

Suppose that we have training data $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$ and we wish to predict y using a nonlinear regression model with two parameters:

$$\hat{y} = a + \exp(x_1 + b)$$

We decide to train our model using gradient descent on the mean squared error (MSE).

1) Write down the expression for the MSE on our training set.

2) Write down the gradient of the MSE

3) Give pseudocode for a (batch) gradient descent function theta = train(X,Y), including

   all necessary elements for it to work.

Q2: 5 marks

Table 6.22. Example of market basket transactions.

Customer ID	Transaction ID	Items Bought
1	0001	$\{a, d, e\}$
1	0024	$\{a, b, c, e\}$
2	0012	$\{a, b, d, e\}$
2	0031	$\{a, c, d, e\}$
3	0015	$\{b, c, e\}$
3	0022	$\{b, d, e\}$
4	0029	$\{c, d\}$
4	0040	$\{a, b, c\}$
5	0033	$\{a, d, e\}$
5	0038	$\{a, b, e\}$

Compute the support for itemsets {e}, {b, d}, and {b, d, e} by treating
each transaction ID as a market basket.

Q3: 5 marks

Given a transaction that contains items {1, 3, 4, 5, 8}, which of the hash tree leaf nodes will be
visited when finding the candidates of the transaction?
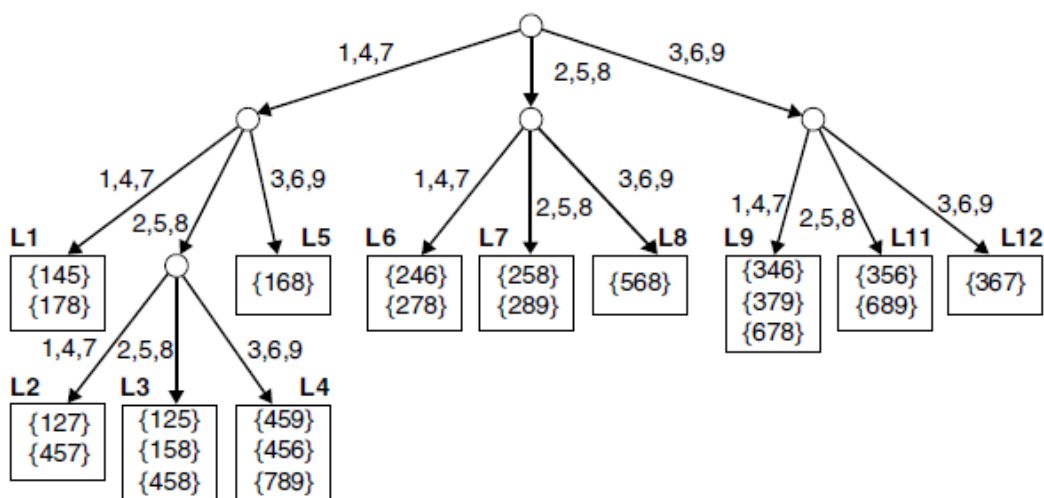
Transaction ID	Items Bought
1	$\{a, b, d, e\}$
2	$\{b, c, d\}$
3	$\{a, b, d, e\}$
4	$\{a, c, d, e\}$
5	$\{b, c, d, e\}$
6	$\{b, d, e\}$
7	$\{c, d\}$
8	$\{a, b, c\}$
9	$\{a, d, e\}$
10	$\{b, d\}$

Q4 : 10 marks

Given the lattice structure shown in the following Figure and the transactions given in Table label each node with the following letter(s):
• M if the node is a maximal frequent itemset,
• C if it is a closed frequent itemset,
• N if it is frequent but neither maximal nor closed, and
• I if it is infrequent.
Assume that the support threshold is equal to 30%.