

Tribhuvan University
Institute of Science and Technology
2079
☆

Bachelor Level / Forth Year / Seventh Semester / Science

Computer Science and Information Technology (CSC 410)

(Data Warehousing and Data Mining)

(NEW COURSE)

Full Marks :- 60

Pass Marks :- 24

Time :- 3 Hrs

Section A

Attempt any **TWO** questions

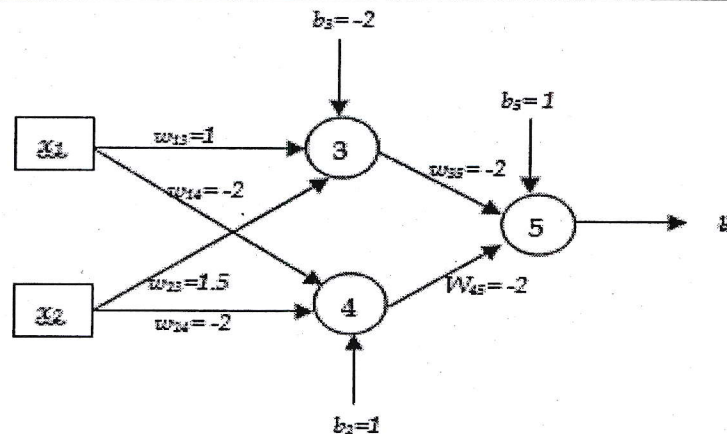
[2 × 10 = 20]

1. Discuss any two drawbacks of Apriori algorithm. Find frequent itemsets and association rules from the transaction database given below using FP-growth algorithm. Assume minimum support is 50% and minimum confidence is 60%. (2+8)

Transaction ID	Items Purchased
1	Sausage, Peanut, Beer
2	Peanut, Beer, Apple
3	Apple, Milk
4	Sausage, Peanut, Apple
5	Sausage, Peanut, Beer, Milk
6	Sausage, Peanut, Beer, Apple

2. When multilayer perceptron is better choice over other classification algorithms? Consider a multilayer feed-forward neural network given below. Let the learning rate be 0.5. Assume initial values of weights and biases as given in the table below. Train the network for the training tuples (1, 1, 0) and (0, 1, 1), where last number is target output. Show weight and bias updates by using back-propagation algorithm. Assume that sigmoid activation function is used in the network. (10)

W_{13}	W_{14}	W_{23}	W_{24}	W_{35}	W_{45}	b_3	b_4	b_5
0.5	0.2	-0.3	0.5	0.1	0.3	0.6	-0.4	0.8



3. Why OLAP operations are used? Discuss various OLAP operation with suitable example of each. (1+9)

Section BAttempt any **EIGHT** questions

[8× 5 = 40]

4. Suppose that we have 5 dimensional data. What will be total number of cuboids generated? If we consider each dimension has 5 levels, what will be the number of cuboids generated? (1+4)
5. Discuss different types of attributes with suitable example of each. (5)
6. Why data normalization is important in data mining? Explain min-max and Z-score normalization approach. (1+4)
7. What are two categories of hierarchical clustering? Divide the following data points into two clusters using agglomerative clustering. (1+4).
 $\{(2,10), ((2,5), (8,4), (5,8), (7,5), (6,4))\}$
8. Discuss the concept of K-means++ and Mini-batch K-means algorithm. (5)
9. What is confusion matrix? Discuss various classification measures along with their mathematical formulae. (5)
10. What are application areas of graph mining? Explain the concept behind inductive logic programming with suitable demonstration. (5)
11. Discuss the concept of text mining with its practical implications. (5)
12. Write down short notes on (2×2.5=5)
 - a. Data Mart
 - b. Market Basket Analysis