IHM2015005

Indian Institute of Information Technology, Allahabad
End term examination – May 2018

Degree Program: Btech / Mtech
Course Title: Data Mining and Warehousing
Date of Examination: 2.5.2018
Course Code: IDMW632C/IDMW~~620~~ E

Semester: 6th/2nd
Time duration: 3 hours
Total Marks: 75
Roll No: _____

Q.1) Prove that each entry in the following table correctly characterizes its corresponding rule constraint for frequent itemset mining. 10 marks

Rule constraint	Antimonotonic	Monotonic
$v \in S$	no	yes
$S \subset V$	yes	no
$min(S) \leq v$	no	yes
$range(S) \leq v$	yes	no

Q.2) What is hierarchical clustering, use category utility function for clustering of the following data.

Color	Nuclei	tails
White	1	2
White	2	2
Black	2	2
Black	2	1

10 marks

Q.3) Using suitable notations derive an expression for margin maximization in case of support vector machine. 10 marks

Q.4) Consider the following data and apply Bayesian classification to predict the class of unknown sample, $X = \{Refund = No, Married, Income = 120K\}$

1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

OR Write short note on:

- Lift
- Information gain
- Directly density reachable
- Gini index

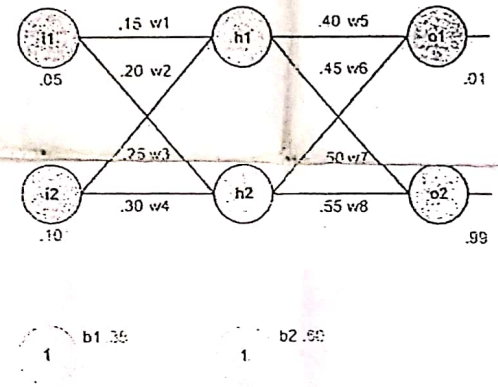
10 marks

Q.5) Most frequent pattern mining algorithms consider only distinct items in a transaction. However, multiple occurrences of an item in the same shopping basket, such as four cakes and three jugs of milk, can be important in transaction data analysis. How can one mine frequent itemsets efficiently considering multiple occurrences of items? Propose modifications to the well-known algorithms, such as Apriori and FP-growth, to adapt to such a situation. 10 marks

Q.6) Apply Back propagation algorithm for classification with a learning rate of 0.5. Given the net input I_j to unit j, then o_j, the output of unit j is o_j.

$$Activation\ Function = o_j = \frac{1}{1 + \exp^{-I_j}} \quad (1)$$

10 marks



Q.7) How Pearson coefficient can be useful for data reduction discuss with an example.

5 marks

Q.8) Compare ROLAP, MOLAP and HOLAP in terms of summarization, data size, aggregation(low or high), application etc.

5 marks

Q.9) Define overfitting in decision tree, how pruning is useful in such cases, discuss with examples. 5 marks

Roll no:

End of exam