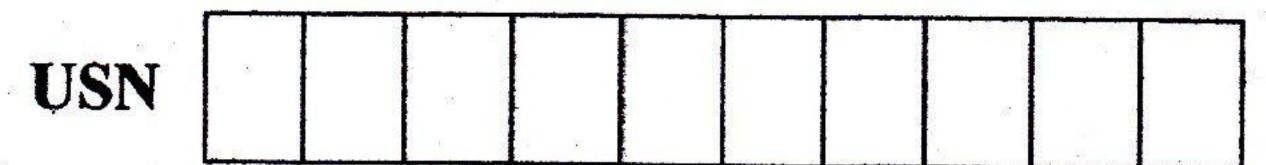
Important Note



Seventh Semester B.E. Degree Examination, Dec. 09/Jan.10

Data Mining

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions, selecting at least TWO questions from each part.

PART - A

Explain the process of knowledge discovery in databases (KDD).

(10 Marks)

Explain the various tasks of data mining.

(10 Marks)

Discuss the four types of data attributes, with suitable examples.

(08 Marks)

What are the various data preprocessing tasks?

(08 Marks)

Consider the following two binary vectors

X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)

Find: i) Hamming distance ii) Simple matching coefficient (SMC) iii) Jauard coefficient. (04 Marks)

Explain the various measures for selecting the best splits, with an example (08 Marks)

Discuss general-to-specific and specific-to-general rule growing strategies. Give suitable examples. (08 Marks)

Write the algorithm for k-nearest neighbor classification algorithm.

(04 Marks)

a. Define the following. Give an example to each.

i) Support of a rule

ii) Confidence of a rule.

(04 Marks)

State and illustrate using lattices the apriory principle with an example.

(08 Marks)

c. Which are the factors affecting the computational complexity of apriory algorithm? Explain them. (08 Marks)

PART - B

Consider the following transaction data set

Tid	1	2	3	4	5	6	7	8	9	10
Items	$\{a,b\}$	$\{b,c,d\}$	${a,c,d,e}$	{a,d,e}	$\{a,b,c\}$	$\{a,b,c,d\}$	{a}	$\{a,b,c\}$	{a,b,d}	{b,c,e}

Construct the FP tree. Show the trees separately after reading each transaction. (08 Marks) Illustrate the limitations of support -confidence frames work for evaluation of an association

rule.

(08 Marks)

Define cross-support pattern. Suppose the support for milk is 70%, support for sugar is 10% and support for caviar is 0.04%. Given $h_c = 0.01$, Is the frequent item set {milk, sugar, caviar} the cross-support pattern? (04 Marks)

Explain the different types of clustering.

(10 Marks)

Explain the basic K-means algorithm of clustering.

(10 Marks)

How can the generalization be performed on set-valued, list valued and sequence valued attributes? Give examples. (10 Marks)

Explain the following:

Description-based retrieval

Content based retrieval for similarity searching in multimedia data.

(10 Marks)

Write short notes on:

- Application of data mining to financial analysis
- Features used to assess data mining systems
- Forms of coupling between data mining systems and data basis /date warehousing systems.

* * * *

Basic measures for text retrieved.

(20 Marks)