MIT ACADEMY OF ENGINEERING

COURSE CODE: CS332T　　　　　　　　　　　　　　　**8 JUNE 2019**

TY BTECH SEMESTER - VI 2018 - 2019 RE - EXAMINATION
DEPARTMENT OF COMPUTER ENGINEERING
MACHINE LEARNING

TIME : 3 HOURS　　　　　　　　　　　　　**MAX MARKS : 100 MARKS**

TOTAL NO OF QUESTIONS: 5　　　　**TOTAL NO OF PRINTED PAGES: 3**

INSTRUCTIONS TO CANDIDATES:

1. Assume suitable data wherever necessary
2. Non programmable scientific calculators are allowed
3. Black figures to the right indicate full marks

1	**a)**	What is Machine learning? Explain structure and need of machine learning. Also explain any four examples of machine learning in detail. (Definition of Machine Learning – 2 mks, ML Structure - 2 mks, need of ML-2 mks, at least four example -2 mks each)	**[14]**	**CO1, CO2**	**L1**
	b)	Summarise the term Data storage and Data privacy. (Explanation Data storage, Data Privacy - 2 mks each)	**[04]**	**CO2**	**L2**
2	**a)**	Examine the role of prior probability, likelihood, evidence, Bayes rule and posterior probability in Naïve Bayes classifiers. (Description of each point with formulae -2 mks each)	**[10]**	**CO3**	**L4**
	b)	Elaborate the different classification algorithm with example. (at least three classification algorithm- 2 mks each)	**[06]**	**CO4**	**L6**
3	**a)**	Identify the requirements of clustering algorithms. Explain Hierarchical clustering algorithm. (at least four requirement – 4 mks, Explanation HC- 2 mks)	**[06]**	**CO3**	**L3**

b) Apply average link and single link algomerative clustering method on given data. Draw the dendogram.
Data: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).
(correct solution for average and single link -3 marks each, Dendogram -2 mks each)

[10]   CO3   L3

4   a) Construct the decision tree given dataset to buy_computer using ID3 algorithm. Use the maximum information gain criterion for splits, Determine the entropy and information gain of each attribute. Transform the decision tree to set of decision rule.
(Explanation of Decision tree:2 mks, Entropy and information gain formulae with correct calculations -8 mks, Correct decison tree construction- 2 mks, Decision rule – 2 mks)

[14]   CO4   L6

ID	Age	Income	Sutdent	Credit Rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

b) Illustrate the association rule? Relate support and confidence wrt association rule mining
(Defination association rule -2 mark, confidence - 1mark, support – 1 mark)

[04]   CO2   L2

c) You are given the transaction data from a fast food restaurant. There are 9 distinct transactions. (order:1 – order:9) and each transaction involves between 2 and 4 meal items. There are a total of 5 meal items that are involved in the transactions. Apply the Apriori algorithm to the dataset of transactions and identify all frequent k-itemsets. Discover all strong association rules of the form:X ∩ Y→ Z and note their support and confidence values.

(finding frequent itemset -6 mks, association rule -4 mks)

[10]   CO3   L4

Meal Order	List of Items
Order:1	I1, I2, I5
Order:2	I2, I4
Order:3	I2, I3
Order:4	I1, I2, I4
Order:5	I1, I3
Order:6	I2, I3
Order:7	I1, I3
Order:8	I1, I2, I3, I5
Order:9	I1, I2, I3

5 a) Inspect is the goal of the support vector machine (SVM). Explain soft and hard margin. How to calculate the distance of point apoint to the margin?
(goal of SVM -2 mks, soft & hard margin with example – 3 mks each, computation -4 mks)

[12]   CO3   L5

b) Explain the working of SVM. Evaluate the term hyperplane, separating hyperplane, margin and support vector with example.
(Working of SVM -3 mks, correct diagram-1 mark, terms evaluation -1 mark each, example -2 mks)

[10]   CO3   L5