B.E (FT) END SEMESTER EXAMINATIONS – NOV / DEC 2022

Computer Science and Engineering

Fifth Semester

CS6301 – MACHINE LEARNING

(Regulation 2018 - RUSA)

Time: 3 Hours | Answer ALL Questions | Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1.	Define machine learning.	2
2.	Relate the parameters and function of an artificial neuron to that of a biological neuron.	2
3.	Justify the need for multi-layer perceptron.	2
4.	Distinguish between RBF network and MLP. Write the two main differences.	2
5.	Justify the need for dimensionality reduction.	2
6.	What are the two main advantages of support vector machine?	2
7.	Identify and write the type of encoding which will be suitable if genetic algorithm is used for i) travelling salesman problem ii) feature selection.	2
8.	Distinguish reinforcement learning from the other three main types of learning.	2
9.	Justify the advantage of ensemble learning with respect to any one approach.	2
10.	Give two applications of vector quantization.	2

PART – B (8 x 8 = 64 marks)

(Answer any 8 questions)

11. i) Explain the steps involved in the design of a learning system. — 4

ii) Apply Candidate elimination algorithm on the following dataset and arrive at the version space: — 4

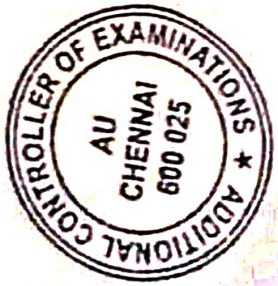
Example	Size	Color	Shape	Class/Label
1	Big	Red	Circle	No
2	Small	Red	Triangle	No
3	Small	Red	Circle	Yes

	4	Big	Blue	Circle	No		
	5	Small	Blue	Circle	Yes		

12.	i) Illustrate the training algorithm for single layer perceptron for NAND gate. ii) The sales of a company (in million dollars) for each year are shown in the table below. x (year) 2005 2006 2007 2008 2009 y (sales) 12 19 29 37 45 Find the least square regression line y = a x + b. Use the least squares regression line as a model to estimate the sales of the company in 2012. 70.4	4 4
13.	i) Write the algorithm for training and testing a multilayer perceptron. ii) How is training of a MLP done in practice for a classification problem?	4 4
14.	i) Explain the two basic concepts used in RBF network. ii) Write the algorithm for training an RBF network.	4 4
15.	i) Apply PCA on the following data:	6

X1	X2
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

ii) Compare and contrast LLE and Isomap. 2



16.	i) Apply naïve Bayes classifier to classify a Red Domestic SUV".	4

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

ii) What type of classifier is knn classifier? Justify. 4

17.	i) Explain the operators and its types in genetic algorithm.	4
	ii) Write the basic genetic algorithm.	4
18.	i) Explain Q-learning.	4
	ii) Distinguish between Q-learning and sarsa.	4
19.	i) Construct a decision tree by applying ID3 algorithm on the following dataset:	6

Color	Type	Doors	Tires	Class
Red	SUV	2	Whitewall	+
Blue	Minivan	4	Whitewall	-
Green	Car	4	Whitewall	-
Red	Minivan	4	Blackwall	-
Green	Car	2	Blackwall	+
Green	SUV	4	Blackwall	-
Blue	SUV	2	Blackwall	-
Blue	Car	2	Whitewall	+
Red	SUV	2	Blackwall	-
Blue	Car	4	Blackwall	-
Green	SUV	4	Whitewall	+
Red	Car	2	Blackwall	+
Green	SUV	2	Blackwall	-
Green	Minivan	4	Whitewall	-

ii) Suggest two methods for building an ensemble classifier using decision trees. 2

20.	i) Suppose we have 4 types of medicines and each has two attributes as shown below:	4

Medicine	Weight	pH
A	1	1
B	2	1
C	4	3
D	5	4

Group these into 2 groups using k-means clustering.

	iii)	Explain the principle of self organizing map.	4
21.	i) ii)	Explain the need for Deep learning networks and RNN. Explain the working of a CNN with neat diagrams.	4 4
22.	i) ii)	Explain the principle of Gaussian mixture model. Write the algorithm for Gaussian mixture model EM algorithm.	8
		PART – C (2 x 8 = 16marks)	
23.	i) ii)	What are steps involved in classification of a heart disease dataset with large number of features? Design a model for the same. Suggest other possible techniques and explain a metric using which you will compare the techniques.	4 4
24.	i) ii)	What are steps involved in classification of Pima Indians diabetes dataset that has missing values? Design a model for the same. Define and explain four metrics against which you will compare the classifier with another classifier.	4 8