C   1100AMT305122202   Pages: 5

Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Fifth Semester B.Tech Degree Regular and Supplementary Examination December 2022 (2019 Scheme)

Course Code: AMT 305

Course Name: INTRODUCTION TO MACHINE LEARNING

Max. Marks: 100   Duration: 3 Hours

PART A

(Answer all questions; each question carries 3 marks)   Marks

1	Distinguish between classification and regression.	3
2	What is meant by k-fold cross validation. Given a dataset with 1600 instances, how the k-fold validation is done with k = 10.	3
3		3

Actual	Predicted
4	5
3	4
15	19
7	10
9	12
20	21

Use the above dataset to calculate, mean square error, mean absolute error and root mean square error

4	Write a short note on logistic regression.	3
5	Draw the activation function ReLu. What are the advantages of using ReLu over a sigmoid function.	3
6	Distinguish between bootstrapping and cross validation.	3
7	What is a meant by maximum margin hyperplane?	3
8	What are the disadvantages of using SVM.	3
9	Explain any three different ways to measure the distance similarity while using k-means clustering algorithm.	3
10	Compare partitioning clustering and hierarchical clustering.	3

PART B

(Answer one full question from each module, each question carries 14 marks)

Module -1

11	a)	Classify the different types of machine learning.	8
	b)	Summarize the various applications of machine learning.	6

12 a) Write a short note on hypothesis space, version space, most specific and most general hypothesis. 6

b) Define VC dimension? Show that an axis aligned rectangle can shatter 4 points in 2 dimensions. 8

Module -2

13 a) The below dataset contains features such as Colour, Type and Origin of cars and the class whether it is stolen, not stolen. 7

Example No	Colour	Type	Origin	Whether 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

Use Naïve Bayes algorithm to determine whether a red domestic SUV car is stolen or not.

b) Use the following data to construct a linear regression model for an auto insurance premium as a function of driving experience. 7

Driving experience (in years)	5	2	12	9	15	6	25	16
Monthly auto insurance premium ($)	64	87	50	71	44	56	42	60

14 a) Explain the procedure to reduce the dimensionality of a dataset using principal component analysis. 6

b) Given the following data on a certain set of patients seen by a doctor, can the doctor conclude that the person having chills, fever, mild headache and without running nose has a flu? 8

Chills	Running Nose	Headache	Fever	Has Flu
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

Construct a decision tree using ID3 algorithm and predict the class for the above case.

Module -3

15 a) Consider a two-class classification problem of predicting whether a photograph contains a man or a woman. Suppose we have a test dataset of 10 records with expected outcomes and set of predictions from our classification algorithm.   8

	Expected	Predicted
1	Man	Woman
2	Man	Man
3	Woman	Woman
4	Man	Man
5	Woman	Man
6	Woman	Woman
7	Woman	Woman
8	Man	Man
9	Man	Woman
10	Woman	Woman

Compute the confusion matrix for the data. Also, compute the accuracy, precision, recall, specificity and f-score.

b) Compute the output of the following neuron if the activation function is,   6

    i)   ReLu

    ii)   Sigmoid

    iii)   Tanh.
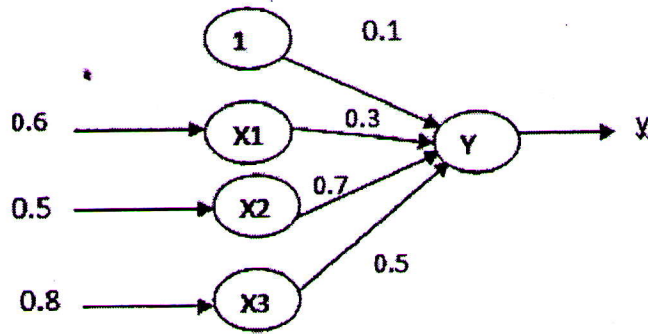
16 a) Given the following data, construct the ROC curve of the data. Compute the   8
        AUC.

| Threshold	TP	TN	FP	FN
1	0	25	0	29
2	7	25	0	22
3	18	24	1	11
4	26	20	5	3
5	29	11	14	0
6	29	0	25	0
7	29	0	25	0

    b) Draw a multilayer feed forward network. Explain how learning happens in the   6
        training phase of a neural network.

Module -4

17 a) Given a set of positively labelled data points { (6,2), (6,-2), (12,2), (12,-2) } and   8
        negatively labelled data points { (2,0), (0,2), (0,-2), (-2,0) }. Identify the
        maximum margin hyperplane. Also, sketch the maximum margin hyper plane
        using a graph.

    b) Explain kernel trick in the context of support vector machine. List any two   6
        Kernel function used in SVM.

18 a) How to estimating the parameters of a probability distribution using the general   6
        MLE method for

    b) State the mathematical formulation to express Soft Margin as a constraint   8
        optimization problem.

Module -5

19 a) By using k-means clustering algorithm divide the below data points into three   8
        groups by using the Euclidean distance. Assume the initial centroids are (1,2),
        (8,9) and (18,18).
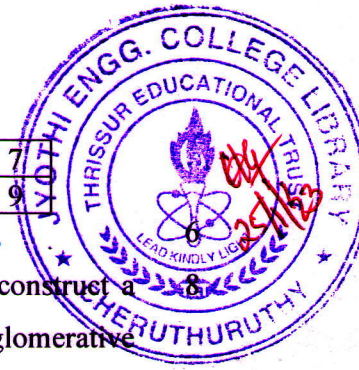
X	16	2	1	17	18	9	3	8	2	19	3	7
Y	17	2	2	19	18	10	3	9	3	17	2	9

b) Explain any two ensemble methods used for learning.

20 a) Given the dataset {A, B, C, D, E} and the below distance matrix, construct a dendrogram by single-linkage hierarchical clustering using the agglomerative method.

	A	B	C	D	E
A	0	9	3	6	11
B	9	0	7	5	10
C	3	7	0	9	2
D	6	5	9	0	8
E	11	10	2	8	0

b) Illustrate any one density-based clustering algorithm.   6

***