C **1100AMT305122301** **Pages: 3**

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Fifth Semester B.Tech Degree Regular and Supplementary Examination December 2023 (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 Define VC dimension and shattering of a dataset. (3)

2 What is meant by most specific and most general hypothesis? (3)

3 Is Logistic regression a classification model? Justify your answer. (3)

4 Formulate Information gain and entropy used in decision tree classifiers. (3)

5 Calculate F measure from the following table. Consider Healthy as positive class. (3)

Classification		Actual	
		Deceased leaf	Healthy leaf
Prediction	Healthy leaf	45	18
	Deceased leaf	12	25

6 What is the need of validation data while training a model? (3)

7 Conceptualise Maximum A-priori Estimation. (3)

8 Derive the equation for margin between support vectors in SVM? (3)

9 Point out the difference between voting and bagging (3)

10 Explain the terms used to define high density areas in density based clustering algorithms? (3)

PART B

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

Module 1

11 a) Define machine learning. With suitable example, describe the various machine learning paradigms. (14)

12 a) Define the terms model selection and generalisation (8)

b) Show that an Axis aligned rectangle can shatter only 4 point in two dimension. (6)

Module -2

13 a) Describe the feature selection method of dimensionality reduction. (6)

b) Consider the below frequency table which shows the frequency of each of the attributes. Apply Naïve Bayes algorithm to predict the fruit type of {Yellow, Sweet, Long}. (8)

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Other fruits	50	100	50	150
Total	800	850	400	1200

14 a) Describe the different regression models and formulate any one error measurements used in regression analysis. (9)

b) State ID3 algorithm, used for decision tree classification. (5)

Module -3

15 a) Define Sensitivity, Specificity, Accuracy and F- measure. Estimate these measures from the following table. The table describes the output of some classification algorithm. Consider 'Rainy Day' as positive class. (10)

Sl. No	Actual Value	Predicted Value
1	Rainy Day	Rainy Day
2	Sunny Day	Sunny Day
3	Sunny Day	Sunny Day
4	Rainy Day	Rainy Day
5	Sunny Day	Sunny Day
6	Sunny Day	Sunny Day
7	Rainy Day	Sunny Day
8	Sunny Day	Rainy Day
9	Sunny Day	Sunny Day
10	Rainy Day	Sunny Day

b) What is the significance of ROC curve in machine learning? (4)

16 a) Describe the working of multilayer perceptron. (8)

b) Draw the perceptron for the following data and calculate the output of the perceptron if the following activations are used? (6)

X1=1	W1=0.7
X2=0	W2=(-0.3)
X3=1	W3=0.4
X4=1	W4=(-0.8)
X5=0	W5=0.5

a) Sigmoid activation fuction

b) TanH activation function

Module -4

17 a) Suppose the dataset (x1,x2,...xn) are independent and identically distributed drawn from exponential distribution exp(λ). Find the maximum likelihood of λ (9)

b) Describe the Bias-variance decomposition. (5)

18 a) Describe soft-margin SVM with neat diagram. (8)

b) Define any 3 kernal functions used for non-linear SVM (6)

Module -5

19 a) Use K-means clustering algorithm for the following data of heights and weights. Choose the first two data points as initial cluster centres. Use Euclidean distance measure and identify the cluster centre in the second iteration. (8)

Height	185	170	168	179	182	188	180	180	183	180	180	177
Weight	72	56	60	68	72	77	71	70	84	88	67	76

b) Illustrate the terms (6)

a) Single linkage b) Complete linkage c)Centroid linkage

20 a) Describe Density based clustering in detail. (8)

b) State any 3 similarity measures used in clustering algorithms (6)

***