Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S5 (S,FE) Examination May 2025 (2019 Scheme)

Course Code: AMT305

Course Name: INTRODUCTION TO MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

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

1	Mention the methodology for implementing a K-class problem.	3
2	Illustrate the concept of VC dimension.	3
3	Mention the basic idea behind naïve bayes classification algorithm	3
4	Differentiate between regression and classification.	3
5	Explain the concept of bootstrapping.	3
6	Express the meaning of corner points in an ROC curve.	3
7	What is bias variance trade off	3
8	Give the equation for soft margin classification and the constraints on it.	3
9	Explain the concept of bagging	3
10	Write any 3 distance measures between data points.	3

PART B

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

Module -1

11	a)	Express the concept of machine learning. Differentiate between various machine learning paradigm?	14
12	a)	Determine the hypothesis space H and version space with respect to the following data D.	4

X	20	32	122	231	32	235	234	23	65	323

Y	0	0	1	1	0	1	1	0	0	1

b) Write in detail about model selection and generalization 6

c) Write a note on the impact of noise in dataset on model creation. 4

Module -2

13 a) Explain the ID3 algorithm. Also discuss about entropy and information gain. 6

b Calculate the information gain of attributes humidity and wind in the 'Play tennis' dataset given below 8

Day	outlook	temperature	humidity	wind	playtennis
D1	sunny	hot	high	weak	no
D2	sunny	hot	high	strong	no
D3	overcast	hot	high	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	no
D7	overcast	cool	normal	strong	yes
D8	sunny	mild	high	weak	no
D9	sunny	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D14	rain	mild	high	strong	no

14 a) Explain the concept behind gradient descent method of optimization. 6

b) Describe the subset selection methods for dimensionality reduction. 8

Module -3

15 a) Define various performance metrics of a machine learning model 6

b) Calculate accuracy, recall and precision from the following data. 3

Confusion Matrix		Actual	
		Yes	No
Predict	Yes	50	30
	No	20	180

c) What is cross validation and mention the benefits of using cross validation in a machine learning model? 5

16 a) Explain the principle behind neural networks. Illustrate the weight optimization procedure in a backpropagation network with an example. 14

Module -4

17 a) Assume that the data follows univariate normal distribution with unknown mean and known variance. 6

Find the Maximum likelihood for the parameter mean. given the data are independent and identically distributed.

b) Find the maximum a posterior estimate for the parameter mean in the above problem.(Prior probability of the mean itself is normal distribution with mean γ and variance β^2) 8

18 a) State the mathematical formulation of the SVM problem 8

b) How do you find a hyperplane for non-linear data in SVM? 6

Module -5

19 a) Write the algorithm for K-means clustering 6

b) Consider 10 points A1(2,5), A2(2,3), A3(4,5), A4(4,2),A5(6,5), A6(8,9), A7(2,1), A8(4,3), A9(1,6) and A10(3,7). 8

Find 3 clusters after 2 epochs considering A1, A4 and A6 as the initial cluster centres. Use Euclidean distance as the distance function.

20 a) Describe Expectation Maximization algorithm for Gaussian mixture 14

***