

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
B.Tech Degree S5 (S,FE) Examinations June 2026 (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 | Define VC (Vapnik-Chervonenkis) Dimension with a simple example. | 3 |
| 2 | How does semi-supervised learning bridge the gap between supervised and unsupervised learning? | 3 |
| 3 | Define supervised learning. Name special cases of supervised learning depending on whether the inputs/outputs are categorical, or continuous. | 3 |
| 4 | List the key assumptions of logistic regression. | 3 |
| 5 | Explain the different types of layers in ANN. | 3 |
| 6 | Explain the purpose of the activation function in a neural network. | 3 |
| 7 | Why MAP estimation might be preferred over MLE. | 3 |
| 8 | Differentiate between Hard margin and soft margin SVM. | 3 |
| 9 | Calculate the dissimilarity between two data points $x_1(3,2,4)$ and $x_2(5,3,5)$ Using (a) Euclidean distance (b) Manhattan distance | 3 |
| 10 | Write down the major differences between K-means clustering and hierarchical clustering. | 3 |

PART B

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

Module -1

- 11 a) Analyze the trade-off between model complexity and generalization in the context of VC dimension. Support your answer with mathematical reasoning. 6
- b) Determine the hypothesis space H and version space with respect to the following data D. 8

x	27	15	23	20	25	17	12	30	6
class	1	0	1	1	1	0	0	1	0

And also write a short note on hypothesis space and version space.

- 12 a) Explain the concept of PAC learning. Derive an expression for PAC learning in such a way that the selected function will have low generalized error. 6

Given the Hypothesis class H with VC dimension $d = 3$, Desired accuracy $\epsilon = 0.1$, Confidence parameter $\delta = 0.05$. Calculate the following

- b) i) The minimum number of training examples needed for PAC learning.
 ii) If the training error is 0.05, what is the maximum possible true error with 95% confidence?

Module -2

- 13 a) Calculate least square regression line using matrix method for the following data points:
 $x: [1, 2, 3, 4, 5]$
 $y: [2.1, 4.0, 6.3, 8.1, 9.9]$ 8
- b) Explain the procedure to reduce the dimensionality of a dataset using Principal Component Analysis. 6
- 14 a) Explain the concept of Information Gain in ID3 decision trees. 6
- b) Identify the first splitting attribute for decision tree by using ID3 algorithm with the following dataset. 8

Input			Output
Age	Income	Gender	Risk
<25	>50	M	High
<25	>50	F	High
>=25	<50	F	High
>=25	>50	F	Low
>=25	>50	M	Low
<25	<50	M	High
<25	<50	M	High
<25	<50	F	High

Module -3

- 15 a) Given the following multi-class confusion matrix for a fruit classification system: 8

Actual Value	Predicted Value		
	Apple	Orange	Banana
Apple	120	15	5
Orange	10	95	5
Banana	5	10	85

Calculate Sensitivity, Specificity, Accuracy and F- measure.

- b) Define the following performance metrics: 6
- i) Precision ii) Recall iii) F1-score iv) Accuracy
- 16 a) Describe the architecture of a multilayer feedforward neural network. 7
- b) Explain the backpropagation algorithm and how it is used to train the network. 7

Module -4

- 17 a) Explain Special cases of MLE. 5
- b) Estimate the parameter ' λ ' of Poisson density function using MLE method. 9
- 18 a) Describe the process of finding the maximum margin hyperplane in SVM classification. 7
- b) Explain how the kernel trick allows SVMs to handle non-linearly separable data. 7
- List any four types of Kernels used in SVM.

Module -5

- 19 a) Explain DBSCAN algorithm for density based clustering. List out its advantages compared to K-means. 6
- b) Apply K-means algorithm to partition the below dataset into two clusters. 8
- A(1,1), B(2,3), C(4,6), D(2,4), E(5,3).
- Take the initial centroid of two clusters as (1, 1) and (2, 4).
- Use Euclidian distance metric.
- Perform up to a maximum of one relocation. Show your step by step work.
- 20 a) Describe the random forest algorithm to improve classifier accuracy. 6
- b) For the given data points, construct the dendrogram using Complete Linkage method. 8

Item	A	B	C	D
A	0	2	6	8
B	2	0	5	7
C	6	5	0	3
D	8	7	3	0
