Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S6 (Hons.) Examinations April 2026 (2023 Admn)

Course Code: CST396

Course Name: ADVANCED TOPICS IN MACHINE LEARNING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 3 marks. Marks

1		What are the key differences between supervised, unsupervised, semi-supervised, and reinforcement learning?	(3)
2		Explain the least squares method in linear regression.	(3)
3		What is clustering in machine learning, and how does it differ from classification?	(3)
4		Describe the K-means clustering algorithm step by step.	(3)
5		Define precision, recall, accuracy, and F1-score.	(3)
6		What is the problem of overfitting in machine learning?	(3)
7		Define the Probably Approximately Correct (PAC) learning model.	(3)
8		What is the Vapnik-Chervonenkis (VC) dimension, and why is it important in learning theory?	(3)
9		What is a Bayesian Belief Network (BBN)?	(3)
10		What is Gibbs sampling.	(3)

PART B

Answer one full question from each module, each carries 14 marks.

Module I

11	a)	What are common criteria used to split nodes in a decision tree. Explain in detail.	(8)
	b)	Given that Hours studied (X) = [1,2,3,4,5] Exam score (Y) = [50,55,65,70,75]. Fit a linear regression model and use it to predict the score for 6 hours of study.	(6)

OR

12	a)	Distinguish between classification and regression.	(4)

b) Construct a decision tree to predict whether someone will play tennis based on weather conditions. (10)

Outlook	Temperature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Module II

13 a) Given the following probabilities: (9)
Prior probabilities: P(Spam) = 0.4, P(NotSpam) = 0.6
Conditional probabilities: P(Free|Spam) = 0.6, P(Free|NotSpam) = 0.3, P(Win|Spam) = 0.7, P(Win|NotSpam) = 0.2, P(Prize|Spam) = 0.8, P(Prize|NotSpam) = 0.5, Classify the sentence: "Win free prize now" as Spam or Not Spam using the Naive Bayes classifier.

b) What are the advantages and disadvantages of hierarchical clustering compared to partitional clustering? (5)

OR

14 a) What are the limitations of the K-means algorithm? (4)

b) What is a Gaussian Mixture Model and how does it differ from K-means clustering? (10)

Module III

15 a) What is the role of sampling in the construction of a Random Forest? List different types of sampling. (4)

b) Construct a confusion matrix for the data and compute the accuracy, precision and recall of the data. (10)

Email	Actual	Predicted
1	Spam	Spam
2	Spam	Not Spam
3	Not Spam	Not Spam
4	Not Spam	Spam
5	Spam	Spam

OR

16 a) How can error estimation techniques guide model selection and hyperparameter tuning? (8)

b) What is k-fold and stratified k-fold cross-validation? How do they differ? (6)

Module IV

17 a) In PAC learning, what do the terms "probably" and "approximately" refer to? (6)

b) Explain with examples (8)

(a) Hypothesis Space

(b) Version Space

(c) Finite Hypothesis Space

(d) Infinite Hypothesis Space

OR

18 a) Explain how the sample complexity differs for finite vs. infinite hypothesis spaces. (6)

b) Illustrate the idea of sampling complexity of PAC using suitable diagrams. (8)

Module V

19 a) How Gibbs sampling work within the MCMC framework? (7)

b) What is an Autoencoder, and how does it differ from a PCA? Explain. (7)

OR

20 a) How does a Variational Autoencoder (VAE) differ from a standard Autoencoder? (7)

b) What is a clique tree (or junction tree), and how does it help with inference in complex graphs? (7)

****