Name: Enrolment No:	UPES UNIVERSITY OF TOMORROW

UPES
End Semester Examination, May 2024

Course: PATTERN RECOGNITION AND ANOMALY DETECTION   Semester: VI
Program: BTECH (CSE+AI&ML-H/NH)   Time   : 03 hrs.
Course Code: CSAI3011   Max. Marks: 100

Instructions: Read all questions carefully.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q1	Describe "pattern"? Briefly discuss applications of Pattern Recognition.	4	CO1
Q2	Define and explain the concept of kernels in SVM, including linear, polynomial, radial basis function (RBF), and sigmoid kernels. Discuss the mathematical formulation of each kernel and their respective characteristics in mapping input data to higher-dimensional feature spaces.	4	CO3
Q3	Explain the steps in "k-means clustering" using a suitable illustration. What are the limitations of this method?	4	CO4
Q4	Explain the different types of anomalies that can occur in data and provide examples illustrating each type.	4	CO4
Q 5	Discuss the significance of false positives and false negatives in different domains, such as medical diagnosis, spam email detection, and fraud detection.	4	CO1
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	Differentiate between Bayesian networks and Markov networks, highlighting their key characteristics, modeling assumptions, and applications.	10	CO3
Q 7	Explain the various types of point anomaly detection techniques used in machine learning and data analysis, highlighting their underlying principles, advantages, and limitations.	10	CO4
Q 8	A researcher investigates whether a new energy drink enhances productivity by conducting a study with 50 employees who consume the drink daily for two weeks. Productivity scores before and after consumption are recorded. With a significance level (α) of 0.05, the null hypothesis (H0) assumes no difference in productivity, while the alternative hypothesis (H1) suggests increased productivity. Before consumption, mean productivity is 75 units with a standard deviation of 10 units, and after consumption, it is 80 units with a standard deviation of 12 units.	10	CO2

	Perform a hypothesis test to determine if there's adequate evidence to support the claim that the energy drink boosts productivity, assuming normally distributed scores. Using a critical value of approximately 1.676. **OR** Investigate the effectiveness of sparse kernel machines for classification tasks, focusing on their ability to handle high-dimensional data efficiently.		
Q9	Discuss the various types of input data encountered in machine learning tasks and their respective characteristics, challenges, and preprocessing requirements.	**10**	**CO1**

SECTION-C
(2Qx20M=40 Marks)

Q10	Consider a Bayesian network modeling student performance in an exam with variables: Study time (S), Previous exam score (P), Sleep hours before the exam (H), Stress level (T), and Exam performance (E). • Draw the graphical model of the Bayesian network, depicting nodes for each variable and directional edges showing causal relationships. • Explain the conditional dependencies in the network, describing how each variable influences or is influenced by others. • Derive conditional probability distributions (CPDs) for each variable based on the network structure, providing necessary probabilities for each node. • Given observed evidence (S=High, P=Good, H=Normal, T=Moderate), calculate the posterior probability distribution of exam performance (E) using the Bayesian network. **OR** There are 800 samples with 500 in Class A and 300 in Class B. Researchers plan to address class imbalance using random sampling, stratified sampling, and oversampling. • How many samples from each class should researchers randomly select to create a balanced dataset? • In a 60:40 split for training and testing sets via stratified sampling, how many samples from each class should be allocated to the training and testing sets to maintain the original class distribution? For oversampling to match Class B's size to Class A's, how many times should each Class B sample be replicated?	**20**	**CO5**
Q11	Write a short note on: • Shallow neural networks (collaborative filtering) • Multilayer perceptron (deep neural networks). • Recurrent Neural Network (RNN). • Long Short-Term Memory (LSTM).	**20**	**CO3**