Name: Enrolment No:	

UPES

End Semester Examination, December 2024

Programme Name : BTech-CSE-MINOR
Semester : VII
Course Name : Pattern Recognition and Anomaly Detection **Time: 03 hrs**
Course Code : CSAI4004P **Max. Marks: 100**
Nos. of page(s) : 2

SECTION A **(4 Marks *5 = 20 Marks)**
Answer all questions

S. No.		Marks	CO
Q 1	Compare two continuous and discrete probability functions with equation and example.	**4**	**CO1**
Q 2	Discuss the roles of regularization, cross-validation, and early stopping in combating overfitting.	**4**	**CO1**
Q 3	What is model combination in pattern recognition, and why is it useful?	**4**	**CO1**
Q 4	List two applications each for quota sampling and stratified sampling in market research.	**4**	**CO2**
Q 5	Explain the concept of Sparse Kernel Machines.	**4**	**CO1**

SECTION B **(10 Marks *4 = 40 Marks)**
Answer all questions

Q 6	Describe RBF Kernel and explain how the kernel trick allows algorithms like Support Vector Machines (SVMs) to operate in high-dimensional spaces without explicitly transforming the data.	**10**	**CO2**
Q 7	You are given the following data for a classification model: • True Positives (TP): 45 • True Negatives (TN): 40 • False Positives (FP): 10 • False Negatives (FN): 5 Calculate the following metrics: o Precision o Recall o F1-Score o Accuracy Show detailed calculations and interpret the results for each metric. Explain the roles of **precision** and **recall** in evaluating classification models. 1. Discuss the significance of the **F1-score** in balancing precision and recall, particularly in the case of imbalanced datasets. 2. Provide an example of a real-world application where: o Precision is more important than recall. o Recall is more important than precision. o F1-score is critical for model evaluation. **Instructions**: • Justify your answers with examples or scenarios.	**10**	**CO1**

	• Use clear mathematical and practical reasoning.		
Q8	Compare and contrast Bayesian Networks and Markov Networks. • Discuss the similarities and differences between Bayesian Networks and Markov Networks in terms of directed vs. undirected graphs, modeling dependencies, and inference. Provide examples of scenarios where one network might be preferred over the other.	**10**	**CO1**
Q 9	You are given a dataset with the following 2 data points and a Gaussian Mixture Model with 2 components (k=2): • **Component 1:** $\mu_1 = [2, 3], \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \pi_1 = 0.6$ • **Component 2:** $\mu_2 = [5, 8], \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \pi_2 = 0.4$ • Data Points: $\mathbf{x}_1 = [3, 4], \mathbf{x}_2 = [6, 9]$ **Tasks:** 1. Calculate the probability density $p(x_i \mid \mu_k, \Sigma_k)$ for each component and each data point using the multivariate Gaussian formula. 2. Compute the responsibilities (r_{ik}) for both data points and components. 3. Interpret the results of the responsibilities.	**10**	**CO2**
SECTION C (20 Marks * 2 = 40 Marks) **Answer all questions**			
Q 10	Use Case: Fraud Detection in Credit Card Transactions • Consider a scenario where a bank wants to detect fraudulent credit card transactions. The bank has historical data with features such as transaction amount, location, merchant category, and time of transaction. a. Modeling the Data: • Describe the steps you would take to preprocess the data for anomaly detection. What are the challenges associated with this dataset? • Which are the possible techniques would you choose to detect anomalies in this context? Give your preference among those techniques with justification. • Provide an example of how false positives might be costly in a fraud detection scenario and how to optimize the trade-off. Discuss the potential role of unsupervised anomaly detection methods for cases where labeled data is not available.	**20**	**CO1**
Q 11	a) Describe the concept of a Gaussian Mixture Model (GMM). • Explain how GMM models data using Gaussian distributions. • Define its key components: mean vector, covariance matrix, and mixing coefficients. b) Derive the likelihood function for GMM and explain how the Expectation-Maximization (EM) algorithm is used to optimize it. • Include the mathematical formulation of the E-step and M-step.	**20**	**CO2**

	• Provide an intuitive explanation of how the algorithm iteratively improves the model. c) Highlight the main differences between GMM and K-Means clustering		