Reg. No.												

Question Paper Code	13742

M.E - DEGREE EXAMINATIONS, APRIL / MAY 2025

Second Semester

Big Data Analytics

24PBDPC203 – MACHINE LEARNING TECHNIQUES

Regulations - 2024

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	State the Law of Total Probability.	2	K2	CO1
2.	What is the primary goal of decision theory in machine learning?	2	K2	CO1
3.	Give an example of a linear model used for classification.	2	K2	CO2
4.	How does information flow in a feed-forward network?	2	K2	CO2
5.	How does Factor Analysis differ from Principal Component Analysis in its fundamental objective?	2	K2	CO3
6.	Differentiate between feature selection and feature extraction as dimensionality reduction methods.	2	K2	CO3
7.	How do Conditional Random Fields (CRFs) differ from Hidden Markov Models (HMMs) in their modeling approach?	2	K2	CO4
8.	For what kind of data is an HMM primarily designed to model?	2	K1	CO4
9.	List the concept of uniform random sampling.	2	K1	CO5
10.	Define sampling.	2	K1	CO6

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

			Marks	K – Level	CO
11.	a)	Elaborate on Baye's Decision Theory for classification. How does it leverage prior knowledge and observed data to make optimal classifications? Discuss its advantages and limitations.	13	K2	CO1
		OR			
	b)	Discuss the fundamental concepts of information theory, including entropy and information gain, and explain their significance in machine learning, particularly in decision tree algorithms.	13	K2	CO1
12.	a)	Describe the boosting technique. How does it create strong learner from weak learners?	13	K2	CO2
		OR			
	b)	Explain the back-propagation algorithm in detail. How does it enable neural networks to learn from data?	13	K2	CO2

13.	a)	Describe the mechanism of Principal Component Analysis (PCA) for dimensionality reduction. Explain the concept of variance maximization and orthogonality in the context of principal components. Discuss the benefits and a major limitation of using PCA.	*13*	*K2*	*CO3*
		OR			
	b)	Explain the K-means clustering algorithm in detail, outlining its objective function and how it attempts to minimize it. Discuss the sensitivity of K-means to initial centroid placement and methods to mitigate this issue.	*13*	*K2*	*CO3*
14.	a)	Discuss the importance of "generalization" in the context of graphical models. Explain how model complexity (e.g., number of parameters, graph density) can affect a graphical model's ability to generalize, and describe strategies to ensure good generalization performance.	*13*	*K2*	*CO4*
		OR			
	b)	Compare and contrast how conditional independence is represented and identified in Directed Graphical Models (Bayesian Networks) versus Undirected Graphical Models (Markov Random Fields). Use specific rules (e.g., d-separation for BNs, graph separation for MRFs) in your explanation.	*13*	*K2*	*CO4*
15.	a)	Describe the Policy Iteration algorithm for solving Reinforcement Learning problems. Compare and contrast its approach with Value Iteration, highlighting their respective strengths and the scenarios where one might be preferred over the other.	*13*	*K2*	*CO5*
		OR			
	b)	Differentiate between deterministic and non-deterministic (stochastic) rewards and actions in Reinforcement Learning. Explain how the presence of non-determinism impacts the design of RL algorithms and the challenges faced by the agent in learning an optimal policy.	*13*	*K2*	*CO5*

PART - C (1× 15 = 15 Marks)

16.	a)	Discuss about Ensemble Learning Algorithm Complexity and Occam’s Razor.	*15*	*K2*	*CO6*
		OR			
	b)	Elaborate the sampling methods used in machine learning algorithms with examples.	*15*	*K2*	*CO6*