

U.S.N.

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

February / March 2024 Semester End Main Examinations**Programme: B.E.****Branch: Artificial Intelligence and Machine Learning****Course Code: 23AM5PCIML****Course: Introduction to Machine Learning****Semester: V****Duration: 3 hrs.****Max Marks: 100**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.			UNIT - I	CO	PO	Marks
	1	a)	How can we develop and implement the Candidate Elimination algorithm to efficiently learn and update hypotheses while adapting to new examples and eliminating inconsistent hypotheses in a given dataset?	CO 1	PO 1	6
		b)	Consider the given data set having the data about which seeds are poisonous. Identify the specific hypothesis using Find-S algorithm.	CO 1	PO 2	10
		c)	List and explain any four main challenges in Machine Learning.	CO 1	PO 1	4

Example	Color	Toughness	Fungus	Appearance	Poisonous
1	GREEN	HARD	NO	WRINKLED	YES
2	GREEN	HARD	YES	SMOOTH	NO
3	BROWN	SOFT	NO	WRINKLED	NO
4	ORANGE	HARD	NO	WRINKLED	YES
5	GREEN	SOFT	YES	SMOOTH	YES
6	GREEN	HARD	YES	WRINKLED	YES
7	ORANGE	HARD	NO	WRINKLED	YES

UNIT - II

- 2 a) Construct a decision tree for the following training samples using ID3 algorithm, by considering "Sensitivity" as the target attribute.

Example	Height	Hair	Eyes	Sensitivity
1	Short	Blond	Blue	Yes
2	Tall	Blond	Brown	No
3	Tall	Red	Blue	Yes
4	Tall	Dark	Brown	No
5	Short	Dark	Blue	No
6	Tall	Dark	Blue	No
7	Tall	Blond	Blue	Yes
8	Short	Blond	Brown	No

CO 2

PO 2

10

- b) Consider the value of k as 5, apply KNN to dataset provided and classify the instance {170,57} accordingly.

Height (CM)	Weight (KG)	Class
167	51	Underweight
182	62	Normal
176	69	Normal
173	64	Normal
172	65	Normal
174	56	Underweight
169	58	Normal
173	57	Normal
170	55	Normal
170	57	?

CO 2

PO 2

10

OR

- 3 a) Elaborate inductive bias in decision tree learning.

CO 2

PO 1

5

- b) Given Positively labeled data points {(2,2), (2, -2), (-2,-2), (-2, 2)} and Negatively labelled data points {(1,1), (1,-1), (-1, -1), (-1,1)}, Classify the data points by applying Non Linear SVM classifier considering:

CO 2

PO 2

10

$$\Phi_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

- c) Derive the gradient descent rule for a locally weighted regression to the target function.

CO 2

PO 1

5

UNIT - III

4	a)	With an example explain EM algorithm in detail.	CO 2	PO 1	10
	b)	For the Bayesian network <div style="text-align: center;"> <pre> graph TD A((A)) --> C((C)) B((B)) --> C((C)) C((C)) --> D((D)) C((C)) --> E((E)) D((D)) --> F((F)) E((E)) --> G((G)) F((F)) --> H((H)) G((G)) --> H((H)) </pre> </div> <div style="margin-left: 40px;"> i. Are D and E necessarily independent given evidence about both A and B? ii. Are A and C necessarily independent given evidence about D? iii. Are A and H necessarily independent given evidence about C? </div>	CO 2	PO 2	10

OR

5	a)	Illustrate any two ensemble learning techniques commonly used in Machine Learning, highlighting the fundamental principles and advantages of each.	CO 2	PO 1	10																
	b)	<p>The frequency of the features {Yellow, Sweet} required to classify the fruits is summarized:</p> <table><tr><th>Fruit</th><th>Yellow</th><th>Sweet</th><th>Total</th></tr><tr><td>Mango</td><td>350</td><td>450</td><td>800</td></tr><tr><td>Banana</td><td>400</td><td>300</td><td>700</td></tr><tr><td>Total</td><td>750</td><td>750</td><td>1500</td></tr></table> <p>Using the concept of Naïve Bayes, predict the type of fruit possessing the properties {Yellow, Sweet}</p>	Fruit	Yellow	Sweet	Total	Mango	350	450	800	Banana	400	300	700	Total	750	750	1500	CO 2	PO 1	10
Fruit	Yellow	Sweet	Total																		
Mango	350	450	800																		
Banana	400	300	700																		
Total	750	750	1500																		

UNIT - IV

6	a)	By considering the support threshold as $S=33.34\%$ & confidence threshold as $C=60\%$, apply Apriori algorithm on the grocery store transactions and answer the following:	CO 3	PO 1	10														
		<table><tr><th>T. ID</th><th>Items</th></tr><tr><td>T1</td><td>HotDogs, Buns, Ketchup</td></tr><tr><td>T2</td><td>HotDogs, Buns</td></tr><tr><td>T3</td><td>HotDogs, Coke, Chips</td></tr><tr><td>T4</td><td>Chips, Coke</td></tr><tr><td>T5</td><td>Chips, Ketchup</td></tr><tr><td>T6</td><td>HotDogs, Coke, Chips</td></tr></table>	T. ID	Items	T1	HotDogs, Buns, Ketchup	T2	HotDogs, Buns	T3	HotDogs, Coke, Chips	T4	Chips, Coke	T5	Chips, Ketchup	T6	HotDogs, Coke, Chips			
T. ID	Items																		
T1	HotDogs, Buns, Ketchup																		
T2	HotDogs, Buns																		
T3	HotDogs, Coke, Chips																		
T4	Chips, Coke																		
T5	Chips, Ketchup																		
T6	HotDogs, Coke, Chips																		

		<ul style="list-style-type: none"> i. Find the candidate and frequent itemset of each database scan. Make sure to enumerate all of the final frequent itemsets. ii. Summarize the generated association rules and highlight the 2 strong rules iterated on highest confidence score. 			
	b)	Classify the nodes A=(2,10), B=(2,5), C=(8,4), D=(5,8), E=(7,5), F=(6,4), G=(1,2), H=(4,9) into 3 clusters using k-means clustering algorithm. Note: Node A, D, G can be considered as the initial seeds.	CO 3	PO 2	10
		UNIT - V			
7	a)	Illustrate the process of Linear Discriminant analysis with suitable equations.	CO 3	PO 1	10
	b)	<p>James has been tasked for analyzing a large dataset containing information about customer behavior in an e-commerce platform. The dataset includes various features such as browsing time, purchase history and demographic information. However, the dataset is high-dimensional and James suspects that there might be redundant or correlated features affecting the performance of the machine learning models.</p> <ul style="list-style-type: none"> i. How can Principal Component Analysis (PCA) be applied to improve the efficiency and interpretability of machine learning models in the context of analyzing customer behavior in our e-commerce platform? ii. Describe the step-by-step process of implementing PCA on the dataset, including the computation of principal components and the variance explained. Discuss the implications of dimensionality reduction in terms of model complexity, training time and the potential for overfitting. iii. Explore how PCA can help identify patterns and underlying structures in the customer behavior data, leading to more insightful and effective model outcomes. Highlight any considerations or challenges that may arise in applying PCA to this specific e-commerce scenario. 	CO 3	PO 3	10
