Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Examination, Summer 2026
Course Code: CSE431, Course Title: Machine Learning
Level: 4 Term: 2 Batch: 64

Time: 02:00 Hrs **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1. A machine learning model classifies ECG signals into three classes: N (Normal), A (Arrhythmia), and M (Myocardial Infarction). The test dataset is imbalanced, where most samples belong to the Normal class. The test results are:

$$Y=[N,N,N,N,N,N,A,A,M,M]$$
$$\hat{Y}=[N,N,N,A,N,N,N,A,N,M]$$

To improve the ECG classification model, three different models were trained. Three models produced the following performance:

Model	Train	Validation	AUC
A	99%	81%	0.74
B	91%	89%	0.92
C	76%	74%	0.69

Based on the above scenario answer the following questions:

a)	Construct the confusion matrix.	[2]	CO3
b)	Evaluate the performance of the Myocardial Infarction (M) class using the given test results by calculating the Accuracy, Precision, Recall, and F1-score, and identifying the evaluation metric(s) that should be prioritized for this ECG classification problem.	[5]	
c)	Evaluate the performance of the three models in terms of underfitting, overfitting, generalization, bias–variance trade-off, AUC, model performance, and an appropriate validation technique before deployment.	[3]	

2. A delivery robot is trained using Reinforcement Learning to navigate a warehouse. During training, the robot receives a reward of **+100** for reaching the destination, **−20** for hitting an obstacle, **−1** for each movement, and **0** for staying idle. After several training episodes, the robot frequently chooses to remain idle instead of moving toward the destination.

a)	Analyze the robot's behavior based on the given reward function.	[2]	
b)	Based on the concept of an optimal policy, explain what policy the robot should learn to maximize cumulative reward. Justify your answer.	[3]	

3. A smart healthcare company is developing two AI-based systems. The first system continuously monitors a patient's heart rate, body temperature, blood pressure, and oxygen saturation over several days to estimate the patient's hidden health condition, which cannot be observed directly. The system analyzes the sequence of observations to determine whether the patient is in a Healthy, Mild Illness, or Critical state. **CO1**

	The second system predicts whether a patient has heart disease using multiple medical attributes such as age, cholesterol level, blood pressure, and ECG measurements.		
	a) Identify the most appropriate technique for determining the hidden health condition.	[2]	
	b) Explain the role of the input layer, hidden layer, and output layer in developing the heart disease prediction system.	[3]	
4.	A city transportation authority wants to identify traffic accident hotspot regions from historical accident locations. The analysts prefer a clustering approach that groups densely populated regions, automatically identifies isolated locations as outliers, and does not require the number of clusters to be specified in advance. The accident locations are: $A(1,1), B(1,2), C(2,1), D(8,8), E(8,9)$ The clustering parameters are $\varepsilon = 1.5$ and MinPts = 3. A new accident location is recorded at $P(9,9)$.		CO2
	a) Classify each point as a core point, border point, or noise point by applying the most appropriate clustering algorithm for the above scenario.	[6]	
	b) Using the identified clustering technique, determine the clusters formed and whether the new point $P(9,9)$ belongs to an existing cluster or remains as noise.	[4]	

5. An e-commerce company wants to classify new customers into two loyalty groups: Premium (P) and Regular (R), based on purchasing behavior. Consider the customer data have already been standardized. Customer Data:

Customer	Purchase Frequency (X_1)	Average Spending (X_2)	Website Visit Time (X_3)	Product Reviews (X_4)	Class
C_1	3	2	1	2	P
C_2	2	3	2	1	P
C_3	1	2	3	2	P
C_4	-2	-1	-2	-3	R
C_5	-3	-2	-1	-2	R

A new customer has the feature vector:
X_new = [2, 1, 2, 1]
After applying PCA, the following eigenvalues are obtained:
$\lambda_1 = 2.6, \lambda_2 = 1.1, \lambda_3 = 0.2, \lambda_4 = 0.1$
The feature vector matrix for the first two principal components is:

$$W = \begin{bmatrix} 0.65 & 0.43 \\ 0.43 & -0.65 \\ 0.54 & 0.32 \\ 0.32 & -0.54 \end{bmatrix}$$

Answer the following questions based on the above scenario:

a)	Determine the minimum number of principal components required to preserve at least 90% of the total variance using the necessary calculations.	[2]	CO2
b)	Calculate the PCA-transformed coordinates (PC_1, PC_2) for each training customer and the new customer using the given feature vector matrix.	[3]	
c)	Predict the class of the new customer using k-Nearest Neighbors ($k = 3$)based on the Euclidean distances between its transformed coordinates and those of each training customer.	[5]	