[This question paper contains 12 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 5547 **J**

Unique Paper Code : 2342013602

Name of the Paper : Machine Learning

Name of the Course : **B.Sc. (H) Computer Science**

Semester : VI

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt all questions from **Section-A.**
3. Attempt any **four** questions from **Section-B.**
4. Attempt **all** parts of a question together.
5. Use of **Scientific Calculator** is allowed.

SECTION A

1. (a) How can logistic regression, originally for binary classification, be extended to handle multiclass problems? Give one example. (3)

   (b) Consider a binary classification model that achieves an accuracy of over 99% on the training data. However, it fails to perform well on the test dataset. Identify whether it is a case of overfitting or underfitting. Justify your answer. Suggest one practical strategy to address this issue. (3)

(c) Consider two clustering algorithms which when applied to the same dataset of N data points.

- Algorithm 1 produces two equally sized clusters (each of size N/2),
- Algorithm 2 produces two clusters with N/4 and 3N/4 data points respectively.

Evaluate which clustering solution is better. List one clustering evaluation metric you would use? Justify your answer. (3)

(d) Why is non-linearity through activation functions introduced in hidden neurons of Neural Network? Give an example of any one non-linear activation function suitable for the hidden layer for the classification problem. (4)

(e) What is the curse of dimensionality problem? Explain any one feature subset selection technique to deal with this problem. (4)

(f) Consider the following dataset containing two numerical attributes, Age and Salary: (4)

ID	Age (in years)	Salary (in rupees)
1	44	72000
2	27	48000
3	30	54000
4	38	61000
5	50	83000
6	37	67000

Explain the potential issues if the above dataset is directly used for k-means clustering without preprocessing. Suggest and apply an appropriate preprocessing step to handle this issue. (4)

(g) For the following test dataset, consider the actual class labels and the class labels predicted by the machine learning model Z provided below: (4)

Instance	A	B	C	Target Class	Predicted Class
1	T	T	1	+	+
2	T	T	6	+	+
3	T	F	5	--	+
4	F	F	4	+	--
5	F	T	7	--	+
6	F	T	3	--	+
7	F	F	8	--	--
8	T	F	7	+	--
9	F	T	5	--	--
10	F	F	10	--	--
11	T	F	1	+	+
12	F	T	9	+	+

Draw the confusion matrix depicting the number of records correctly/ incorrectly classified. Evaluate the performance of the machine learning model Z in terms of accuracy, and F1-Score.

(h) Given the following scenarios, identify whether each case falls under Supervised Learning, Unsupervised Learning, or Reinforcement Learning.

Justify your answer in one line for each. (5)

(i) A retail company uses customer purchase history to segment customers into groups.

(ii) A mobile app learns to identify spam messages.

(iii) A robot learns to walk by trial and error and gets a reward when it moves forward.

(iv) A weather prediction model trained on past temperature and humidity data.

(v) An AI system groups articles from different news sources based on topic similarity.

SECTION B

2. (a) Using the data given below, build a logistic regression model to predict whether a customer will purchase (1) or not purchase (0) a product based on their Age and Income, using the gradient descent algorithm. Assume the initial values of the model parameters as $\theta_0 = 0$, $\theta_1 = 0$, $\theta_2 = 0$, and the learning rate as 0.01.Perform one iteration of the gradient descent algorithm to update the model parameters. (5)

Age	Income (in 1000s)	Purchased (Y)
25	40	0
30	60	0
45	85	1
35	75	1
40	65	1

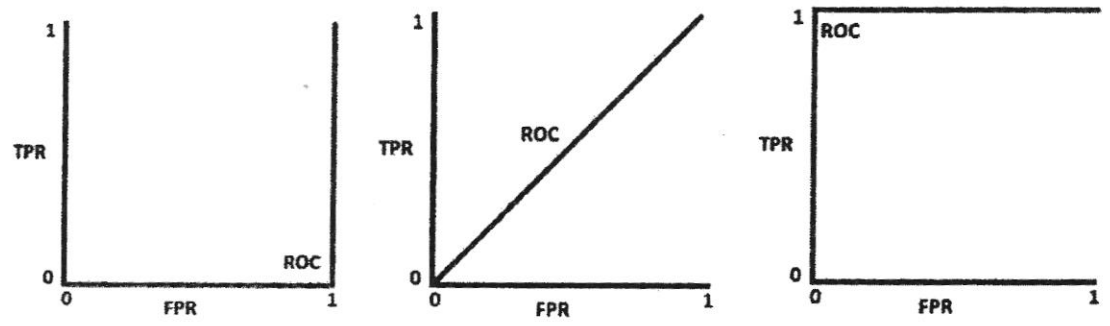
(b) Consider the following dataset that records student behavior and whether they (5)

Study Hours	Attendance	Sleep Quality	Part-time Job	Pass/Fail
Low	Poor	Bad	Yes	Fail
High	Good	Good	No	Pass
Medium	Good	Good	Yes	Pass
Low	Poor	Bad	No	Fail
High	Good	Bad	No	Pass
Medium	Poor	Good	Yes	Fail
High	Good	Good	Yes	Pass
Low	Good	Bad	Yes	Fail
Medium	Good	Good	No	Pass
High	Poor	Good	No	Pass

Using the Naive Bayes classifier, predict whether the following student will pass or fail:

- Study Hours = Medium, Attendance = Good, Sleep Quality = Good, Part-time Job = No

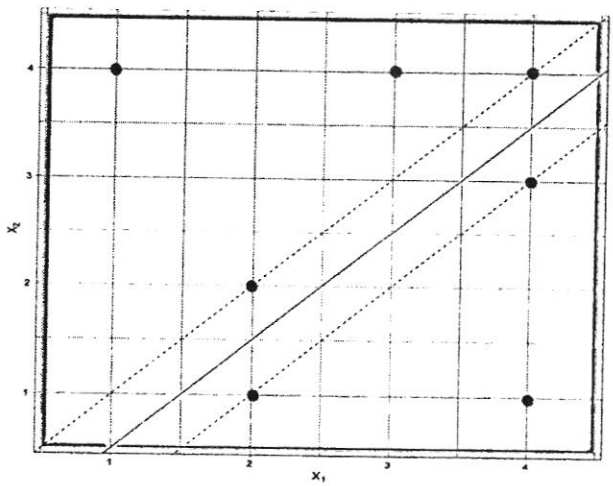
(c) What is an ROC curve? Why is it used in classification problems? Given the ROC curves below for three different classifiers, determine the AUC (Area Under the Curve) for each and identify the best-performing model: (5)



3. (a) Consider the following two-dimensional dataset: (5)

Obs.	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

The dataset is graphically represented below, with a decision boundary separating instances of the Red and Blue classes:



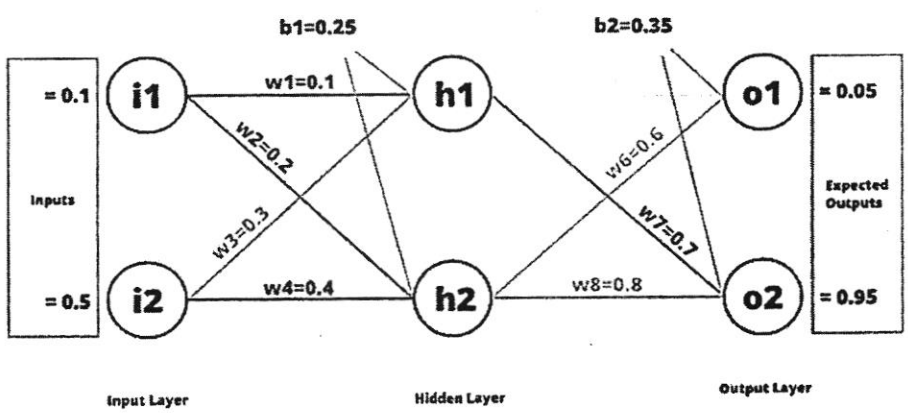
(i) Based on the figure above, provide the equation of the optimal separating hyperplane for the Support Vector Machine (SVM).

(ii) Identify the support vectors for the SVM classifier applied to this dataset.

(iii) Define the term margin in the context of SVM.

(b) Use forward propagation to determine the output and prediction error for the following neural network (use ReLU for hidden and output layers):

(5)



(c) (i) Write the mathematical expression of the sigmoid function and state its output range. (5)

(ii) Why is Mean Squared Error (MSE) not commonly used as the cost function in logistic regression? Which cost function is used for logistic regression?

(iii) What does the output value of logistic regression represent?

4. (a) Consider the following dataset for binary classification problem: (6)

Instance	A	B	C	Target Class
1	T	T	1	+
2	T	T	6	+
3	T	F	5	-
4	F	F	4	+
5	F	T	7	-
6	F	T	3	-
7	F	F	8	-
8	T	F	7	+
9	F	T	5	-

Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose from A and B?

(b) Explain Lasso Regression in the context of Logistic Regression. Explain how the regularization parameter λ controls bias-variance tradeoff. (5)

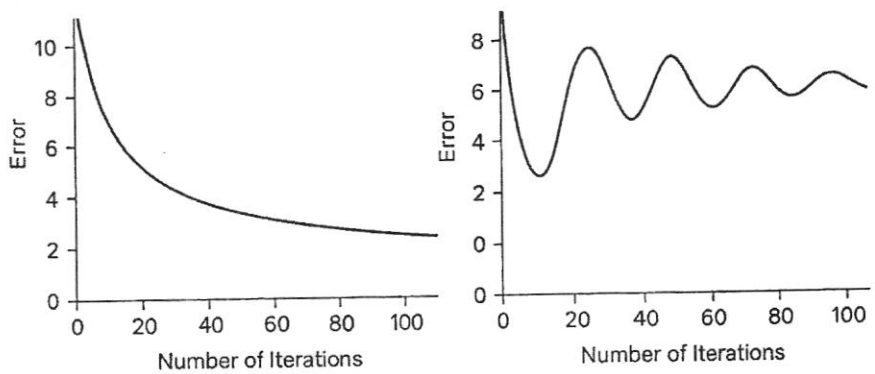
(c) For evaluating the performance of a classifier, how does the holdout method differ from k-fold cross-validation? For k-5 and datapoints- D1, D2, D3, D4, D5, D6, D7, D8, D9, and D10 in the dataset, mention one possible dataset distribution between training and test partition for k-fold cross-validation. (4)

5. (a) Given below is the data of five students who took a proficiency test as well as language course. Proficiency test score Language Course Performance (6)

Proficiency test score	Language Course Performance
95	85
85	95
80	70
70	65
60	70

(i) Use the least square approximation to estimate the linear equation that best predicts language course performance, based on proficiency test scores?

(ii) Considering the above scenario in which these linear regression model parameters are learned using gradient descent with two different learning rates (one high, one low), identify which of the following curves represents the higher and lower learning rates. Justify your answer.



(b) How does the PCA (Principal Component Analysis) algorithm help reduce dimension in machine learning? Write the steps of the PCA algorithm.

(5)

(c) A data scientist builds a linear regression model to predict a student's final exam score based on the number of study hours and attendance rate. After training the model, the following metrics are reported: (4)

- Mean Squared Error (MSE): 36.0
- Coefficient of Determination (R^2): 0.82

(i) Define both Mean Squared Error (MSE) and Coefficient of Determination (R^2). Write the mathematical formulas for each.

(ii) In the context of model evaluation, explain when R^2 is preferred over MSE.

6. (a) Consider the following data set: {4, 8, 12, 20, 32, 36, 48}. Assume k =2, and initial cluster centers for k-means clustering as 32 and 48. Perform the k-means clustering to arrive at the final set of clusters. Also, at the end of every iteration, compute the SSE. (7)

(b) Consider the following optimization problem for Support Vector Classifiers: (5)

$$\underset{\beta_0,\beta_{11},\beta_{12},\ldots,\beta_{p1},\beta_{p2},\epsilon_1,\ldots,\epsilon_n,M}{\text{maximize}} \quad M$$

$$\text{subject to } y_i\left(\beta_0+\sum_{j=1}^{p}\beta_{j1}x_{ij}+\sum_{j=1}^{p}\beta_{j2}x_{ij}^2\right)\geq M(1-\epsilon_i)$$

$$\sum_{i=1}^{n}\epsilon_i\leq C,\ \ \epsilon_i\geq 0,\ \ \sum_{j=1}^{p}\sum_{k=1}^{2}\beta_{jk}^2=1.$$

In Support Vector Classifiers (SVC), the optimization problem includes a regularization parameter C that controls the tradeoff between maximizing the margin and minimizing classification errors. Explain how different values of C affect the bias-variance tradeoff in SVC:

(i) When C is large

(i) When C is small

(c) List two scenarios where K-means clustering technique may not perform well. (3)

7. (a) The DM Pizza Parlour sells pizzas with optional toppings: pepperoni, pineapple, and pickled-onion. Suppose, you have tried five pizzas (P1 to P5) and kept a record of which you liked: (5)

	Pepperoni	pineapple	pickled-Onion	liked
P1	True	True	True	False
P2	True	False	False	True
P3	False	True	True	False
P4	False	True	False	True
P5	True	False	False	True

Show binarization of the above data and use it to calculate Euclidean distances, to demonstrate how the k-Nearest-Neighbor (k-NN) classifier with majority voting would classify a tuple $< False, True, True >$, *for* $k = 3$.

(b) Use the distance matrix given below to perform agglomerative hierarchical clustering using single link and show the dendrogram. (5)

	P1	P2	P3	P4	P5
P1	0				
P2	0.24	0			
P3	0.22	0.15	0		
P4	0.37	0.20	0.05	0	
P5	0.34	0.14	0.28	0.29	0

(c) Explain why a single layer perceptron cannot be used to solve XOR problem. (5)

(1500)