This question paper contains 8 printed pages]

Roll No. | | | | | | | | | | |

S. No. of Question Paper : **1342**

Unique Paper Code : **32347607**

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

Name of the Paper : **Machine Learning**

Semester : **VI**

Duration : **3 Hours** Maximum Marks : **75**

(Write your Roll No. on the top immediately on receipt of this question paper.)

Attempt *all* questions from Section A.

Attempt any *four* questions from Section B.

Attempt *all* parts of a question together.

Use of Scientific Calculator is allowed.

Section–A

1. (*a*) A company applies K-means clustering to segment customers. Explain how the initial choice of centroids can affect the final clusters. 2

(*b*) Consider the following data that a healthcare organization is using to predict patient risk levels based on blood pressure and cholesterol data for five individuals. Normalize the feature–Blood Pressure using the Min-Max normalization technique : 3

P.T.O.

Person	Blood Pressure (mmHg)	Cholesterol (mg/dL)	Risk Level
A	120	180	Low
B	145	220	High
C	130	200	Low
D	160	250	High
E	135	210	Low

(*c*) List and briefly explain any *two* evaluation metrics used to compare the performance of regression models. 3

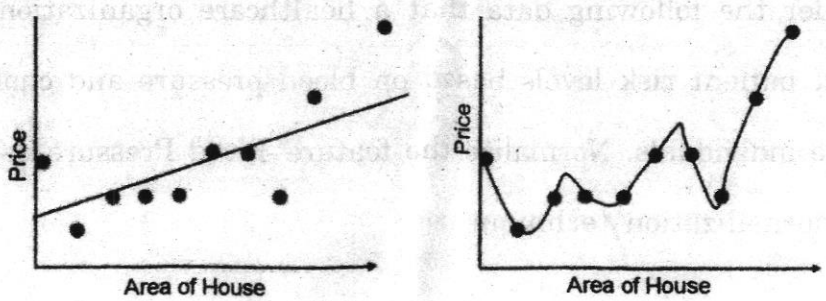
(*d*) Consider the following dataset and use k-Nearest Neighbor (k-NN) classifier to predict the success (Class = Yes/No) of a new product :

A = 4 (advertisement spending in thousands) and B = 3 (positive reviews)

Assume the value of k as 3 and Euclidean distance as the proximity metric : 4

A	B	Class
2	3	Yes
5	5	No
1	1	Yes
6	2	No
8	5	Yes

(*e*) Consider the following two models built for a linear regression task. Compare the performance of the model in terms of bias and variance. 4



(*f*) Given the following scenarios, identify whether each case falls under supervised or unsupervised learning. Justify your answer in one line for each task : 4

(*i*) Predicting if a loan will be approved.

(*ii*) Organizing news articles based on content similarities without using labelled categories.

(*iii*) Predicting the number of copies a music album will sell.

(*iv*) Grouping customers based on purchasing behaviour.

(*g*) A company is building a machine learning model to predict whether a customer will purchase a product based on browsing history and demographic features : 5

(*i*) Evaluate the suitability of using linear regression *versus* logistic regression for this problem.

(*ii*) Based on the appropriate model, clearly define the hypothesis and cost functions.

(*iii*) Based on the appropriate model, write the mathematical expression for Ridge (L2) regularization.

(*h*) A hospital has developed a Machine Learning model, MedPredict, to classify whether patients are at high risk (+) or low risk (–) for a particular heart condition, based on attributes : cholesterol level (High/Low), blood pressure (High/Normal), and age (numerical). A test dataset of 12 patients with their actual and predicted risk classifications is shown below : 5

Patient ID	Cholesterol	BP	Age	Actual Risk	Predicted Risk
1	High	High	65	+	+
2	High	High	70	+	+
3	High	Normal	45	–	+
4	Low	Normal	60	+	–
5	Low	High	80	–	+
6	Low	High	50	–	+
7	Low	Normal	75	–	–
8	High	Normal	77	+	–
9	Low	High	66	–	–
10	Low	Normal	90	–	–
11	High	Normal	55	+	+
12	Low	High	85	+	+

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

(*i*) Consider a classification model trained on a dataset of 100 instances using 5-Fold Cross Validation. How many instances are used for training and test partition in each fold ? The model achieves the following accuracies in each fold : 5

- Fold 1 : 80%
- Fold 2 : 82%
- Fold 3 : 78%
- Fold 4 : 81%
- Fold 5 : 79%

Compute the mean accuracy of the model. Briefly mention the difference between 5-Fold Cross Validation and Leave-One-Out Cross Validation ((LOOCV).

Section–B

2. (*a*) Using the data comprising one training example below, build a logistic regression model to predict the output y 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

$$x_1 = 1,\ x_2 = 2 \text{ and label } y = 1.$$

Also, answer the following :

(*i*) What could happen if the learning rate is too small ?

(*ii*) What could happen if the learning rate is too large ?

(*b*) Consider the following dataset that records whether the given transaction is fraudulent or not : 5

Device	Location	Amount	Fraudulent
Mobile	Local	High	Yes
Mobile	Abroad	Low	No
Laptop	Local	Low	No
Laptop	Abroad	High	Yes
Mobile	Abroad	High	Yes
Mobile	Abroad	High	No
Laptop	Local	High	No

Using the Naïve Bayes classification technique, predict whether the following transaction is fraudulent or not, based on the given feature values : {*Device : Mobile, Location : Abroad Amount : High*}.

3. (*a*) Consider the following dataset used for binary classification using a Support Vector Machine (SVM) : 6

- Positively labelled data points :

  {(4, 1), (4, –1), (8, 1), (8, –1)}

- Negatively labelled data points :

  {(1, 0), (0, 1), (0, –1), (–1, 0)}

Answer the following questions :

(*i*) Identify the support vectors with respect to the SVM classifier applied on the above data.

(*ii*) Sketch the optimal and marginal hyperplanes formed by the SVM for this classification problem.

(*iii*) Write the equation of Gaussian kernel function of SVM.

(*b*) Compute the derivative of the following function : 4

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

4. (*a*) A company wants to build a decision tree model to predit whether a person will buy a product using the features : Age and Student using the following training dataset : 5

Person	Age	Student	Buys
1	Young	No	No
2	Young	Yes	Yes
3	Old	Yes	Yes
4	Old	No	No
5	Middle	Yes	Yes

Calculate the Information Gain when splitting on Age and Student. Based on the calculations, which attribute would the decision tree induction algorithm choose for the first split ?

(*b*) Construct the truth table for a 2-input NAND logic gate. Demonstrate how a single-layer perceptron can be used to model this logic gate. 5

5. (*a*) A two-layered neural network is trained to classify handwritten digits. Explain how weights are updated during training using backpropagation by listing the steps of gradient descent backpropagation algorithm. 5

(*b*) Consider the following dataset providing car's engine size (in litres) and its fuel efficiency (miles per gallon) : 5

Engine Size (Litres)	Mileage (mpg)
1.2	35
1.6	32
2	28
2.4	24

Use the Least Squares Method to fit a linear regression line to learn the relationship between these two attributes.

6. (*a*) Cluster the following data using agglomerative hierarchical clustering with complete linkage and show the dendrogram : 6

Points	x	y
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4

(*b*) Derive the gradient descent update rule to find the value of optimal parameters θ_0 and θ_1 for the simple linear regression model using Mean Squared Error (MSE) as the cost function. 4

7. (*a*) Cluster the following data into two clusters (k = 2) using the k-means algorithm, with objects O2 and O5 as the initial cluster centers. Show the clusters at the end of two iterations. Also, compute the SSE at the end of 2nd iteration : 5

Objects	X-coordinate	Y-coordinate
O1	2	4
O2	4	6
O3	6	8
O4	10	4
O5	12	4

(*b*) How does the PCA (Principal Component Analysis) algorithm help reduce the number of dimensions in machine learning ? Write the steps of the PCA algorithm. 5