[This question paper contains 8 printed pages.]

Your Roll No.....

E

Sr. No. of Question Paper : 4835

Unique Paper Code : 32347607

Name of the Paper : Machine Learning

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

Year of Admission : (Admission of 2019)

Semester : VI

Duration : 3 Hours Maximum Marks : 75

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. **Section A** is compulsory.
3. Attempt any **four** questions from **Section B**.
4. Use of scientific calculator is allowed.

Section A (Compulsory)

1. (a) Explain the cost function of logistic regression. (3)

P.T.O.

(b) Give a formula for the binary sigmoid activation function. Also obtain the first derivative of this function. (3)

(c) What is K-means clustering algorithm used for? (3)

(d) Explain two applications of machine learning. (3)

(e) Consider the following data. Find the information gain if A2 is selected as the root of the decision tree. (6)

A1	A2	A3	Class
True	Hot	High	No
True	Hot	High	No
False	Hot	High	Yes
False	Cool	Normal	Yes
False	Cool	Normal	Yes
True	Cool	High	No
True	Hot	High	No
True	Hot	Normal	Yes
False	Cool	Normal	Yes
False	Cool	High	Yes

(f) Can a single layer perceptron solve the XOR problems? Justify your answer. (3)

(g) Find the line of best fit using least square regression method, given the following data :

x	y
2	6
6	13
10	28
12	30

Predict the value of y for the test data x = 11. (4)

(h) Given two situations : (4)

(i) A publishing house wants to predict the number of copies of a book would be sold. They want to use the popularity rating of the author, success index of her/his previous books and one other variable.

(ii) Based on the blood test reports, a researcher builds a model to identify the type of diabetes (type 1, type 2, gestational diabetes) that a patient has.

Identify each of these situations as a classification problem or a regression problem. Justify your answer.

P.T.O.

(i) Consider the Confusion matrix given below:

		Actual Values	
		True	False
Predicted Values	True	10	11
	False	4	20

Compute :

(i) Accuracy

(ii) Precision

(iii) Recall (6)

Section B

2. (a) Differentiate between supervised, unsupervised and reinforcement learning. (6)

(b) Using a diagram, depict the following scenarios for a machine learning model :

(i) High variance, low bias

(ii) Low variance, high bias (4)

3. (a) It is found that a classification model performs with a high accuracy on training data, but with new instances, it generalizes poorly. What could be the problem in this model? Explain using a suitable diagram. Give two possible solutions to this problem. (6)

(b) Define and elaborate the terms : (4)

(i) Hypothesis space

(ii) Inductive bias

4. (a) Apply the K-nearest neighbor algorithm on the following dataset : (4)

a	b	Class
7	7	False
7	4	False
3	4	True
1	4	True

Predict the class for X (a = 3, b = 7). Assume K = 3 (Use Euclidean distance to compute the distance)

P.T.O.

(b) What is the disadvantage of the filter approach for feature selection?

List the steps for feature selection using Principal Component Analysis (PCA). (6)

5. (a) Design a neural network for the Boolean function AND (for two variables) with the help of a neat diagram. (4)

(b) Explain the back-propagation algorithm for a multilayer perceptron. (6)

6. (a) What do you understand by the following terms w.r.t. Support Vector Machine (SVM) learning algorithm?

(i) Support vectors

(ii) Marginal distance (4)

(b) For the following data, apply one iteration of K-means clustering to partition the data in 2 clusters. Assume points C1 = (1,2,3) and C2 = (3,4,5) as the initial cluster centers. Show the generated clusters and cluster centers after the first iteration. (Use Manhattan distance to compute the distance). (6)

A	B	C
1	2	2
2	4	1
2	3	1
6	3	8
5	2	9
1	4	3

7. (a) What is the basic assumption of a Naive Bayes classifier? (2)

(b) Consider the following training data set for car theft :

Sample No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Use the Naive Bayes classification rule to classify a Red Domestic SUV. (8)

P.T.O.

8. (a) Consider the following linear regression problem :

x	1	2	4	3	5
y	1	3	3	2	5

Given the initial values of θ_1, θ_2 as 0.5, 0.5 and learning rate as 0.1. Compute the next set of values for regression coefficients, using gradient descent method. (6)

(b) Answer the following for 5-fold cross validation on a training data set of 45 tuples :

(i) How many rounds of learning will be performed?

(ii) Statethe size of the training and testing set for each round of learning.

(iii) How would the accuracy of the classifier be calculated in 5-fold cross validation? (4)