AUTUMN END SEMESTER EXAMINATION-2022
5th Semester B.Tech (Deptt. Elective-II)
MACHINE LEARNING
CS3035

(For 2021 (L.E), 2020 & Previous Admitted Batches)

Time: 3 Hours   Full Marks: 50

Answer any SIX questions.
Question paper consists of four SECTIONS i.e. A, B, C and D.
Section A is compulsory.
Attempt minimum one question each from Sections B, C, D.
The figures in the margin indicate full marks.
*Candidates are required to give their answers in their own words as far as practicable
and <u>all parts of a question should be answered at one place only</u>.*

SECTION-A

1.   Answer the following questions. $[1 \times 10]$

   (a) Sketch the basic block diagram of a learning machine.

   (b) The Manhattan distance between two points (10, 10)
       and (30,30) is:
       A. 20
       B. 30
       C. 40
       D. 50

   (c) K-Nearest Neighbor is a _____ , _____ algorithm.
       A. Non-parametric, eager
       B. Parametric, eager
       C. Non-parametric, lazy
       D. Parametric, lazy

   (d) Give an example scenario where a multiple linear
       regression model is necessary.

   (e) State the difference between Generalized Linear Model
       and Linear model?

(f) What should you do when your model is suffering from low bias and high variance?

(g) How do you initialize weights in a neural network?

(h) In a binary classification problem, out of 30 data points 12 belong to class I and 18 belong to class II. What is the entropy of the data set?
A. 0.97   B. 0   C. 1   D. 0.67

(i) Which of the following algorithms is not an ensemble learning algorithm?
A. Random Forest   B. Adaboost
C. Gradient Boosting   D. Decision Trees

(j) Explain the phrase "Curse of Dimensionality".

SECTION-B

2. (a) Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. Derive the gradient descent training rule for weight update equation. [4]

(b) Using KNN algorithm and the given data set, predict the label of the test data point (3,7), where K=3 and Euclidean distance. [4]

X	Y	Label
7	7	1
7	4	1
3	4	2
1	4	2

3. (a) Fit a straight line trend by the method of least square from the following data and find the trend values. [4]

Year	1950	1951	1952	1953	1954	1955	1956	1957
Value	346	411	392	512	626	640	611	796

(b) Explain L1 and L2 regularization with detailed mathematical equations and explain its importance. [4]

SECTION-C

4. (a) Discuss the following terms with examples: Confusion Matrix, Accuracy, Precision, Recall. [4]

   Suppose I have 10,000 emails in my mailbox out of which 200 are spams. The spam detection system detects 150 mails as spams, out of which 50 are actually spams. What is the precision and recall of my spam detection system?

   (b) Explain the below given terms with suitable examples [4]
   (i) linearly and non-linearly separable datasets
   (ii) over-fitting and under-fitting

5. (a) Why do we use a logistic function in Logistic Regression? How does the Maximum Likelihood estimation help in computing the parameter of logit function? [4]

   (b) Divide the given sample data set into two(2) clusters using K means Algorithm by using Euclidean distance. Use Cluster 1 (185, 72) and Cluster 2 (170, 56) as initial cluster centers. [4]

Height	Weight
185	72
170	56
168	60
179	68
182	72
188	77

6. (a) Write Short notes on [4]
   I. Bagging
   II. Boosting
   III. Random Forests
   IV. Adaboost

   (b) Explain the Principal Component Analysis (PCA) and reduce the following dataset step-by-step from 2 dimensions to 1. [4]

Feature	Example 1	Example 2	Example 3	Example 4
x	4	8	13	7
y	11	4	5	14

SECTION-D

7. (a) Why is SVM more accurate than Logistic regression? [4]

Plot a hyper plane for the given data set (1,1) (2,1) (1,-1) (2,-1) (4,0) (5,1) (5,-1) (6,0) by using SVM.

(b) State the characteristics of an artificial neural network. [4] Construct a feed-forward network with five input nodes, three hidden nodes and four output nodes with necessary mathematical expressions and explanations.

8. (a) Consider the following set of training examples: [4]

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	_	T	F
4	+	F	F
5	_	F	T
6	_	F	T

I. What is the entropy of this collection of training examples with respect to the target function classification?

II. What is the information gain of a2 relative to these training examples?

III. Draw decision tree for the given data set.

(b) Implement AND function using McCulloch–Pitts [4] neuron (take binary data). Consider the below given truth table for AND function.

X1	X2	Y
1	1	1
1	0	0
0	1	0
0	0	0

*****