OR

(b) Explain the following : (CO1/CO2)

(i) Information gain and Gini Index with help of mathematical equations and examples.

(ii) Concept of K-fold cross validation and LOOCV with the figure.

H Roll No. .....

TCS–509

B. TECH. (CSE)

(FIFTH SEMESTER)

MID SEMESTER

EXAMINATION, Oct., 2023

MACHINE LEARNING

Time : 1:30 Hours

Maximum Marks : 50

Note : (i) Answer all the questions by choosing any ***one*** of the sub-questions.

(ii) Each question carries 10 marks.

1. (a) What do you understand by Supervised, Unsupervised and Reinforcement Machine Learning ? Explain the real time scenarios with examples where all these types of Machine Learning can be implemented. (CO1)

P. T. O.

OR

(b) Compute Mean, Median, Mode, Range, Average Deviation, Absolute Deviation, Squared Deviation, Standard Deviation, Total Sum of Squares for the following dataset : (CO1)

{18, 22, 33, 11, 9, 4}.

2. (a) Construct the regression tree using the following dataset with 10 instances and 3 attributes : (CO1/CO2)

Sl. No.	Assessment	Assignment	Project	Result (%)
1	Good	Yes	Yes	95
2	Average	Yes	No	70
3	Good	No	Yes	75
4	Poor	No	No	45
5.	Good	Yes	Yes	98
6	Average	No	Yes	80
7	Good	No	No	75
8	Poor	Yes	Yes	65
9	Average	No	No	58
10	Good	Yes	Yes	89

OR

(b) Consider the following data set of experience and salary of five employees. Compute the y-intersect and slope of the best-fitting line for Linea Regression : (CO1/CO2)

Experience	Salary
10	21
14	33
12	27
10	22
8	23

3. (a) How can we identify outliers in a dataset using SVM ? Explain the following figures : (CO1/CO2)

Figure A : Define the best fit in line Figure A, and justify the reason.

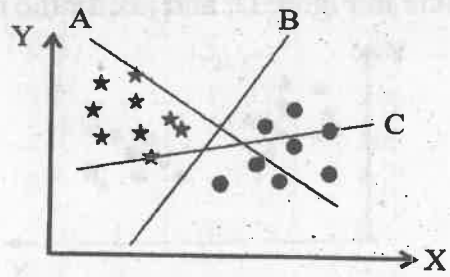

Figure A

Figure B : Define, why we cannot choose Line A and Line B as the best fit line in figure B.

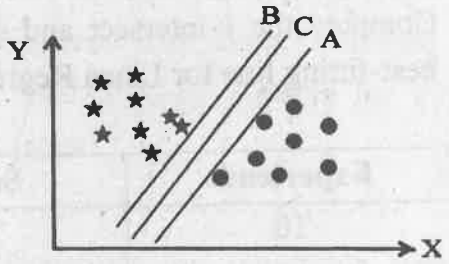

Figure B

Figure C : Define the best fit line in Figure C, and justify the reason.

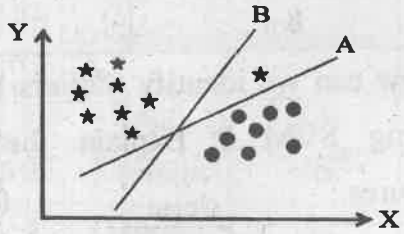

Figure C

Figure D : Draw and define the best fit line in Figure D, and justify the reason.

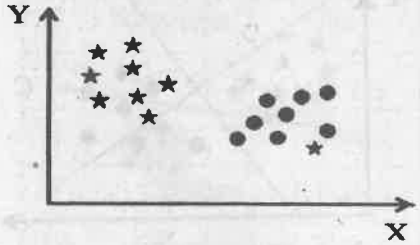

Figure D

Figure E : Draw and define the best fit line in Figure E, and justify the reason.

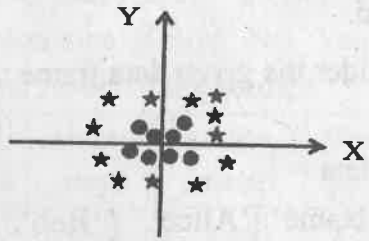

Figure E

OR

(b) Consider the following dataset with 10 data instances and having 2 attributes : (CO2)

Sl. No.	GPA	No. of projects done	Award
1	9.5	5	Yes
2	8	4	Yes
3	7.2	1	No
4	6.5	5	Yes
5	9.5	4	Yes
6	3.2	1	No
7	6.6	1	No
8	5.4	1	No
9	8.9	3	Yes
10	7.2	4	Yes

Given the test instance (GPA = 7.8, no. of projects done = 4) and K = 3. Apply KNN to model the problem for predicting the award.

4. (a) Consider the given data frame : (CO1/CO2)

(i) data = {

(ii) 'Name':['Alice', 'Bob', 'Charlie', 'David', 'Emily'],

(iii) 'Age' : [25, 30, None, 28, 24],

(iv) 'Gender' : ['Female', 'Male', 'Male', 'Male', 'Female'],

(v) 'Math_Score' : [85, 92, 78, 88, 760],

(vi) 'Science_Score' : [90, None, 85, 92, 88],

(vii) 'Passed_Exam' : ['Yes', 'Yes', 'No', 'Yes', 'No']

(viii) }

Using pandas perform the following :

(1) Create a DataFrame

(2) Handle missing values

(3) Drop rows with missing values in other columns

(4) Encoding categorical variables

(5) Display the preprocessed data

OR

(b) Consider the dataset. Train it using Random forest model. Given the test sequence (Good, No, Yes, Poor), predict the result of the student : (CO1/CO2)

Sl. No.	Assess-ment	Assign-ment	Pro-ject	Semi-nar	Res-ult
1	Good	Yes	Yes	Good	Pass
2	Average	Yes	No	Poor	Fail
3	Good	No	Yes	Good	Pass
4	Average	No	No	Poor	Fail
5	Average	No	Yes	Good	Pass
6	Good	No	No	Poor	Pass
7	Average	Yes	Yes	Good	Fail
8	Good	Yes	Yes	Poor	Pass

5. (a) Consider XNOR Boolean function that has 4 patterns (00, 01, 10, 11) in 2-dimensional space. Construct RBFNN that classifies the input pattern : (CO1/CO2)

$00 \rightarrow 1$

$01 \rightarrow 0$

$10 \rightarrow 0$

$11 \rightarrow 1$