Get the R2-value for the test data.

Test data	
X	Y
8	11
9	8.5
11	15
16	18
12	11

(OR)

b.i. Write general procedure followed for training and test ML model. 4 3 2 2

ii. Write note on K-fold cross validation. Represent pictorially for data set size N = 250 and K = 4. 8 3 2 2

30. a. What is principle component? How PCA method result in reduced dimensionality. 12 3 3 1

(OR)

b. Explain the need for kernel in SVM. List the types of kernel function in SVM. What is the difference between hand and soft margin classifier? 12 3 3 1

31. a. What is a dentogram? What are hierarchical clustering methods? List the types of linkages used in groups with their formula. 12 4 2 2

(OR)

b. Apply K-means algorithm in separating into 2 groups. 6 4 2 2

Medicine	Feature 1	Feature 2	Group
A	1	1	1
B	2	1	1
C	4	3	2
D	5	4	2

Take initial centroids $C_1 = (1, 1)$, $C_2 = (2, 1)$.

ii. Write K-medoid algorithm. What sample size of N = 10 and K = 2 apply the algorithm. 6 4 2 2

32. a. Represent neural network with a diagram. Represent the basic unity of neural network. 12 4 2 2

(OR)

b. Write about inductive bias used in decision tree learning. Issues in decision tree learning. 12 4 2 2

* * * * *

Reg. No.															

B.Tech. DEGREE EXAMINATION, JUNE 2023

Fifth & Sixth Semester

18CSE392T – MACHINE LEARNING – I

(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

Note:

(i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

(ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours Max. Marks: 100

	Marks	BL	CO	PO

PART – A (20 × 1 = 20 Marks)

Answer **ALL** Questions

1. Identify the type of ML whose goal is to learn a mapping from input X to output Y. 1 2 1 1

(A) Supervised (B) Unsupervised

(C) Semi-supervised (D) Reinforcement

2. Find the trace of the matrix 1 2 1 1

$$\begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

(A) 3 (B) 6

(C) 9 (D) 1

3. If a dataset is having 'K' features, for a requirement of 10 data points for any combination of features. How many data points are needed 1 2 1 1

(A) $2^K \times 10$ (B) $2^K + 10$

(C) $K2^{10}$ (D) K

4. Learning curve is a 1 2 1 1

(i) Diagnostic tool for understanding a model

(ii) Performance over varied data set size

(A) Only (i) (B) Both (i) and (ii)

(C) Only (ii) (D) Neither (i) nor (ii)

5. Pick the odd with response to ML platform 1 2 1 1

(A) Microsoft azure (B) IBM Watson

(C) H_2S (D) One-AI

6. Regressor models predicted output $\hat{y} = a + bx_1 + cx_2$. It has _____ parameters. 1 2 1 1

(A) 3 (B) 2

(C) 1 (D) No parameter

7. Logistic regression uses — 1 2 1 1
   (A) Sigmoid function (B) Linear regression
   (C) LASSO regression (D) Regularizer

8. Percentage of split for training and testing in hold out may be — 1 2 1 1
   (A) 80% and 20% (B) 70% and 30%
   (C) It can be (A) or (B) (D) 50% and 50%

9. Ridge regression handles — 1 2 1 1
   (A) Over fitting (B) Under fitting
   (C) Increases models performance (D) Decreases models performance

10. Impact of preprocessing step results in — 1 2 1 1
   (A) Dimensionality reduction (B) Computational cost and memory reduction
   (C) Both (A) and (B) (D) Removes variance of a model

11. SVM stands for — 1 2 1 1
   (A) Support vector machine (B) Soft vector machine
   (C) Soft vector model (D) Supervised vector machine

12. Baye's rule helps to find — 1 2 1 1
   (A) PDF (B) A posterior probability
   (C) Likelihood probability (D) A primi probability

13. Clustering is used in — 1 2 1 1
   (i) Medical images
   (ii) Market segmentation
   (A) Both (i) and (ii) (B) Only (i)
   (C) Only (ii) (D) Neither (i) nor (ii)

14. Pick the clustering algorithm that is most sensitive to outliers — 1 2 1 1
   (A) K-means (B) K-modes
   (C) K-median (D) K-medoids

15. Euclidean distance between data points (5, 6) and (6, 5) is — 1 2 1 1
   (A) $\sqrt{2}$ (B) 2
   (C) $2\sqrt{2}$ (D) 1

16. Identify the bottom-up clustering — 1 2 1 1
   (A) Agglomerative (B) Divisive
   (C) Feature reduction (D) Feature grouping

17. Entropy value of zero implies all member belong to _____. — 1 2 1 1
   (A) Same class (B) Different class
   (C) Inter class (D) Intra class

18. ID3 represents — 1 2 1 1
   (A) Iterative Dichotomister 3 (B) Iterative Decision tree 3
   (C) Internal Decision tree 3 (D) Iterative Differential tree 3

19. Perceptron can be represented as — 1 2 1 1
   (A) Summation and activation (B) Summation and deactivation
   (C) Differentiation and activation (D) Differentiation and deactivation

20. Random forest consumes — 1 2 1 1
   (i) Less memory
   (ii) More memory
   (iii) Collection of DT
   (A) Only (i) (B) Only (ii)
   (C) Both (ii) and (iii) (D) Both (i) and (iii)

PART – B (5 × 4 = 20 Marks) — Marks BL CO PO
Answer **ANY FIVE** Questions

21. When a ML module is said to be over fitted and under fitted. — 4 3 1 2

22. Explain the explosive nature of data dimension increased in machine learning. — 4 2 2 2

23. Write Python-Scikit script for linear regression on training and testing (take 4 data instance, one explanatory variable and response variable). — 4 4 2 2

24. Given $\beta = 0.7762, \bar{y} = 12.9, \bar{x} = 11.2$. Find the y-intercept $\propto$. The regressor model $y = \alpha + \beta x$. — 4 4 2 2

25. Write about Naïve Baye's classifier. — 4 3 2 2

26. Compute the distance matrix for the data points (Use Euclidean distance). — 4 3 2 2

	x	y
P_1	0	2
P_2	2	0
P_3	3	1
P_4	5	1

27. Write about random forest method. — 4 2 2 2

PART – C (5 × 12 = 60 Marks) — Marks BL CO PO
Answer **ALL** Questions

28. a. Compare supervised, unsupervised and reinforcement learning methods. — 12 3 1 1

(OR)

b. Write with example the method of learning and policy in reinforcement learning. — 12 3 1 1

29. a. For the given training data, fit the linear line and obtain $\hat{y}$. — 12 3 2 2

Training data	
X	Y
6	7
8	9
10	13
14	17.5
18	18