School of Computer Science and Engineering

Winter Semester 2023-24

Continuous Assessment Test – II

SLOT: C1+TC1

Programme Name & Branch: 5 Yr Int. MTech (CSE)

Course Name & Code: Machine Learning for Data Science (MDI4001)

Class Number (s): 2477, 2478

Faculty Name (s): Dr. Anisha M. Lal, Dr. Jyotismita Chaki

Exam Duration: 90 Min.   Maximum Marks: 50

General instruction(s):

Kindly allow scientific calculator in the examination hall.

Q. No.	Question	Max Marks
1.	Consider the following dataset. <table><tr><td>Group</td><td>X</td><td>Y</td></tr><tr><td>A</td><td>1</td><td>4</td></tr><tr><td>A</td><td>2</td><td>5</td></tr><tr><td>A</td><td>3</td><td>5</td></tr><tr><td>A</td><td>3</td><td>4</td></tr><tr><td>B</td><td>6</td><td>1</td></tr><tr><td>B</td><td>4</td><td>0</td></tr><tr><td>B</td><td>5</td><td>2</td></tr><tr><td>B</td><td>5</td><td>1</td></tr></table> a) Draw the decision boundary using SVM (calculations are not needed). [2] b) Consider the SVM classifier. Is feature scaling needed for the above dataset? Justify your decision with proper visualization. [8]	10
2.	Consider the following dataset. <table><tr><td>Class</td><td>X1</td><td>X2</td></tr><tr><td>A</td><td>1</td><td>1</td></tr><tr><td>A</td><td>-1</td><td>1</td></tr><tr><td>A</td><td>-1</td><td>-1</td></tr><tr><td>A</td><td>1</td><td>-1</td></tr><tr><td>B</td><td>2</td><td>0</td></tr><tr><td>B</td><td>0</td><td>2</td></tr></table>	10

B	-2	0
B	0	-2

Draw the hyperplane which will separate class A and class B by using SVM.

You can consider the following equations to convert data from one feature space to another feature space (iff needed).

$$\Phi\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{cases} 6 - X + (X - Y)^2 & \text{if } X = \sqrt{X^2 + Y^2} \geq 2 \\ \begin{pmatrix} X \\ Y \end{pmatrix} & \text{otherwise} \end{cases}$$

3.	Consider the following training dataset and discuss the difference between random forest and extremely randomized trees.	10

X1	X2	X3	X4	Class
A1	B1	C1	D1	1
A2	B2	C2	D2	2
A3	B3	C3	D3	1
A4	B4	C4	D4	1
A5	B5	C5	D5	2

4.	a) Consider the following dataset and create a CART decision tree. [The answer should contain all steps]. [9] b) Validate the decision tree by using the first sample. [1]	10

Age	Job	Credit	Loan approval
Young	No	Fair	No
Young	No	Good	no
Young	Yes	Good	Yes
Young	No	Fair	No
Middle	No	Fair	No
Middle	No	Good	No
Old	Yes	Good	Yes
Old	Yes	Excellent	Yes
Old	No	Fair	No

5.	Discuss the concept of stacking ensemble learning with respect to the dataset mentioned in Qs.1	10