VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

Reg. Number:	22BAI1266

Continuous Assessment Test (CAT) – I - FEB 2024

Programme	:	B.Tech.(CSE)	Semester	:	Winter 23-24
			Code		BCSE209L
Course Code & Course Title	:	Machine Learning	Class Number	:	CH2023240501696, CH2023240501692, CH2023240503074, CH2023240503363
Faculty	:	Dr. Syed Ibrahim S P (50392), Dr. Sajidha S A (50063), Dr. Mohan R (51012), Dr. Raghukiran (50593)	Slot	:	C1+TC1
Duration	:	1½ Hours	Max. Mark		50

General Instructions:

Write only your registration number on the question paper in the box provided and do not write other information.

- Use statistical tables supplied from the exam cell as necessary
- Use graph sheets supplied from the exam cell as necessary
- Only non-programmable calculator without storage is permitted

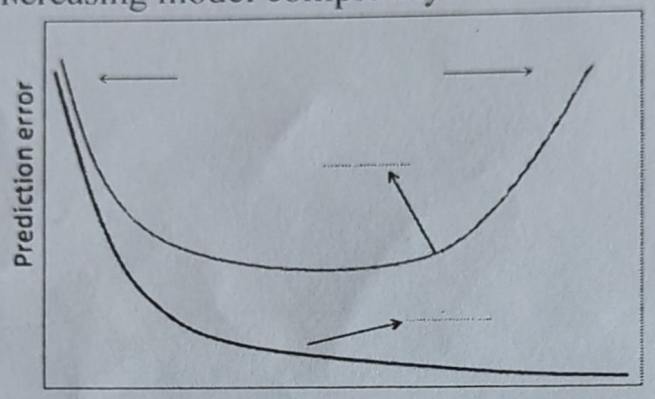
Answer all questions

Q. No	Sub Sec.	Description	Marks
1.	i)	Identify to which machine learning technique does each of the problem belong to. (5 marks) **Problem** a) User preference of news interest can change frequently, therefore recommending news to users based on reviews and likes could become obsolete quickly. The system must be able to track the change in the readers behaviour and recommend the best possible news appropriately where learning must happen based on rewards and punishments. b) Helpdesk assigns tags to each of the conversations with the customers. This is done for quick and easy navigation between previous customers' requests and for grouping conversations by topics. This process should be automated to reduce manual work. c) Every day a lot of employees work in different supermarket chains with money, so the owner wants to get a full picture of employees' performance, being able to evaluate the efficiency of operational costs. Imagine that the business owner intends to analyse employees' performance and identify who is not working very hard.	10

d) Create a model that will be able to predict when and how many products to buy for a restaurant, considering the expiration date of different products. To make a workable model, we'd need to feed it with the historical data: The number of restaurant dishes that were sold during the past periods (grouped by days, weeks, etc.), Holiday info (these days have other specifics), Marketing campaign info.

e) Facebook uses a type of learning which works on the basis of learning from feedback and giving appropriate rewards to optimise large scale production systems to personalise suggestions, deliver more meaningful notifications to users, optimise video streaming quality.

ii) The following figure depicts training and validation curves of a learner with increasing model complexity. (5 marks)



a) Which of the curves is more likely to be the training error and which is more likely to be the validation error? Indicate on the graph by filling the dotted lines. (2 marks)

b) In which regions of the graph are bias and variance low and high? Indicate clearly on the graph with four labels: "low variance", "high variance", "low bias", "high bias". (2 marks)

c) In which regions does the model overfit or underfit? Indicate clearly on the graph by labelling "overfit" and "underfit". (1 mark)

2. NASA wants to discriminate between Martians(M) and Humans(H) based on the following characteristics: Green $\in \{N, Y\}$,Legs $\in \{2,3\}$,Height $\in \{S,T\}$,Smelly $\in \{N,Y\}$. Our available training data is as follows: — 20

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

a) Generate a classification model having a tree structure using appropriate Supervised machine learning algorithm. (18 marks)

b) From the tree generated in 2(a), identify the species for the given anonymous data - {Y,3, Y, Y}. (2 marks)

3. Consider the following set of training examples

Table 1. Training data

X1	X2	Class (y)
0.08	0.72	100
0.1	1	100
0.26	0.58	200
0.35	0.45	200

10

a) For the given data in Table 1, design a suitable neural network with an appropriate activation function and show the training steps. (7 Marks)

b) Compare the training data accuracy level after 1st and 2nd epoch. (3 Marks)

4. Consider the following set of training examples:

Table 2. Training Data

X1	X2	Class (y)
2	7	H
3	1	H
3.5	1	H
4.5	5	M
5	6	M

10

Identify the class of the anonymous data (4,2) using the neighbourhood approach using the data given in Table1. Choose the number of neighbours to be 3 in order to decide the class.

**********All the best **********