INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
UNA [HP]
An Institute of National Importance under MoE
Saloh, Una (HP) – 177 209
Website: www.iiitu.ac.in

AY 2022-23
School of Computing
CURRICULUM: IIITUGCSE20
End Semester Examination
07, Dec.'22
(09:00 AM – 12:00 PM)

Degree	B. Tech.	**Branch**	CSE/IT/ECE
Semester	V		
Subject Code & Name	CSSE11: Machine Learning		
Time: 3 Hours	**Answer All Questions**		**Maximum: 100 Marks**

Sl. No.	Question	Marks
1.a	Pick a learning task from real world. Describe it informally in a paragraph in English. Now, describe it by stating as precisely as possible the task, performance measure, and training experience in terms of machine learning. Finally, propose a target function to be learned and a target representation. Discuss the main tradeoffs in formulating this learning task.	(5)
1.b	Can machine learning be termed as problem of function approximation in mathematics? Explain the reasoning in detail with suitable example.	(5)
1.c	Discuss some emerging trends in the field of machine learning.	(5)
1.d	Explain the difference between Classification and Regression.	(5)
2.a	Define the following terms: i. Inductive learning hypothesis ii. Version space iii. Consistent Hypothesis iv. General Boundary v. Specific boundary	(5)
2.b	What is regularization? Define cost function for Regularized linear regression.	(5)
2.c	Is Feature Selection necessary for simple linear regression? Explain with an example.	(5)
2.d	What is cross validation? How does it improve the accuracy of the outcome?	(5)
3.a	List down any five problem domains in which Decision Trees are most suitable.	(5)
3.b	ISRO wants to discriminate between Martians (M) and Humans (H) based on the following features: Green E {N,Y}, Legs E {2,3}, Height E {S,T}, Smelly E {N,Y}. The training data is as follows:	(5)

Species	Green	Legs	Height	Smelly
M	N	3	S	Y
M	Y	2	T	N
M	Y	3	T	N
M	N	2	S	Y
M	Y	3	T	N
H	N	2	T	Y
H	N	2	S	N
H	N	2	T	N
H	Y	2	S	N
H	N	2	T	Y

i. Which attribute will be the root of the decision tree?
ii. What is the information gain due to the attribute found in the question 3.b.i.?
(log values can be kept intact in calculations.)

3.c	What are the alternative measures for selecting attributes as root node in building decision trees? Explain each with formula.	(5)
3.d	A dataset has 3 binary features A, B, and C which take values 0/1. A data scientist wants to learn a function which counts the number of features which have value 1. Draw the decision tree which represents this function. How many leaf nodes does it have?	(5)
4.a	Discuss and define different distance measures used in clustering?	(5)
4.b	What is K in K-means algorithm and give its significance? List the pros and cons of K-means clustering algorithm.	(5)
4.c	Cluster the objects 1 to 10 applying k Medoids clustering using Manhattan Distance as the distance metric for calculating the distance between two points (x1,y1) and (x2,y2).	(5)
4.d	What is the purpose of using cluster analysis in data science? List any five applications of clustering.	(5)
5.a	Define Markov Decision Process of Reinforcement Learning in detail.	(5)
5.b	How to define States in Reinforcement Learning? What is the difference between a Reward and a Value for a given State?	(5)
5.c	How would one predict who will renew a Netflix subscription next month? What data is required to solve this? Can predictive models be built that predicts the subscription? If so, specify the algorithms?	(5)
5.d	Which algorithm is used to recommend the following seen on advertisement: 'People who bought this also bought…' Explain in detail with suitable workflow diagram of the same process.	(5)

Object	x	y
1	9	6
2	10	4
3	4	4
4	5	8
5	3	8
6	2	5
7	8	5
8	4	6
9	8	4
10	9	3