GURU NANAK INSTITUTE OF TECHNOLOGY
An Autonomous Institute under MAKAUT
2022
MACHINE LEARNING
CS802B

TIME ALLOTTED: 3 HOURS **FULL MARKS: 70**

The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable

GROUP – A
(Multiple Choice Type Questions)

Answer any ***ten*** from the following, choosing the correct alternative of each question: **10×1=10**

			Marks	CO No
1.	(i)	Which of the following are ML methods? a. Based on human supervision b. Supervised Learning c. Semi-reinforcement Learning d. All of the above	1	CO1
	(ii)	The most widely used metrics and tools to assess a classification model are: a. Confusion matrix b. Cost-sensitive accuracy c. Area under the ROC curve d. All of the above	1	CO1
	(iii)	Which of the following is a disadvantage of decision trees? a. Factor analysis b. Decision trees are robust to outliers c. Decision trees are prone to be overfit d. None of the above	1	CO1
	(iv)	Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? a. Decision Tree b. Random Forest c. Regression d. Classification	1	CO3
	(v)	In what type of learning labelled training data is used a. unsupervised learning b. supervised learning c. reinforcement learning d. active learning	1	CO2
	(vi)	Data used to build a machine learning model is called---------- a. training data b. validation data c. test data d. hidden data	1	CO2
	(vii)	The problem of finding hidden structure in unlabeled data is called… a. supervised learning b. unsupervised learning c. reinforcement learning d. none of the above	1	CO2

(viii)	Following is powerful distance metrics used by Geometric model a. Euclidean distance b. Manhattan distance c. both a and b d. Square distance		1	CO3
(ix)	A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is a. nominal b. ordinal c. categorical d. boolean		1	CO4
(x)	Machine learning is a subset of ..... a. Deep Learning b. Artificial Intelligence c. Data Learning d. None of the above		1	CO4
(xi)	A Machine Learning technique that helps in detecting the outliers in data. a. Clustering b. Classification c. Anomaly Detection d. All of the above		1	CO2
(xii)	What are the three types of Machine Learning? a. Supervised Learning b. Unsupervised Learning c. Reinforcement Learning d. All of the above		1	CO3

GROUP – B
(Short Answer Type Questions)
Answer any ***three*** from the following: **3×5=15**

			Marks	CO No
2.		Mention stepwise the workings of K-means algorithm.	5	CO1
3.	(a)	What is the role of Mean Squared Error (MSE) in ML? Write in brief.	3	CO2
	(b)	Define the roles of gradient descent in linear regression.	2	CO2
4.	(a)	What are the categories of Hyperparameters in ML model?	4	CO1
	(b)	Write difference between parameter and hyperparameter.	1	CO3
5.		What are the types of kernel available in ML. Write with its formula (any five).	5	CO1
6.		Write a short note on Normal Distribution concept in ML.	5	CO3

GROUP – C
(Long Answer Type Questions)
Answer any ***three*** from the following: **3×15=45**

			Marks	CO No
7.	(a)	What are various types of data are handled in ML? Write down with example.	6	CO2
	(b)	What are the significances of logistic function? Show with an example (diagram).	5	CO4
	(c)	Write short note on precision, recall, accuracy, F-measure in ML.	4	CO3

8. (a) 12 CO4

Age	Competition	Type	Profit
Old	Yes	s/w	Down
Old	No	s/w	Down
Old	No	h/w	Down
Mid	Yes	s/w	Down
Mid	Yes	h/w	Down
Mid	No	h/w	Up
Mid	No	s/w	Up
New	Yes	s/w	Up
New	No	h/w	Up
New	No	s/w	Up

Target attribute is "Profit". Now find out the "gain" of three attributes Age, Competition, and Type using **decision tree classifier** algorithm formula.

(b) Draw a comparative study of clustering and classification with example. 3 CO4

9. (a) 8 CO2

Weight	Height	Class
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

On the basis of given data, classify the below set as normal or underweight using KNN algorithm.

57	170	?

(b) Briefly explain hyperplane of SVM in ML. 2 CO2

(c) State linear and non-linear SVM with example 5 CO1

10. (a) 8 CO3

Height	Weight
185	72
170	56
168	60
179	68
182	72
188	77
180	71
180	70
183	84
180	88
180	67
177	76

Use **K-means clustering** algorithm form two clusters (k1, k2) and put those data points to the appropriate clusters.

(b) 7 CO3

No.	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

If the weather is sunny, then the Player should play or not? Using **Naïve Bayes classifier** algorithm, solve the problem.

11. (a) How does Random Forest algorithm work? State with an example. 7 CO4

(b) Mention the advantages and disadvantages of this algorithm 3 CO3

(c) Write down the formulae of four distance measurements namely Euclidean, Squared Euclidean, Manhattan, Cosine distance measurements. 5 CO2