GUJARAT TECHNOLOGICAL UNIVERSITY
PGDDS – SEMESTER I - EXAMINATION –WINTER-2021

Subject Code: 1618003   **Date: 19/03/2022**
Subject Name: Fundamentals of Machine Learning
Time: 10:30 AM to 01:00 PM   **Total Marks: 70**

Instructions:

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

			MARKS
Q.1	**(a)**	Define following terms with respect to machine learning: (i) precision (ii) overfitting (iii) concept learning	03
	(b)	What are the issues in machine learning as far as designing of learning system is concerned?	04
	(c)	A sample dataset about stolen cars is given below:	07

Sample No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Yellow	Sports	Domestic	No
3	Red	Sports	Imported	Yes
4	White	Sports	Imported	No
5	White	Family	Imported	Yes
6	White	SUV	Domestic	No
7	Yellow	SUV	Imported	Yes
8	Yellow	Family	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Domestic	Yes

Find the probability of a stolen car which is white in color, sports type and origin domestic. Use naïve bayes classification.

Q.2	**(a)**	Define following terms with respect to machine learning: (i) hypothesis (ii) hypothesis space (iii) hypothesis space search	03
	(b)	Briefly explain the steps to prepare the version space for given training dataset.	04
	(c)	Is given statement true or false? Reduced error pruning is following pre-pruning approach. Explain one-by-one step to perform reduced error pruning to get optimized decision tree.	07

OR

	(c)	With suitable example, show one-by-one step of backpropagation algorithm on multilayer feedforward artificial neural network.	07
Q.3	**(a)**	Define following terms with respect to machine learning: (i) target function (ii) inductive learning hypothesis (iii) version space	03
	(b)	Explain inductive bias with suitable example.	04
	(c)	What is semi-supervised learning? Explain the steps of self-training algorithm.	07

OR

Q.3 **(a)** Define following terms with respect to machine learning:
(i) consistent hypothesis (ii) underfitting (iii) mean square error **03**

(b) Compare supervised learning with unsupervised learning. **04**

(c) What is perceptron? What is perceptron training rule? What are its limitations and by which rule these limitations can be overcome? **07**

Q.4 **(a)** For which type of problem characteristics, decision tree algorithm is suitable? **03**

(b) Define following terms with respect to machine learning:
(i) gradient descent (ii) learning rate (iii) cost function (iv) epoch **04**

(c) A sample dataset about diagnostic is given below: **07**

Sample No.	Fever	Vomiting	Diarrhea	Shivering	Diagnosis
1	no	no	no	no	healthy
2	average	no	no	no	influenza
3	high	no	no	yes	influenza
4	high	yes	yes	no	salmonella poisoning
5	average	no	yes	no	salmonella poisoning
6	no	yes	yes	no	bowel inflammation
7	average	yes	yes	no	bowel inflammation

Using gain ratio, find the purest attribute that can be at root of the decision tree.

OR

Q.4 **(a)** Define: (i) prior probability (ii) posterior probability (iii) maximum likelihood **03**

(b) State Bayes theorem and its applications. **04**

(c) What is EM algorithm? Explain it in detail. **07**

Q.5 **(a)** What is Bayesian Belief Network? **03**

(b) Briefly explain minimum description length principle. **04**

(c) Define: precision, recall. **07**
Make confusion matrix from given data and find precision, recall, accuracy and F-score from it:
No. of samples in test data having class value POSITIVE = 150
No. of samples in test data having class value NEGATIVE = 600
No. of POSITIVE samples of test data classified as POSITIVE by model = 100
No. of NEGATIVE samples of test data classified as NEGATIVE by model = 550

OR

Q.5 **(a)** Briefly explain Bayes optimal classifier. **03**

(b) What is recommender system? How machine learning methods can be useful in it? **04**

(c) Define: sensitivity, specificity. **07**
Make confusion matrix from given data and find sensitivity, specificity, accuracy and F-score from it:
No. of samples in test data having class value RED = 250
No. of samples in test data having class value GREEN = 450
No. of RED samples of test data classified as RED by model = 200
No. of GREEN samples of test data classified as GREEN by model = 400

****