Course: MCSE1.4.1 MACHINE LEARNING (ELECTIVE I)
Date: 22nd Nov.2021

Semester: I
FIRST TEST

Year: Oct.-Jan. 2022
Total Marks: 50

Time: 120 minutes

1)	Explain the final design of the checkers learning system. Illustrate the sequence of design choices made for the checkers program	10
2)	Define Consistent Hypothesis and Version Space. Write the Candidate-Elimination algorithm. Will the candidate elimination algorithm converge to the correct hypothesis if the training data contains error? Justify your answer using an example.	06
3)	Consider the following data which represents the target concept "Smile?"	

Eyes	Nose	FColor	Hair?	Smile?
Square	Square	Green	Yes	No
Round	Triangle	Green	No	Yes
Round	Triangle	Purple	Yes	No
Square	Triangle	Yellow	Yes	No
Round	Square	Yellow	No	Yes
Square	Square	Purple	Yes	No

Calculate the number of concepts, number of syntactically distinct hypothesis. number of semantically distinct hypothesis. Compute the version space.

4)	Explain the inductive biased hypothesis space. unbiased learner and the futility of Bias Free Learning. Describe the three types of learning algorithms.	05
5)	Discuss Inductive Bias in Decision Tree Learning. Differentiate between two types of biases. Why prefer Short Hypotheses?	05
6)	Construct the Decision tree for the following training data using Gain Ratio as the measure of impurity:	10

Visitor	Meal	Cost	Vegan	Allergic Reaction
Alice	Dinner	Cheap	No	Yes
Alice	Lunch	Expensive	Yes	Yes
Bob	Lunch	Cheap	No	No
Alice	Lunch	Expensive	No	No
Bob	Dinner	Expensive	Yes	Yes
Alice	Dinner	Cheap	No	Yes
Alice	Lunch	Expensive	Yes	Yes
Bob	Dinner	Expensive	Yes	No
Alice	Dinner	Cheap	No	Yes
Alice	Lunch	Expensive	Yes	Yes
Alice	Lunch	Expensive	Yes	Yes
Bob	Lunch	Cheap	No	No
Alice	Lunch	Expensive	No	No
Bob	Dinner	Expensive	Yes	Yes

7)	Discuss the issues in learning decision trees. Explain briefly how are they overcome?	10

Total No. of Printed Pages:3

M.E.(Computer Science and Engineering) Semester-I
EXAMINATION MARCH 2022
Machine Learning

[Duration : Three Hours]   [Total Marks : 100]

Instructions:-

          1) Attempt any five questions.
          2) Make suitable assumptions if required

Q.1   a) Consider the problem of online shopping of groceries. Explain the steps in designing   10
          a program to learn to purchase groceries using Internet. List the available design
          choices at each step.

    b) Consider the following data: 10

Training	running nose	coughing	reddened skin	Classification
d1	+	+	+	positive(ill)
d2	+	+	-	positive(ill)
d3	+	-	+	Negative(healthy)
d4	-	+	+	Negative(healthy)
d5	-	-	+	Negative(healthy)
d6	-	-	-	Negative(healthy)

    a) Calculate the number of syntactically distinct hypothesis
    b) Compute a maximally specific hypothesis using FIND-S algorithm
    c) Compute a version space containing all hypotheses consistent with the above
          examples using candidate elimination algorithm

Q.2   a) Train the given data using Decision Tree algorithm. Classify the test data instances as   14
          going to be stuck in Traffic Jam or not.

Weather	Time of week	Time of day	Traffic Jam
Clear	Workday	Morning	Yes
Clear	Workday	Lunch	No
Clear	Workday	Evening	Yes
Clear	Weekend	Morning	No
Rainy	Workday	Morning	Yes

Rainy	Workday	Lunch	Yes

			Yes
Rainy	Workday	Evening	
			No
Rainy	Weekend	Morning	
			No
Rainy	Weekend	Lunch	
			No
Rainy	Weekend	Evening	
			Yes
Snowy	Workday	Morning	
			Yes
Snowy	Workday	Lunch	
			Yes
Snowy	Workday	Evening	
			Yes
Snowy	Weekend	Morning	

Test Data:

Clear	Weekend	Lunch
Snowy	Weekend	Evening

b) With suitable illustration explain the following solutions to overfitting in decision trees:
   i) Stop growing the tree
   ii) Pruning using cross-validation   6

Q.3   a) List various issues in machine learning.   2

         b) For each of the following learning problems, identify the class of tasks, performance   12
         measures and the source of experience. Suggest an appropriate target function to
         model each:
            i)   Autonomous Mars Rover
            ii)   Fruit disease prediction
            iii) Signature recognition

         c) Will the output of candidate elimination algorithm change if we change the ordering
         of the training examples? Justify with the help of an example.
                                                                                                     6

Q.4   a) Illustrate Bayesian Belief Networks with the help of an example.

         b) Discuss the advantages and disadvantages of Bayesian Learning. A spam filtering   8
         system has the probability of 0.95 to classify correctly a mail as spam and 0.10   7
         probability of giving false positives. It is estimates that 0.5% of the mails are actual
         spam mails. Suppose that a new mail is classified as not-spam, what is the probability
         that the mail is actually not-spam?

         c) What is the Naïve assumption in Naïve Bayes Classifier? For the following set of   5
         attributes (graph depicting causality), can we apply Naïve Bayes Classifier? Justify.

2
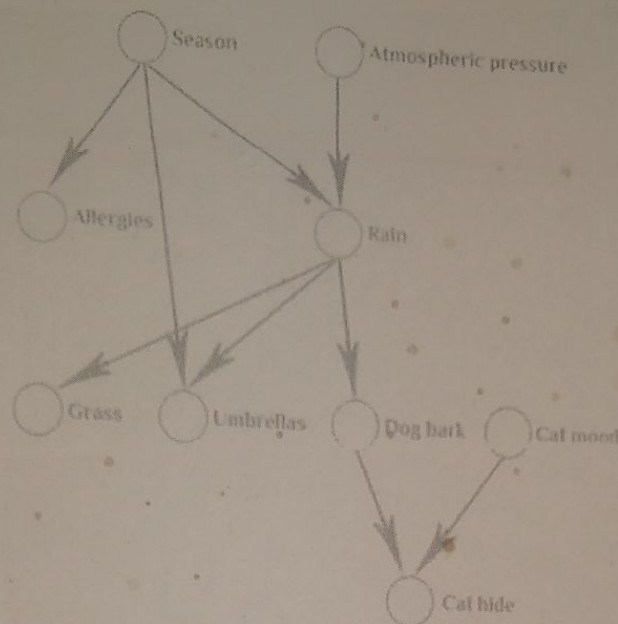
Q.5   a) Write a short note on PAC learnability   7

b) Discuss the Sample Complexity of Finite Hypothesis space. Comment of the   7
minimum number of training samples needed for a PAC learnable consistent learner,
given the following set of Boolean literals, error and probability thresholds.
1. $n= 100, \delta= \epsilon= 0.05$
2. $n=100, \delta= \epsilon = 0.03$
3. $n= 100, \delta = \epsilon= 0.01$

c) Discuss the Sample Complexity of an Infinite Hypothesis Space.   6

Q.6   a) Give three computer applications for which machine learning approaches seem   6
appropriate and three applications for which they seem inappropriate.

b) What is version space? What are the elements of the version space? How are they   4
ordered?

c) What do you understand by the minimum description length principle? Explain how it   5
deals with the hypothesis space.

d) List the problems under which decision tree learning is best suitable.   5