CBCS SCHEME

USN | | | | | | | | | | |

21CS54

Fifth Semester B.E. Degree Examination, Dec.2023/Jan.2024
Artificial Intelligence and Machine Learning

Time: 3 hrs. Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

1 a. Identify the Turing test approach to provide a satisfactory operational definition of Intelligence. **(04 Marks)**

b. Make use of the state space of the vacuum world and define the components to solve this problem. **(06 Marks)**

c. Illustrate the properties and the algorithm for Breadth-first search technique. **(10 Marks)**

OR

2 a. Explain the concepts of thinking rationally and acting rationally. **(04 Marks)**

b. Explain the tree search and graph search algorithms. **(06 Marks)**

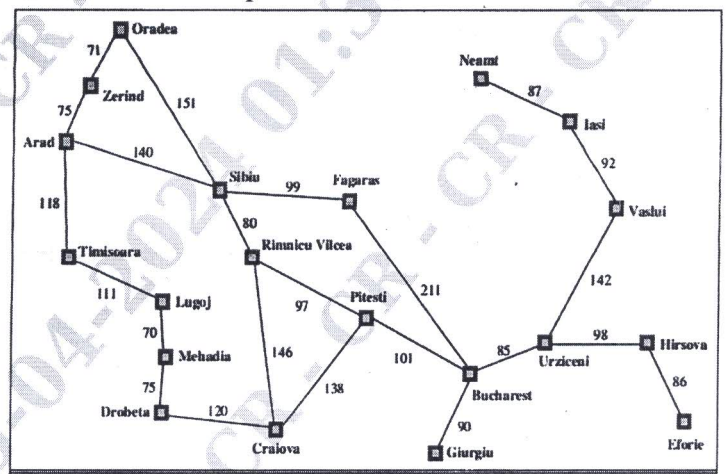
c. Explain problem solving agents alongwith the algorithm and illustrate the incremental formulation of 8-Queens problem. **(10 Marks)**

Module-2

3 a. Identify the differences between supervised and unsupervised learning. **(04 Marks)**

b. Explain the types of Big data. **(06 Marks)**

c. Apply A* algorithm to find the best path from Arad to Bucharest. [Refer Fig.Q3(c)].



A simplified road map of part of Romania.

Arad	366	Mehadia	241
Bucharest	0	Neamt	234
Craiova	160	Oradea	380
Drobeta	242	Pitesti	100
Eforie	161	Rimnicu Vilcea	193
Fagaras	176	Sibiu	253
Giurgiu	77	Timisoara	329
Hirsova	151	Urziceni	80
Iasi	226	Vaslui	199
Lugoj	244	Zerind	374

Values of h_{SLD}—straight-line distances to Bucharest.

Fig.Q3(c) **(10 Marks)**

OR

4 a. Explain the machine learning process model along with diagram. **(06 Marks)**

b. Consider the table given below which contains the machine learning course registration done by both boys and girls. There are 50 boys and 50 girls in the class and the registration of the course is given in the table. Apply Chi-square test and find out whether any differences exist between boys and girls for course registration.

Table 4(b)

Gender	Registered	Not Registered	Total
Boys	35	15	50
Girls	25	25	50
Total	60	40	100

(06 Marks)

c. Apply the heuristic search algorithm on the given 8 puzzle problem to reach the goal state from the given initial state.

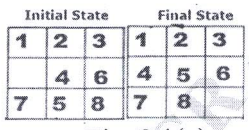


Fig.Q4(c) **(08 Marks)**

Module-3

5 a. Consider the training dataset of 4 instances shown in the table below. Apply Find-S algorithm to find the final hypothesis.

Table 5(a)

CGPA	Interactiveness	Practical Knowledge	Communication Skills	Logical Thinking	Interest	Job Offer
≥9	Yes	Excellent	Good	Fast	Yes	Yes
≥9	Yes	Good	Good	Fast	Yes	Yes
≥8	No	Good	Good	Fast	No	No
≥9	Yes	Good	Good	Slow	No	Yes

(08 Marks)

b. Explain why Instance based learners are called lazy learners and compare instance based learning and model based learning. **(06 Marks)**

c. Explain the types of Regression methods with diagram. **(06 Marks)**

OR

6 a. Consider the student performance training dataset of 8 data instances in the below table. Based on the performance of a student, classify the test instance (6.1, 40, 5) to check whether the student will pass or fail in that course using KNN approach (K = 3).

Table 6(a)

S.No.	CGPA	Assessment	Project Submitted	Result
1	9.2	85	8	Pass
2	8	80	7	Pass
3	8.5	81	8	Pass
4	6	45	5	Fail
5	6.5	50	4	Fail
6	8.2	72	7	Pass
7	5.8	38	5	Fail
8	8.9	91	9	Pass

(12 Marks)

b. Explain version space and the candidate elimination algorithm explaining the algorithm steps. **(08 Marks)**

Module-4

7 a. Explain the advantages and disadvantages of decision trees. **(06 Marks)**

b. Explain validating and pruning of decision trees. **(06 Marks)**

c. Explain Bayes optimal classifier and solve to find whether a patient is diagnosed as COVID positive or COVID negative using the table given below.

Table 7(c)

$P(h_i/T)$	P(COVID positive)	P(COVID negative/h_i)
0.3	0	1
0.1	1	0
0.2	1	0
0.1	1	0

(08 Marks)

OR

8 a. Explain the procedure to construct a decision tree using ID3 algorithm. **(06 Marks)**

b. Explain Bayes theorem, Maximum A Posteriori (MAP) Hypothesis (h_{MAP}) and Maximum Likelihood (ML) Hypothesis (h_{ML}). **(06 Marks)**

c. Illustrate the algorithm of Naïve Bayes and explain the popular variants of Bayesian classifier. **(08 Marks)**

Module-5

9 a. Explain the different activation functions used in ANN. **(06 Marks)**

b. Illustrate the various types of Artificial Neural Networks. **(08 Marks)**

c. Illustrate the applications and challenges of Clustering algorithms. **(06 Marks)**

OR

10 a. Explain the perceptron model and the algorithm. **(08 Marks)**

b. Consider the following set of data given in the below table. Cluster it using K-means algorithm with the initial value of objects 2 and 5 with the coordinate values (4, 6) and (12, 4) as initial seeds.

Table 10(b)

Objects	X-coordinate	Y-coordinate
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4

(12 Marks)

* * * * *

Artificial Intelligence & Machine Learning

VTU Scheme & Solution

Dec 2023/ Jan 2024

semester : V sub code : 21CS54

Q.	Solution	Marks
1a.	Turing test explanation when a Computer passes the test with human interrogator capabilities of a Computer to pass the test i) Natural Lang. processing ii) Knowledge representation (iii) Automated reasoning (iv) Machine Learning etc.	1 3
1b.	State space of the Vacuum world showing the links to denote the actions L - Left and R - Right, S - suck. Components values to be mentioned States, Initial state, No. of actions, goal, path cost.	3 3
1c.	properties of BFS. Completeness - yes, optimality - if path cost non decreasing function of the depth node. Time complexity & Space complexity = $O(b^d)$ BFS alg : Complete alg. Frontier ← Insert (child, Frontier)	4 6

2a). Thinking rationally: "The laws of thought" approach, logic etc. 2

Action rationally: "The rational agent approach rational agent: def. & expl. 2

2b. Tree search algorithm: Complete alg.

function Tree-Search (problem) returns a solution or failure. 3

- -
- -

Graph search algorithm: with the explored set

function graph-search (problem) returns a solution or failure. 3

2c). It is a kind of goal-based agent problem-solving-agent algorithm with

- - -

goal ← Formulate-goal (state)

problem ← Formulate-problem (state, goal)

seq ← Search (problem)

Incremental formulation:

States - any arrangements of 0 to 8 squares (queens), Initial state, actions, Transition model, goal test.

3a. Any 4 differences blw supervised & unsupervised learning.

Supervised Learning	Unsupervised Learning
1. possess supervisor component	NO supervisor
2. Uses labelled data	Uses unlabelled data
3. Assign labels/categories	Performs grouping or clustering
4. classification & regression alg.	clustering alg.

4

3b. i) structured data. ii) Recorded data (iii) Data Matrix (iv) graph data v) Ordered data. (vi) Unstructured data. (vii) semi-structured data.

Explanation of any 6 - Each 1 Mark. 6

3c. Initial state → Arad - Graph showing After expand Arad, Sibiu, Rimnicu, Vilcea, Fagaras. 10

Path Arad → Sibiu → Fagaras → Bucharest

4a) i) Understanding the business ii) Understanding data (iii) preparation of data, (iv) modelling (v) Evaluate (vi) deployment. 4

Diagram 2

4b. $x^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 4.166$, categories = $2-1=1$ & 1

$0.04 < 0.05$

∴ the result is significant

4c. $f(x) = g(x) + h(x)$

$g(x)$ = no. of steps from current state from the initial state

$G(x)$ - no. of ways to reach the goal state from the current state &

5a. Step 1 : Initialize G to most specific hypothesis

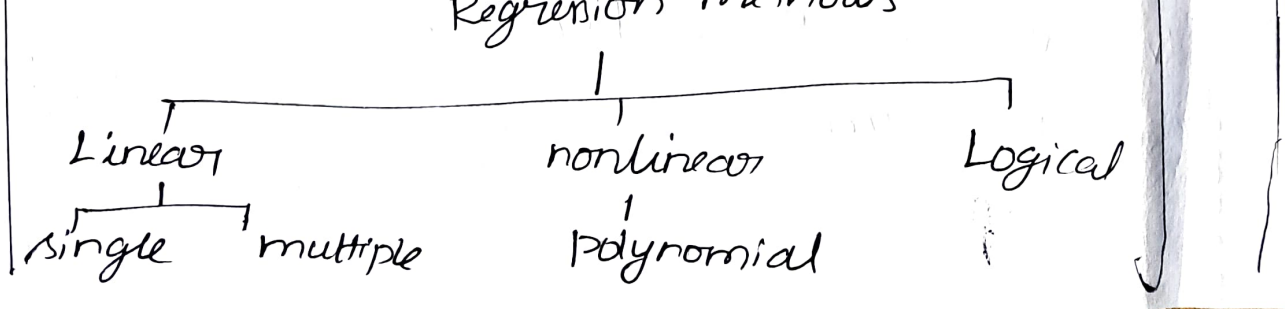
$h = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

step 2 & step 3 generalizing the hypothesis & Final hypothesis generated with Find-S

$h = \langle \geq 9, yes, ?, Good, ?, ? \rangle$

5b. Instance based learning is referred to as lazy learning since it does not generalize any model from the training data set but just keeps the training data 6 set as a knowledge base until a new instance is given

e).



6a. Step 1: Calculate the euclidean distance between the test instance (6.1, 40 & 5) and each of the training instances Table. 8

Step 2: Sort the distances in the ascending order and select the first nearest traing data instances to the test instance

Instance	Eluclidean distance	class
4	5.001	Fail
5	10.057	Fail
7	2.022	Fail

3

Step 3: predict the class of the test instance by majority voting as 'Fail' 1

b). Version space contains the subset of the hypotheses from the hypothesis space that is consistent with all training instances in the training dataset. 1

candidate elimination alg. 7

Input: Set of Instance in the traing dataset.

output: Hypothesis G & S.

7a. Advantages: Easy, simple, I/O attributes can be discrete or continous predictor variables, model a high degree of non-linearity, eucicle to train 3

disadvanges: Difficult to distermine how deeply a decision tree

can be grown, error on mission attributes then decision tree is unstable if data has continuous valued attributes, handling is complex, complex decision tree may be overfitting etc. 3

7b. Inductive bias in ID3 algorithm, overfitting problem with decision trees.

i). Prune the trees & construct an optimical decision tree. ii) Alter the tree using training set, statistical tests (iii) use a principle called minimum description length. 4

Mention three pruning methods - MEP, EBP, MDL 2

c). $\sum_{h_i \in H} P(\text{COVID negative} / h_i)\, P(h_i|T) = 0.3$

$\sum_{h_i \in H} P(\text{COVID positive} / h_i)\, P(h_i|T) = 0.4$ 6

$\max \sum_{h_i \in H} P(l_i|h_i)\, P(h_i|T) = \text{COVID Positive}$

$\max_{l_i} \sum_{h_i \in H} P(l_i|h_i)\, P(h_i|T)$ 2

8a). Algorith:

i) Compute entropy - info for the whole training dataset based on the target attribute ... 6

Compute procedure.

8b. Bayes theorem explanation and

$$P(H/E) = \frac{P(E/H)\,P(h)}{P(E)}$$ 2

MAP (h_{MAP}) explanation and

$h_{MAP} = \max_{h \in H} P(i/)\,P(h)$ 2

ML (h_{ML}) explanation $h_{ML} = \max_{h \in H} P(E/Gi)$ 2

H → Hypothesis E – Evidence.

8c. Algorithm: i) Compute the prior probability for the target class

Variants: i) Bernauli Naive Bayes Classifier
ii) Mutinominal Naive Bayes Classi
iii) Muti – class Naive Bayes Classifier. 4

9a. i) Identity or Linear function, ii) Binary step function. (iii) Bipolar step function (iv) sigmoidal function (v) Bipolar sigmoid (vi) Ramp (vii) Tan – hyperbolic tangent func. 6

9b) i) Feed forward – Neural n/w explain 2
ii) Fully connected Neural n/w explain 3
(iii) Multilayer perception (MLP) 3

c) Applications: i) grouping, (ii) profiling
(iii) Retrieval of info. (iv) Identifying eg. genes. 3

Challenges of clustering:

8b. Bayes theorem explanation and

$$P(H/E) = \frac{P(E/H)\,P(h)}{P(E)}$$

2

MAP (h_{MAP}) explanation and

$$h_{MAP} = \max_{h \in H} P(i/)\,P(h)$$

2

ML (h_{ML}) explanation $h_{ML} = \max_{h \in H} P(E/G)$ 2

$H \rightarrow$ Hypothesis E - Evidence.

8c. Algorithm: i) Compute the prior probability for the target class

Variants: i) Bernauli Naive Bayes classifier
ii) Mutinominal Naive Bayes classi
iii) Muti-class Naive Bayes classifier. 4

9a. i) Identify or Linear function, ii) Binary step function. (iii) Bipolar step function (iv) sigmoidal function (v) Bipolar sigmoid (vi) Ramp (vii) Tan- hyperbolic tangent func. 6

9b) i) Feed forward - Neural n/w explain 2
ii) Fully connected Neural n/w explain 3
(iii) Multilayer perception (MLP) 3

c) Applications: i) grouping, (ii) profiling
(iii) Retrieval of info. (iv) Identifying eg. genes. 3

Challenges of clustering:

10a) Diagram showing the perceptron model — 2

Explanation — 2

Algorithm - set initial weights $w_1 \dots w_n$ & bias θ to a random value in the range $[-0.5, 0.5]$ for each epoch i) Compute weighted sum...

10.b) Initial cluster Table. — 2

Cluster 1	Cluster 2
(4,6)	(12,4)

Iteration 1: Compare all the data points on samples with the centroid and assign to the nearest sample — 5

Cluster 1	Cluster 2
(4,6)	(10,4)
(2,4)	(12,4)
(6,8)	—

Iteration 2:

Cluster 1	Cluster 2
(4,6)	(10,4)
(2,4)	(12,4)
(6,8)	—
centroid (4,6)	centroid (11,4)