INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

End-Autumn Semester Examination 2023-24

Date of Examination: 16/11/2023 **Session:** AN **Duration:** 3 Hrs **Full Marks:** 101

Subject No.: AI61005 **Subject:** Artificial Intelligence: Foundations and Algorithms

Department/Center/School: Centre of Excellence in Artificial Intelligence

Specific charts, graph paper, log book etc., required: NO

Special Instructions: **Answer PART-A and PART-B in separate answer scripts.**

Answer all parts of a question in same place.

PART-A

1) Consider Table below, which contains data about house listings. We wish to build a decision tree classifier to predict whether it is acceptable to buy a listed house. We consider the two classes 'acc' and 'unacc'.

No.	Pricing	Condition	Rooms	Safety	Acceptability
P1	High	Great	Three	High	acc
P2	Low	Good	Three	Low	acc
P3	Low	Great	Two	High	acc
P4	Low	Good	Two	High	unacc
P5	High	Bad	Three	Low	unacc
P6	High	Great	Three	Low	unacc
P7	Low	Good	Two	Low	unacc
P8	High	Good	Three	Low	?
P9	Low	Great	Two	Low	?

(a). Calculate the Entropy of the root node of a decision tree with the given training data. [2]
(b). Build the decision tree using the ID3 algorithm using information gain. Show all steps. Assume log base is 2. [10]
(c). Use the decision tree to classify the unknown points in the Table (P8 and P9). Show the calculations. [2]
(d). Information Gain has the disadvantage of forming several small but pure partitions. State an approach to overcome this. [3]
(e). Assume there are three binary features used to represent a training instance and it is a two class classification problem. Training set contains 15 positive and 15 negative examples. What is the minimum and maximum number of leaves in the tree? Show the computation. [3]

2) Consider the data in the following Table which consists of house prices based on the area and distance from the city. We wish to build a linear regression estimator based on these data.

Sr. No	Area of land (m^2)	Dist. from city (Km)	Price (million INR)
P1	25	12	13
P2	30	10	18
P3	28	25	19
P4	21	2	17
P5	35	15	?

(a). Write the hypothesis function, and the equations for updating the different parameters of this estimator in the iterative approach. [3]
(b). Use the training data to calculate the parameters after two iterations of gradient descent. Take the initial values of the parameter as 0.5 each and learning rate as 0.001. [6]

(c). Use the parameter values so trained to predict the price of the unknown example [1]
(d). State a reason why linear regression is not ideal for classification. How can Logistic Regression be used to overcome this problem? [4]
(e). Why is mean squared error not used in Logistic Regression? [3]
(f). Suppose we have to classify a book into one of four classes. Explain how to use logistic regression for this task? [3]

3) You are in charge of scheduling for computer science classes that meet Mondays, Wednesdays and Fridays. There are 5 classes that meet on these days and 3 professors who will be teaching these classes. You are constrained by the fact that each professor can only teach one class at a time.
The classes are:
- Class 1 - Programming: meets from 8:00-9:00am
- Class 2 - Artificial Intelligence: meets from 8:30-9:30am
- Class 3 - Natural Language Processing: meets from 9:00-10:00am
- Class 4 – Information Retrieval: meets from 9:00-10:00am
- Class 5 - Machine Learning: meets from 9:30-10:30am

The professors are:
- Professor A, who is available to teach Classes 3 and 4.
- Professor B, who is available to teach Classes 2, 3, 4, and 5.
- Professor C, who is available to teach Classes 1, 2, 3, 4, 5.

(a). Formulate this problem as a CSP problem in which there is one variable per class, stating the domains, and constraints. Constraints should be specified formally and precisely, in an explicit way. [4]
(b). Draw the constraint graph associated with your CSP [2]

(c). Search for a solution using basic backtracking. Only check whether any new assignment violates no constraint with previous assignments. As a tie breaker assign a class to a professor based on <u>alphabetical order</u>. **Continue up to the first time you try and fail to assign <u>any value</u> for Class 5. [4]**

Fill out this worksheet (Table 1) as you draw your search tree. There may be more rows than you need.

(i). Every time you assign a variable or remove a variable from the propagation queue, fill out a new row in the table. (The same variable might appear in more than one row, especially if you have to backtrack.)

(ii). In that row, indicate **which variable you assigned or de-queued**; write its **assigned** value if it has one (e.g. X=x), otherwise just write its name (X). In the second column, list the **values that were just eliminated from neighboring variables** as a result. If no values were just eliminated, write **NONE** instead.
(iii). If your search has to backtrack after assigning or de-queuing a variable: first, **<u>finish listing</u>** all values eliminated from neighboring variables in the current row. Next, check the backtrack box in that row. Then, continue with the next assignment in the following row as usual.
(iv). At some point, you might add several variables to your propagation queue at once. Break ties by adding variables to your propagation queue **<u>in alphabetical order</u>**.

	Var assigned or dequeued	**List all values eliminated from neighbouring variables**	**Backtrack?**
Ex	X	$Y \neq B$ $Z \neq C$ (example)	☑
1			☐
2			☐

Table 1
(d). Fill the worksheet (Table 1) but now implement backtracking with Forward checking and Minimum Value Ordering. Break ties **<u>in alphabetical order</u>**. [5]

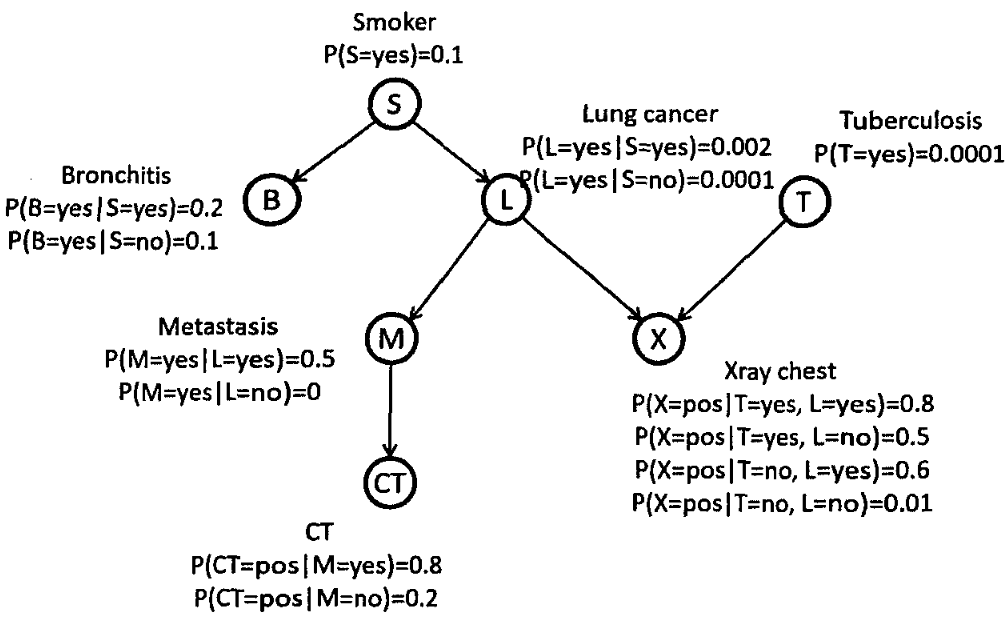
Part B [46 marks]

Answer Part B in a separate booklet.

1. (a) Self-supervised learning is used to generate representations in NLP. (4 marks)
   i. Show one example of a training data (comprising an input and the corresponding output) which can be used to construct a Word2vec model.
   ii. Give one example of a training data that is used to train a contextual representation model.

   (b) Which of these issues in NLP is/are addressed by subword embedding? (i) Ambiguity, (ii) Sparsity, (iii) Variability, (iv) Out of Vocabulary Words. (2 marks)

2. Consider the following Bayesian network and the associated CPTs.



   (a) State Bayes Rule (2 marks)

   (b) Compute $P(S = yes|B = Yes)$ (2 marks)

   (c) Compute the probability that a smoker does not have bronchitis, has lung cancer, does not have Tuberculosis and chest XRay is positive. (2 marks)

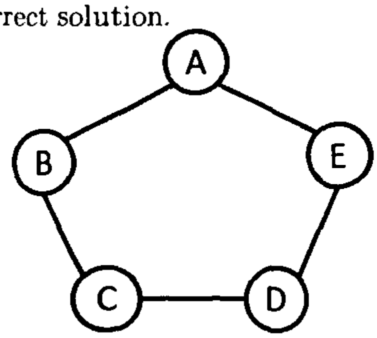
   (d) Compute the probability that a smoker does not have bronchitis, has lung cancer, and chest XRay is positive. (2 marks)

   (e) State whether the following conditional Independence relations hold. Write T if the independence holds, F if it does not. (3 marks)
   i. B and X given S
   ii. B and T
   iii. B and T given X

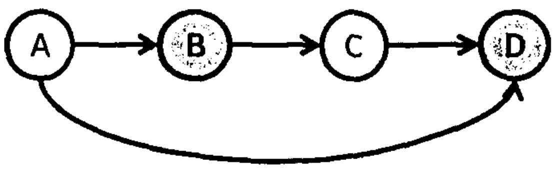
(f) Suppose I gave you the following Bayes net with nodes and edges, but I did not provide you with the direction of the edges. Given, (5 marks)

1. B is independent of E
2. E is independent of C given D

Assign a direction to every edge (by adding an arrowhead at one end of each edge) to ensure that the Bayes' Net structure implies the assumptions provided. You cannot add new edges. The Bayes' nets can imply more assumptions than listed, but they must imply the ones listed. There may be more than one correct solution.



(g) Consider the following Bayes Net, where we have observed that B = +b and D = +d. (4 marks)



$P(A)$	
$+a$	0.5
$-a$	0.5

$P(B\|A)$		
$+a$	$+b$	0.8
$+a$	$-b$	0.2
$-a$	$+b$	0.4
$-a$	$-b$	0.6

$P(C\|B)$		
$+b$	$+c$	0.1
$+b$	$-c$	0.9
$-b$	$+c$	0.7
$-b$	$-c$	0.3

$P(D\|A,C)$			
$+a$	$+c$	$+d$	0.6
$+a$	$+c$	$-d$	0.4
$+a$	$-c$	$+d$	0.1
$+a$	$-c$	$-d$	0.9
$-a$	$+c$	$+d$	0.2
$-a$	$+c$	$-d$	0.8
$-a$	$-c$	$+d$	0.5
$-a$	$-c$	$-d$	0.5

Consider doing Gibbs sampling for this example. Assume that we have initialized all variables to the values +a, +b, +c, +d. We then unassign the variable C, such that we have A = +a, B = +b, C = ?, D = +d. Calculate the probabilities for new values of C at this stage of the Gibbs sampling procedure.

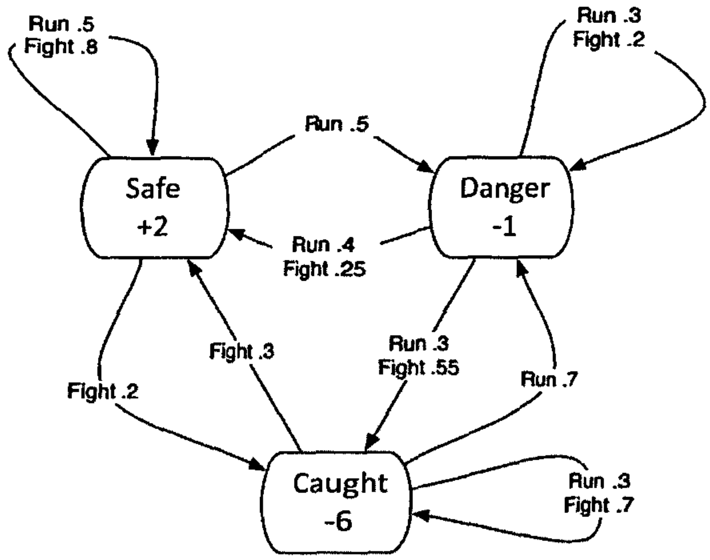
3. A boy is being chased around the school ground by a bully and must choose whether to Fight or Run. There are three states:

Safe (S) : he is safe for the moment.

Danger (D) : the bully is about to catch up.

Caught (C) : the bully catches up with him and snatches his tiffin.

He begins in state S. The graph of the MDP is given below:



(a) Suppose you run value iteration with a discount factor $\gamma = 0.8$. Fill up the table with the results of value iteration for the next time step. (6 marks)

t	$V^t(S)$	$V^t(D)$	$V^t(C)$
1	2	-1	-6
2			

(b) Assuming the above values of the states $V^2(S)$, $V^2(D)$ and $V^2(C)$ computed in (a), what is the implied policy π at this step? That is, write the values of $\pi(S)$, $\pi(D)$ and $\pi(C)$. (4 marks)

(c) Now, suppose the boy adopts the policy of Run from all states, that is, $\pi_1(S) = \pi_1(D) = \pi_1(C) = \textit{Run}$. Evaluate the value of each state given this policy π_1. That is, calculate $V_{\pi_1}(S)$, $V_{\pi_1}(D)$ and $V_{\pi_1}(C)$. (6 marks)

(d) Starting with π_1 as in (c), do a policy update to find the new policy π_2 (one step of policy iteration). That is, find the value of $\pi_2(S)$, $\pi_2(D)$ and $\pi_2(C)$. (4 marks)