Indian Institute of Technology (Indian School of Mines) Dhanbad
Department of Computer Science and Engineering
Principles of Artificial Intelligence (CSO303)
End Semester Exam, Date: November 21, 2024

Timing: 09:00 AM to 12:00 PM   Monsoon Semester (2024-2025)   Maximum marks: 100

1. (a) Compare and contrast the sigmoid, tanh, and ReLU activation functions in terms of their mathematical properties, advantages, disadvantages, and use cases.

   (b) Consider a multilayer perceptron with one hidden layer using the sigmoid activation function. Derive the gradient of the loss function with respect to the weights of: i. The output layer and ii. The hidden layer. Explain the significance of the chain rule in this derivation.

   (c) Discuss the limitations of the minimax algorithm in large-scale games like chess. How can techniques like heuristic evaluation functions and iterative deepening be integrated to make minimax practical for such complex games? Provide an example.

(6+6+3)

2. (a) Explain the difference between constraint-based and score-based approaches for learning the structure of Bayesian Networks. Provide one advantage and one disadvantage of each approach.

   (b) Define the concept of "premature convergence" in Genetic Algorithms. Propose two strategies to mitigate this issue.

   (c) GAs are often criticized for their high computational cost. Explain why this is the case and suggest atleast three ways to reduce the computational burden without compromising solution quality.

(5+5+5)

3. (a) You are given a map-colouring CSP with eight regions: A, B, C, D, E, F, G, and H. Each region must be coloured using one of three colours: Red, Green, or Blue. The adjacency constraints are as follows:

   i. A is adjacent to B, C, D.
   ii. B is adjacent to A, E, F.
   iii. C is adjacent to A, E, G.
   iv. D is adjacent to A, F, H.
   v. E is adjacent to B, C, G.
   vi. F is adjacent to B, D, H.
   vii. G is adjacent to C, E, H.
   viii. H is adjacent to D, F, G.

   Solve this CSP using the Backtracking algorithm with Degree Heuristic for variable ordering. Show each step in the assignment process, including where backtracking occurs. Additionally, explain how using the Least Constraining Value (LCV) heuristic and Minimum Remaining Value (MRV) could affect the outcome and efficiency of the solution.

   (b) Consider a CSP with seven variables A, B, C, D, E, F, and G, each having a domain of [1,2,3,4,5]. The constraints are:

   i. $A + B = C$
   ii. $B < D$
   iii. $C \neq E$
   iv. $D + E = F$
   v. $F > G$

Scanned with OKEN Scanner

vi. $G \neq A$

Solve this CSP using constraint propagation with arc consistency (AC-3) to reduce the domains of each variable. Show each step in the domain reduction process. After completing arc consistency, use backtracking to find a solution. Document each assignment and any backtracking steps and discuss how arc consistency helped reduce the search space before backtracking.

(8+7,

4. (a) Consider the following first-order logic (FOL) sentences:

i. $\forall x, y\, [P(x,y) \wedge Q(y,x) \Rightarrow R(x,y)]$

ii. $\forall x, y\, [S(x, \text{Bob}) \wedge S(y,x) \Rightarrow P(x,y)]$

iii. $\forall x, y\, [S(x,y) \Rightarrow Q(y,x)]$

iv. $\forall x, y\, [T(x,y,x) \Rightarrow Q(x,y)]$

v. $\forall x, y\, [T(x,x,y) \Rightarrow Q(x,y)]$

Facts:

T(Alice, Dawn, Alice)

T(Eve, Carl, Eve)

T(Alice, Bob, Dawn)

T(Carl, Carl, Alice)

S(Bob, Alice)

S(Carl, Bob)

S(Dawn, Carl)

S(Carl, Dawn)

S(Alice, Dawn)

S(Eve, Carl)

Use backward chaining to find **ALL** answers for the following queries. When matching rules, proceed from top to bottom, and evaluate subgoals from left to right.

Queries:

i. $\exists x\, Q(\text{Alice}, x)$ (Find all x such that $Q(\text{Alice}, x)$ holds.)

ii. $\exists x, y\, R(x,y)$ (Find all pairs (x,y) such that $R(x,y)$ holds.)

(b) Determine whether the expressions p and q unify with each other in each of the following cases. If so, give the most general unifier; If not, give a brief explanation (Assume that the upper case letters are (object, predicate, or function) constants and that the lower case letters are variables).

i. $p = F(G(v), H(u,v)); q = F(w, J(x,y))$

ii. $p = F(x, F(u,x)); q = F(F(y,A), F(z, F(B,z)))$

iii. $p = F(x1, G(x2,x3), x2, B); q = F(G(H(A,x5),x2), x1, H(A,x4), x4)$

(c) Convert the following English statements to statements in first order logic.

i. every boy or girl is a child

ii. every child gets a doll or a train or a lump of coal

iii. no boy gets any doll

iv. no child who is good gets any lump of coal

v. Jack is a boy

Using the above five axioms construct a proof by resolution of the statement: "If Jack doesn't get a train, then Jack is not a good boy".

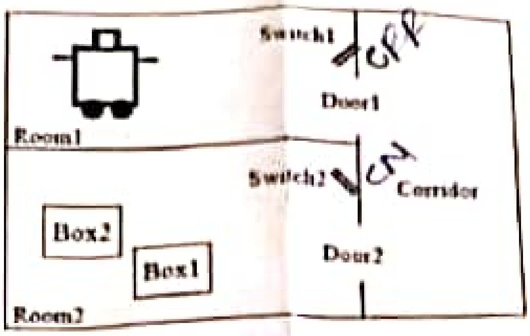
(5+5+5

5. Consider the world of Shakey the robot, as shown below. Shakey has the following six actions available:

i. Go(x,y), which moves Shakey from x to y. It requires Shakey to be at x and that x and y are locations in the same room. By convention a door between two rooms is in both of them, and the corridor counts as a room.

ii. Push(b,x,y), which allows Shakey to push a box b from location x to location y. Both Shakey and the box must be at the same location before this action can be used.

iii. ClimbUp(b,x), which allows Shakey to climb onto box b at location x. Both Shakey



and the box must be at the same location before this action can be used. Also Shakey must be on the Floor.

iv. ClimbDown(b,x), which allows Shakey to climb down from a box b at location x. Shakey must be on the box and the box must be in location x before this action can be used.

v. TurnOn(s,x), which allows Shakey to turn on switch s which is located at location x. Shakey must be on top of a box at the switch's location before this action can be used.

vi. TurnOff(s,x), which allows Shakey to turn off switch s which is located at location x. Shakey must be on top of a box at the switch's location before this action can be used.

(a) Using STRIPS syntax, define the six actions from above. In your action definitions, use only the following predicates: Box(b) to mean that box b is a box; In(x,r) to mean that location x is in room r; At(x,y) to mean that the object x is at location y; ShakeyOn(x) to mean that Shakey is on the object x, Switch(s) to mean that s is a switch; and SwitchOn(s) to mean that the switch s is on. You may also use the constants Shakey and Floor in the action definitions.

(b) In the above problem, using STRIPS, define the initial state depicted. Use only the predicates from part (a) and the constants Box1, Box2, Switch1, Switch2, Floor, Shakey, Room1, Room2, Corridor, L_{Door1}, L_{Door2}, $L_{ShakeyStart}$, $L_{Switch1}$, $L_{Box1Start}$, $L_{Box2Start}$, $L_{Switch2}$. The L_x constants are intended to represent the locations of x.

(c) Provide a totally ordered plan for Shakey to turn off Switch2 using the actions and the initial state defined in (a) and (b).

(6+2+2)

6. (a) Discuss one advantage and one disadvantage of using distance-weighted k-NN compared to standard k-NN. Discuss how using Cosine Similarity as a distance metric in k-NN could affect classification outcomes when applied to high-dimensional, sparse data.

(b) State the Naive Independence Assumption in Naive Bayes. Explain why this assumption might not hold in real-world datasets, with example. Discuss how the independence assumption affects the computational efficiency of Naive Bayes.

(c) The following dataset helps predict whether a company employee will work remotely based on their position, department, and commute distance.

POSITION	DEPARTMENT	COMMUTE DISTANCE	WORKS REMOTELY?
Engineer	IT	Long	Yes
Manager	HR	Short	No
Engineer	Finance	Medium	Yes
Staff	IT	Short	No
Manager	Finance	Long	Yes
Staff	HR	Medium	No
Engineer	Finance	Long	Yes
Staff	IT	Short	No

Using entropy to guide your splitting decisions, construct a decision tree that predicts if an employee will work remotely. Show all calculations and each step of the decision-making process.

(4+3+8)

7. (a) Explain the difference between TF-IDF and word embeddings in text representation, giving an example of each. Discuss the advantages of contextual embeddings over traditional word embeddings. Identify one limitation of contextual embeddings and suggest a possible improvement.

(b) Consider the story below:

Oliver had a busy day planned. He first stopped by his favorite café to grab a coffee. While waiting in line, he checked his watch and realized he was running late for his meeting at the office. He quickly paid for his coffee, took a sip, and headed out. As he reached the subway station, he found that the train was delayed due to maintenance. Frustrated, he called his assistant, Lisa, to inform her that he might be late and asked her to take notes in his absence. On the train, Oliver overheard a conversation about an upcoming art exhibition. Intrigued, he made a mental note to check it out later. When he finally arrived at the office, he thanked Lisa for covering for him and dove into his work.

Each inference below is based on a plausible interpretation of the story. For each inference, briefly explain whether that inference was primarily based on syntactic, semantic, pragmatic, discourse, or world knowledge. (Do not answer world knowledge unless none of the other categories are appropriate.) (2 marks for each inference)

- **Inference 1:** Oliver intended to get a coffee quickly because he was aware that he was running late for his meeting.
- **Inference 2:** When Oliver called Lisa to inform her that he would be late, he expected her to understand that she should take notes in his absence without him having to explain every detail.
- **Inference 3:** Oliver was likely interested in the art exhibition because he mentally noted to check it out later.
- **Inference 4:** The story implies that Oliver thanked Lisa specifically for taking notes and covering for him while he was late.
- **Inference 5:** Oliver's frustration at the delayed train suggests that he was relying on the subway to arrive at the office on time.

(5