COL333-COL671 Major Exam

Please write answers legibly using dark blue/black ink.

Please write your name, entry number and page number on all sheets. The response for each question must begin on a fresh page and that all sub-parts of a question must appear together.

The space requirement for any question is typically one page. Please do not exceed three pages for any question.

You may use a non-programmable calculator only for numerical calculations.

Please work individually and submit responses based on your own efforts.

Cases of copying in the answer scripts will be awarded zero points for this exam and Disciplinary guidelines will be followed.

Wishing you good luck for the exam.

Name and Entry No.: _____

Question	Points	Score
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
Total:	80	

1. (10 points) Consider a model, similar to MDPs, where instead of maximizing the expected discounted rewards, we wish to minimize the expected time to reach a specific state called the goal state. Work out the update equations for *value iteration* solution for this model using the quantities introduced below. Also, provide the equation to extract the optimal policy once the value iteration has converged. Clearly explain and show how you arrive at your solution. You may use the following quantities in your solution.

   - Let p_{ij}^a denote $p(s_t = j | s_{t-1} = i, a)$, which is the likelihood of arriving at state j if action a is executed at the current state i.
   - Let i_{goal} indicate the goal state.
   - Let $J^*(i)$ denote the expected time to the goal state starting from i if the agent follows the optimal policy.
   - Let $\pi^*(i)$ denote the optimal policy indicating the optimal action to take at state i.
   - Let $J^k(i)$ denote the value for state i on the k^{th} iteration .

   Finally, assume that each transition takes exactly one time step and the value function is initialized to 0.

Name and Entry No.: _____

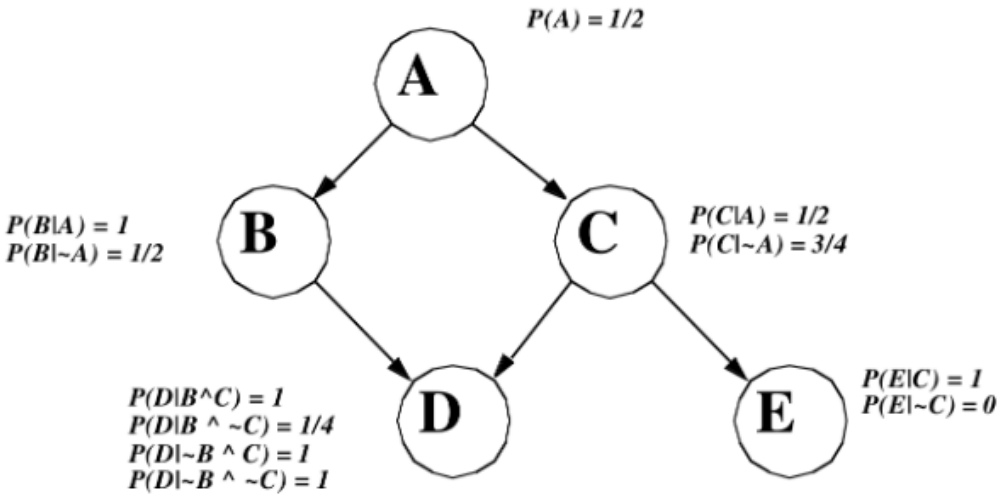
2. (10 points) A cleaning robot must vacuum a house on battery power. It can at anytime either *Clean* the house, *Wait* (without doing anything), or *Recharge* its battery. The robot measures its battery level as either *High* or *Low*. If the robot *Recharge*s its battery (whether in the *High* or *Low* battery state), the battery returns to the *High* state, and the robot receives a reward of (0). Executing the *Wait* action does not affect the battery state (the battery state remains unchanged) and the robot receives a reward of (+1). If the robot executes the *Clean* action, the outcome depends on the battery state. If the battery level is *High*, the battery drops to a *Low* level with a probability of (1/3) and with the probability of (2/3) the battery is still *High* after the robot cleans. However, if the battery level is *Low*, the battery can be exhausted with probability (1/2) in which case, the human must physically collect the robot and recharge it. In effect, the robot's battery ends up in the *High* state but receives a reward of (−10) due to the need for human intervention. With the remaining probability (1/2) the battery remains *Low*. In case the battery does not get exhausted during the *Clean* action, then the house gets cleaned and the robot receives a reward of (+3). The robot's rewards are based on the state-action-state triples in this problem.

   (a) Identify the states and actions for this problem. Draw the Markov process showing the states and transitions via actions and rewards from the problem statement above.

   (b) Assume you have an initial estimate of zero for each state's utility. What are the estimates after *one round* of value iteration given the problem description? You may assume a discount factor of $\gamma = 0.9$. Please show the working to arrive at your result. Provide the expression for estimating the value for a state before simplifying. Which state does the robot prefer to be in?

   (c) Now, consider that the robot does not have access to the transition model and the reward model. The robot interacts with the environment and observes the following states, actions and rewards:
   $\{High, Clean, (+3), High, Clean, (+3), Low, Clean, (+3), Low, Clean, (-10)\}$.
   Perform Q-learning using the sequence. Evaluate Q-values for each state-action pair after the robot receives each reward in the sequence. Use a learning rate of $\alpha = 0.2$ and a discount factor of $\gamma = 0.9$. Assume that all Q-value estimates are initialized to 0. Please show the working to arrive at your result. Provide the expression for estimating the Q-value before simplifying. You may show the estimation of only the non-zero Q-values at each stage. At the end of Q-learning, how many state-action pairs possess a non-zero Q-value? Further, identify the state in which the robot prefers the *Clean* action (over other action choices) and in which state it would not prefer taking the *Clean* action?

3. (10 points) Consider the Bayesian Network in the following figure. All variables are binary. The symbol $\sim$ indicates negation and $\wedge$ indicates a conjunction of variables. Justify your response in each case.



(a) What is $P(\sim A \wedge \sim B \wedge \sim C \wedge \sim D \wedge \sim E)$?

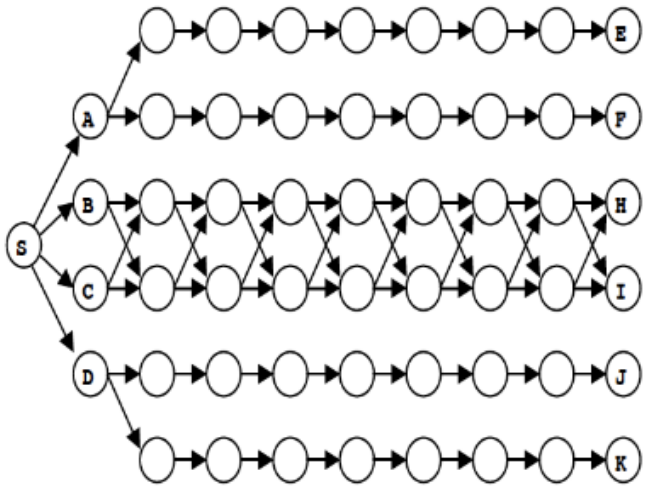
(b) What is $P(D| \sim B)$?

(c) What is $P(C|A \wedge B)$?

(d) What is $P(A|D \wedge E)$?

Name and Entry No.: _____

4. (10 points) Consider the following Bayesian Network.



In the diagram, some of the variables have been named as A, B, C, D, E, F, H, I, J, K and S. You may denote the remaining un-named variables as X, Y, Z etc. as required in your solution. For the following question, please provide a justification to arrive at your answer.

(a) Is B conditionally independent of C given $\{S, A, D\}$?

(b) Is B conditionally independent of C given $\{S, A, D, H\}$?

(c) Is B conditionally independent of C given $\{S, A, D, H, I\}$?

(d) Is B independent of C?

(e) Now assume that all the variables shown in the figure are binary. Further, each conditional probability table implied by every arc has each probability as 0.5. Call this Bayesian Network as BN_1. Can BN_1 be simplified by only removing some of the arcs such that the resulting Bayesian Network (say BN_2) represents the same distribution as BN_1? If yes, then draw the resulting bayesian network removing the maximum number of arcs.

5. (10 points) An autonomous vehicle is moving on a road. The vehicle is computer-controlled and must decide whether to *accelerate*(a) or *brake*(b), so as to drive collision-free on the road. The computer has access to the vehicle's state via two sensors: *speed*(s) which returns the integer-valued speed of the vehicle as $\{0, +1, +2, \dots\}$ and *distance*(d) that measures the distance to the nearest vehicle on the road and provides integer-valued output as $\{0, +1, +2, \dots\}$. Both sensors are assumed noise-free. Our goal is to perform (approximate) Q-learning in this setting. Assume the following integer-valued set of features $\{f_{ad}, f_{as}, f_{bd}, f_{bs}\}$ derived from the sensor values (s or d) paired with the action taken (a or b). For example, if the sensor reading is $\{(d = 1), (s = 2)\}$ and the action taken is a, then the feature values are $\{f_{ad} = +1, f_{as} = +2, f_{bd} = 0, f_{bs} = 0\}$.

   (a) The following episode is observed. Assume that the feature weights are initialized as $\{w_{ad} = +1, w_{as} = 0, w_{bd} = 0, w_{bs} = 0\}$. and the learning rate is 0.5. Compute the feature weights after each step. Show working to arrive at your answer.

Step	Initial sensor readings	Action	Reward	Final sensor readings
1	$\{(d = 0), (s = 2)\}$	a	-2	$\{(d = 1), (s = 0)\}$
2	$\{(d = 1), (s = 0)\}$	b	0	$\{(d = 1), (s = 0)\}$

   (b) Assume that the current sensor readings are $\{(d = 1), (s = 1)\}$. Given the learned weights, which action should be taken? Break ties alphabetically if needed. Show working to arrive at your answer.

Name and Entry No.: _____

6. (10 points) This problem concerns predicting the status S of a satellite orbiting the Earth. The probe has two states: *collecting data* C or *recharging* R. Each day, it sends a signal back to earth about its status. Due to interference caused by solar flares, the signal is occasionally corrupted, or even lost altogether. Hence, the signal G received by a ground station on earth can be of the following types: *collecting data* c, *recharging* r, and *no signal received* n. Certain conditional probabilities are known and appear below. Let the subscript t denote the discrete time instant. At time instant 0, there is no information about the state and each state is assumed to be equally likely.

S_t	C	R
$P(S_t \mid S_{t-1} = C)$	$\frac{3}{4}$	$\frac{1}{4}$
$P(S_t \mid S_{t-1} = R)$	$\frac{2}{3}$	$\frac{1}{3}$

G_t	c	r	n
$P(G_t \mid S_t = C)$	$\frac{3}{4}$	0	$\frac{1}{4}$
$P(G_t \mid S_t = R)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

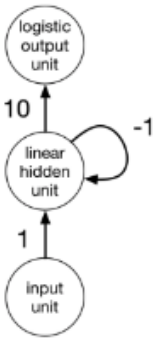
(a) The ground station receives the first observation as $G_0 = r$. What is the likelihood that the probe is collecting data? Please show your work to arrive at the solution.

(b) The ground station receives the second observation as $G_1 = n$. What is the most likely state of the probe after the first and the second observations. Please show your work to arrive at the solution.

7. (10 points) Consider the following short questions.

   (a) Consider the problem of training a neural network that takes one scalar input x as input and produces one scalar output y through the following non-linear function, $y = w_0 + w_1 sin(w_2 + w_3 x)$. Here, $\{w_0, w_1, w_2, w_3\}$ denote the weight parameters. The training data is of the form $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. We wish to minimize the sum of squared error $\sum_{i=1}^{n} \delta_i^2$ where $\delta_i = y_i - w_o - w_1 sin(w_2 + w_3 x_i)$. Draw the computation graph for this network showing the primitive operations involved. Let η denote the learning rate. Derive the four weight update equations for the parameters for gradient descent training.

   (b) Consider the recurrent network drawn below. All of the biases are assumed to be zero. The length of the input sequence can be assumed as even and each input is assumed to be integer-valued. Derive the function computed by the output unit at the final time step (the derivation of intermediate outputs can be ignored). Further, qualitatively explain what the network computes.



Name and Entry No.: _____

8. (10 points) Assume that it is your birthday today. Your hostel mate wants a treat and is asking you to buy her/him an ice cream. Both of you agree to the following strategy: you will select the place (IIT campus or Hauz Khas market), your hostel mate will select the shop. Ice creams at the shop can only be collected through a vending machine. But that machine is faulty and is handing out ice-creams uniformly at random, irrespective of customer's choice. Your hostel mate wants the most expensive outcome, but you want to spend the least given the agreed strategy. The table below lists the ice cream prices (in brackets). Prices that are not known (X and Y) are assumed to be non-negative.

Place	Shop Name	Ice Cream Flavours
IIT Campus	Mother Dairy	Mango (Rs. 50), Orange (Rs. 20)
	Café Coffee Day	Chocolate (Rs. 50), Vanilla (Rs. 30)
Hauz Khas	Giani's Shop	Strawberry (Rs. 25), Fruit (Rs. 35)
	Kulfi Shop	Kesar Kulfi (Rs. 36), Pista Kulfi (Rs. X), Badam Kulfi (Rs. Y)

(a) Draw the corresponding game tree and provide values for all nodes that do not depend on X and Y.

(b) What values of X will make you pick IIT for the treat (instead of Hauz Khas) regardless of the price of Y? Justify (1-2 lines).

(c) If the price of Badam Kulfi is at most Rs. 30, what values of X will result in an ice cream from Giani's shop regardless of the exact price of Badam Kulfi? Justify (1-2 lines).