CS 188 Summer 2022 Introduction to Artificial Intelligence — Written HW 4

Due: Monday 07/25/2022 at 11:59pm (submit via Gradescope).

Policy: Can be solved in groups (acknowledge collaborators) but must be written up individually

Submission: It is recommended that your submission be a PDF that matches this template. You may also fill out this template digitally (e.g. using a tablet). **However, if you do not use this template, you will still need to write down the below four fields on the first page of your submission.**

First name	
Last name	
SID	
Collaborators	

For staff use only:

Q1.	Quadcopter: Data Analyst	/31
Q2.	Quadcopter: Pilot	/27
Q3.	MangoBot Human Detector	/8
Q4.	Markov Decision Process	/20
	Total	/86

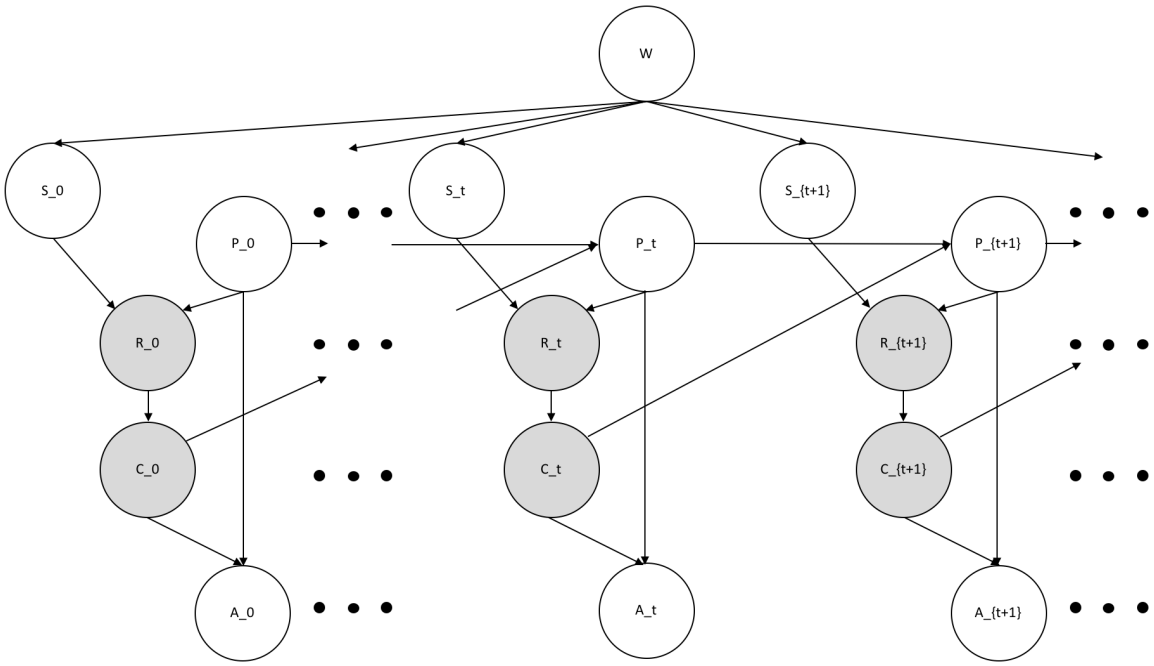
Q1. [31 pts] Quadcopter: Data Analyst

As in the last homework, we will consider a Bayes net for quadcopter flight. Our Bayes net has the following variables: W (weather), S (signal strength), P (true position), R (reading of the position), C (control from the pilot), and A (smart alarm to warn pilot if that control could cause a collision).

We will also consider the quadcopter flight over time. Here, flight can be considered in discrete time-steps: $t \in 0, 1, 2, ..., N-1$ with, for example, P_t representing the true position P at discrete time-step t. Suppose the weather (W) does not change throughout the quadcopter flight.

One key thing to note here is that there are edges going between time t and time $t+1$: The true position at time $t+1$ depends on the true position at time t as well as the control input from time t.

Let's look at this setup from the perspective of Diana, a data analyst who can only **observe** the output from a data-logger, which stores **R (reading of position) and C (control from the pilot)**.



(a) Hidden Markov Model

(i) [4 pts] List all the hidden variables and observed variables in this setup. In a few sentences, how is this setup different from the vanilla Hidden Markov Model you saw in lecture? You should identify at least 2 major differences.

Hidden variables:

Observed variables:

Differences:

(ii) [3 pts] As a data analyst, Diana's responsibility is to infer the true positions of the quadcopter throughout its flight. In other words, she wants to find a list of true positions $p_0, p_1, p_2, ..., p_{N-1}$ that are the most likely to have happened, given the recorded readings $r_0, r_1, r_2, ..., r_{N-1}$ and controls $c_0, c_1, c_2, ..., c_{N-1}$.

Write down the probability that Diana tries to maximize in terms of a **joint probability**, and interpret the meaning of that probability. Note that the objective that you write below is such that Diana is solving the following problem: $\max_{p_0,p_1,...,p_N}$ (maximization objective).

Maximization objective:

Explanation:

(iii) [3 pts] Morris, a colleague of Diana's, points out that maximizing the joint probability is the same as maximizing a **conditional probability** where all evidence ($r_0, r_1, r_2, \dots$ and $c_0, c_1, c_2, \dots$) are moved to the right of the conditional bar. Is Morris right?

◯ Yes, and I will provide a proof/explanation below.

◯ No, and I will provide a counter example below.

(b) The Markov Property

(i) [5 pts] In this setup, conditioned on all observed evidence, does the sequence $S_0, S_2, ..., S_{N-2}$ follow the Markov property? Please justify your answer.

(c) Forward Algorithm Proxy

Conner, a colleague of Diana's, would like to use this model (with the R_t and C_t observations) to perform something analogous to the forward algorithm for HMMs to infer the true positions. Let's analyze below the effects that certain decisions can have on the outcome of running the forward algorithm.

Note that when we say to **not include** some variable in the algorithm, we mean that we marginalize/sum out that variable. For example, if we do not want to include W in the algorithm, then we replace $P(S_t|W)$ everywhere with $P(S_t)$, where $P(S_t) = \sum_W P(S_t|W)P(W)$.

(i) [4 pts] He argues that since W (weather) does not depend on time, and is not something he is directly interested in, he does not need to include it in the forward algorithm. What effect does not including W in the forward algorithm have on (a) the accuracy of the resulting belief state calculations, and on (b) the efficiency of calculations? Please justify your answer.

Accuracy:

Efficiency:

(ii) [3 pts] He also argues that he does not need to include hidden state A (smart alarm warning) in the forward algorithm. What effect does not including A in the forward algorithm have on (a) the accuracy of the resulting belief state calculations, and on (b) the efficiency of calculations? Please justify your answer.

Accuracy:

Efficiency:

(iii) [3 pts] Last but not least, Conner recalls that for the forward algorithm, one should calculate the belief at time-step t by conditioning on evidence up to $t-1$, instead of conditioning on evidence from the entire trajectory (up to $N-1$). Let's assume that some other algorithm allows us to use evidence from the full trajectory ($t = 0$ to $t = N-1$) in order to infer each belief state. What is an example of a situation (in this setup, with the quadcopter variables) that illustrates that incorporating evidence from the full trajectory can result in better belief states than incorporating evidence only from the prior steps?

◯ If the signal strength is bad before $t-1$, but gets better later.
◯ If the signal strength is good up to $t-1$, and the signal is lost later.
◯ There isn't such example because using evidence up to $t-1$ gives us the optimal belief.

(d) Policy Reconstruction

Emily, another colleague of Diana's, would like to use this model to reconstruct the pilot's policy from data. Let's analyze below the effects that certain decisions can have on the outcome of doing policy reconstruction.

(i) [2 pts] Emily states that the probabilistic model for the pilot's **policy** is entirely captured in one Conditional Probability Table from the Bayes Net Representation. Which table do you think this is, and explain why this table captures the pilot's policy.

Table:

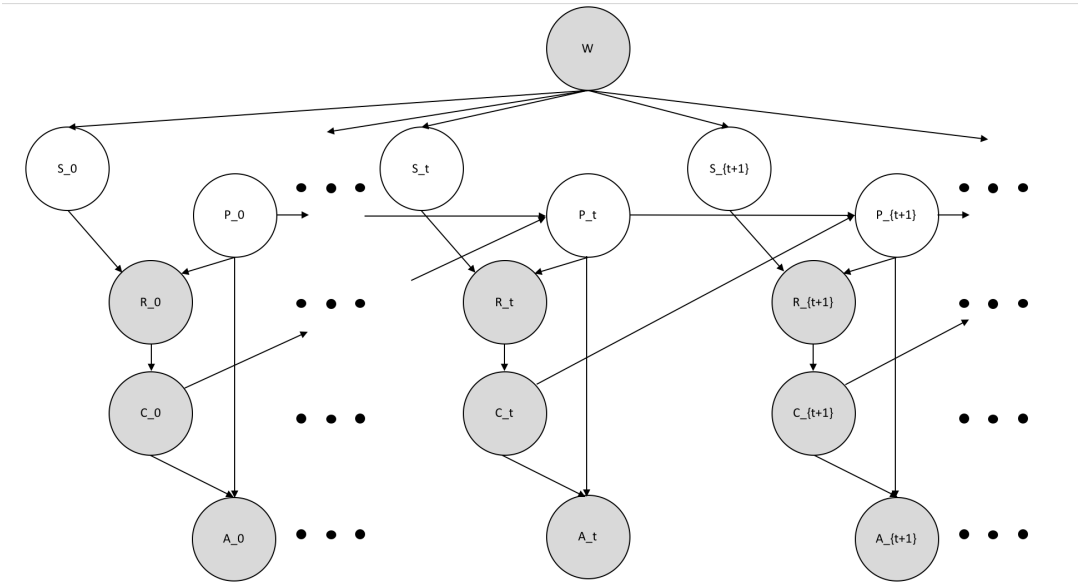
Explanation:

(ii) [4 pts] Emily argues that if we were given a lot of data from the data logger, we could reconstruct the probabilistic model for the pilot's policy. Is she right?

- ◯ Yes, and I will provide an overview of how to reconstruct the pilot's policy from the data.
- ◯ No, and I will provide a list of reasons for why we cannot reconstruct the policy.

Q2. [27 pts] Quadcopter: Pilot

In this question, we look at the same setup as the previous homework, but we now look at it from the perspective of Paul, a quadcopter pilot who can **observe W (weather), R (reading of position), C (control from the pilot), and A (smart alarm warning)**. As before, suppose weather (W) does not change throughout the quadcopter's flight.



(a) Forward Algorithm: The real deal

(i) [2 pts] Now that the only hidden states are S_t and P_t, is this graph a well-behaving HMM (where $E_{t+1} \perp\!\!\!\perp E_t \mid X_{t+1}$ and $X_{t+1} \perp\!\!\!\perp E_t \mid X_t$, recall that X is the hidden variable and E is the evidence variable)? Please explain your reasoning.

(ii) [4 pts] What is the time-elapsed **prediction** formula from time-step t to time-step $t+1$? Be sure to include all hidden states and observed states, and show how to assemble the formula from the conditional probability tables corresponding to the graph. Denote $B(S_t, P_t) = P(S_t, P_t \mid W, R_{0:t}, C_{0:t}, A_{0:t})$. Find $B'(S_{t+1}, P_{t+1}) = P(S_{t+1}, P_{t+1} \mid W, R_{0:t}, C_{0:t}, A_{0:t})$. Hint: refer to the "Filtering algorithm" lecture slides.

- ◯ $B'(S_{t+1}, P_{t+1}) = \max_{s_t} \sum_{p_t} P(S_{t+1} \mid W) * P(P_{t+1} \mid p_t, c_t) * B(S_t, P_t)$
- ◯ $B'(S_{t+1}, P_{t+1}) = \sum_{s_t} \sum_{p_t} P(S_{t+1} \mid S_t) * P(P_{t+1} \mid p_t) * B(S_t, P_t)$
- ◯ $B'(S_{t+1}, P_{t+1}) = \sum_{s_t} \sum_{p_t} P(R_{t+1} | S_{t+1}, P_{t+1}) P(S_{t+1} \mid S_t) * P(P_{t+1} \mid p_t) * B(S_t, P_t)$
- ◯ $B'(S_{t+1}, P_{t+1}) = \sum_{s_t} \sum_{p_t} P(S_{t+1} \mid S_t) * P(P_{t+1} \mid p_t) P(P_{t+1} \mid c_t) * B(S_t, P_t)$
- ◯ $B'(S_{t+1}, P_{t+1}) = \sum_{s_t} \sum_{p_t} P(S_{t+1} \mid W) * P(P_{t+1} \mid p_t, c_t) * B(S_t, P_t)$
- ◯ $B'(S_{t+1}, P_{t+1}) = \sum_{s_t} \max_{p_t} P(S_{t+1} \mid W) * P(P_{t+1} \mid p_t, c_t) * B(S_t, P_t)$

(iii) [4 pts] How do we include the observation **update** at time-step $t+1$? Be sure to include all hidden states and observed states, and show how to assemble the update from the conditional probability tables corresponding to the graph. Denote $B'(S_{t+1}, P_{t+1}) = P(S_{t+1}, P_{t+1} \mid W, R_{0:t}, C_{0:t}, A_{0:t})$, find $B(S_{t+1}, P_{t+1})$

- ◯ $B(S_{t+1}, P_{t+1}) = P(R_{t+1} \mid S_{t+1}, P_{t+1}) * P(C_{t+1} \mid R_{t+1}) * P(A_{t+1} \mid C_{t+1}) * B'(S_{t+1}, P_{t+1})$
- ◯ $B(S_{t+1}, P_{t+1}) = P(R_{t+1} \mid S_{t+1}, P_{t+1}) * P(C_{t+1} \mid R_{t+1}) * B'(S_{t+1}, P_{t+1})$
- ◯ $B(S_{t+1}, P_{t+1}) = P(R_{t+1} \mid S_{t+1}, P_{t+1}) * P(C_{t+1} \mid R_{t+1}) * P(A_{t+1} \mid C_{t+1}, P_{t+1}) * B'(S_{t+1}, P_{t+1})$
- ◯ $B(S_{t+1}, P_{t+1}) \propto P(R_{t+1} \mid S_{t+1}, P_{t+1}) * P(C_{t+1} \mid R_{t+1}) * P(A_{t+1} \mid C_{t+1}, P_{t+1}) * B'(S_{t+1}, P_{t+1})$
- ◯ $B(S_{t+1}, P_{t+1}) \propto P(R_{t+1} \mid S_{t+1}, P_{t+1}) * P(C_{t+1} \mid R_{t+1}) * B'(S_{t+1}, P_{t+1})$

(b) Consider a simpler scenario where we only track the 2D position (x, y) of the quadcopter. Paul, the pilot, wants to infer the quadcopter's true position P as accurately as possible.

- x, y **each** can take on values $\in \{0, 1, 2\}$.
- We have four controls: forward, backward, left, and right.
- Let variable E_R be Paul's estimate of the current position, and this variable depends on the reading R. The utility is based on the difference between the estimate of current position E_R and the actual position P: $U(P, E_R) = -||P - E_R||_x - ||P - E_R||_y$, in dollars.
- We consider only one time step. In that time step the **reading R is** $(1, 0)$ and that the weather is cloudy.
- Under cloudy weather, the signal strength can take on 2 values with equal probability: weak and strong. The signal strengths correspond to the following errors in readings:
  - Weak: The reading R returns a random number (for each position element) sampled uniformly from the domain of possible positions.
  - Strong: The reading R is identical to the true position.

Answer the following questions:

(i) [2 pts] Among the hidden variables S and P, Which variable should intuitively have the greatest VPI? Explain your answer. You should not do any calculations for this part.

Paul's coworker offers to tell him the signal strength (S) in exchange for some cash.

(ii) [3 pts] Suppose the signal strength is strong. Given the current reading R, what is the Maximum Expected Utility after knowing this information of S?

.

(iii) [3 pts] Suppose the signal strength is weak. Given the current reading R, what is the Maximum Expected Utility after knowing this information of S?

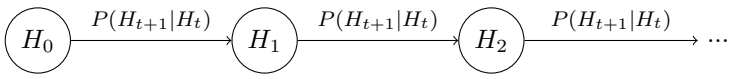
(iv) [3 pts] Considering the possibility of both signal strength, how much should Paul should pay to know this information of S?

(c) **(i)** [3 pts] Suppose your coworker only tells you the signal strength with probability q, and with probability $1 - q$, they don't tell you the signal strength even after payment. How much would you be willing to pay in this scenario? Your result should contain q.

(ii) [3 pts] How much would you pay to know the true position (P)?

Q3. [8 pts] MangoBot Human Detector

Your startup company MangoBot wants to build robots that delivers packages on the road. One core module of the robot's software is to detect whether a human is standing in front of it. We model the presence of humans with a Markov model:



where $H_t \in \{0, 1\}$ corresponds to a human being absent or present respectively. The initial distribution and the transition probabilities are given as follows:

H_0	$P(H_0)$
0	p
1	$1-p$

H_t	H_{t+1}	$P(H_{t+1}\|H_t)$
0	0	0.9
0	1	0.1
1	0	0.8
1	1	0.2

(a) Express the following quantities in terms of p:

(i) [1 pt] $P(H_1 = 1) =$

(ii) [1 pt] $\lim_{t\to\infty} P(H_t = 0) =$

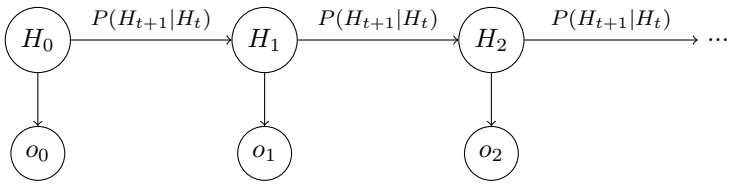
(b) The first-order Markov assumption in the model above can be inaccurate in real-world situations. Some potential ways to improve the model are listed below. For each option, determine whether it is possible to rewrite the process as a first-order Markov process, potentially with a different state representation.

(i) [1 pt] H_t depends not only on H_{t-1} but also on H_{t-2}. ◯ Yes ◯ No

(ii) [1 pt] H_t depends not only on H_{t-1} but also on $H_{t-2}, H_{t-3}, \cdots, H_{t-k}$ for some fixed $k \geq 3$. ◯ Yes ◯ No

(iii) [1 pt] H_t depends not only on H_{t-1} but also on $H_{t-2}, H_{t-3}, \cdots, H_1, H_0$. ◯ Yes ◯ No

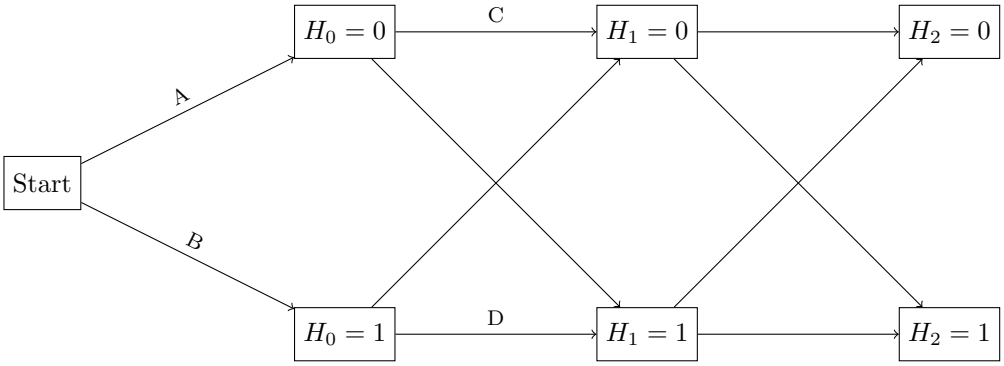
To make things simple, we stick to the original first-order Markov chain formulation. To make the detection more accurate, the company built a sensor that returns an observation O_t each time step as a noisy measurement of the unknown H_t. The new model is illustrated in the figure, and the relationship between H_t and O_t is provided in the table below.



H_t	O_t	$P(O_t\|H_t)$
0	0	0.8
0	1	0.2
1	0	0.3
1	1	0.7

(c) Based on the observed sensor values $o_0, o_1, \cdots, o_t$, we now want the robot to find the most likely sequence $H_0, H_1, \cdots, H_t$ indicating the presence/absence of a human up to the current time.

(i) [1 pt] Suppose that $[o_0, o_1, o_2] = [0, 1, 1]$ are observed. The following "trellis diagram" shows the possible state transitions. Fill in the values for the arcs labeled A, B, C, and D with the product of the transition probability and the observation likelihood for the destination state. The values may depend on p.



(ii) [1 pt] There are two possible most likely state sequences, depending on the value of p. Complete the following (Write the sequence as "x,y,z" (without quotes), where x, y, z are either 0 or 1):
Hint: it might be helpful to complete the labelling of the trellis diagram above.

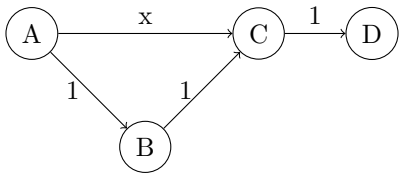
- When $p <$ _____, the most likely sequence H_0, H_1, H_2 is _____.
- Otherwise, the most likely sequence H_0, H_1, H_2 is _____.

(d) [1 pt] True or False: For a fixed p value and observations $\{o_0, o_1, o_2\}$ in general, H_1^*, the most likely value for H_1, is always the same as the value of H_1 in the most likely sequence H_0, H_1, H_2. ◯ True ◯ False

Q4. [20 pts] Markov Decision Process

Throughout this homework, we use $V(s)$ to denote the value of a state. This is the same as $U(s)$ used in lecture to denote the utility of a state. "Value" and "utility" mean the same thing in a Markov decision process.

(a) [5 pts] Consider the following deterministic MDP with four states A, B, C and D:



The edges designate actions between states, the weights on those edges are the rewards, and the discount factor is $\gamma = 1$. Let k be the **first** iteration of Value Iteration at which the value function converges for some x for a particular state (i.e. $V_k(s) = V^*(s)$). Use the convention from lecture where $V_0(s)$ is the value at initialization, $V_1(s)$ is the value after one iteration, etc. For each state A, B, C, and D, list **all** possible values of k. In the case a value function for a particular state never converges, set $k = \infty$ for that state.

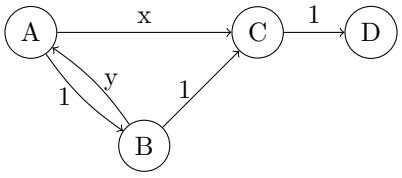
(a) State A, $k =$ ☐

(b) State B, $k =$ ☐

(c) State C, $k =$ ☐

(d) State D, $k =$ ☐

(b) Now consider the following deterministic MDP with four states A, B, C and D:



The edges designate actions between states, the weights on those edges are the rewards, and the discount factor is again $\gamma = 1$. Furthermore assume that $x, y \geq 0$.

(i) [5 pts] Let k be the **first** iteration of Value Iteration for some nonnegative x and y at which the value function converges for a particular state ($V_k(s) = V^*(s)$). For each state A, B, C and D list **all** possible values of k. In case a value for a particular state never converges set $k = \infty$ for that state.

(a) State A, $k =$ ☐

(b) State B, $k =$ ☐

(c) State C, $k =$ ☐

(d) State D, $k =$ ☐

(ii) [6 pts] Suppose we perform Policy Iteration and that k is the **first** iteration for which the policy is optimal for a particular state (i.e. $\pi_k(s) = \pi^*(s)$). On top of $x, y \geq 0$ also assume that $x + y < 1$ and that tie-breaking during policy improvement is alphabetical. The initial policy is given in the table below.

State s	Policy $\pi_0(s)$
A	C
B	C
C	D
D	D

For each state A, B, C and D, find k; if the policy never converges set $k = \infty$ for that state.

(a) State A, $k =$ ☐

(b) State B, $k =$ ☐

(c) State C, $k =$ ☐

(d) State D, $k =$ ☐

Th following two questions are conceptual.

(c) [2 pts] Which of the following statements are guaranteed to be correct for any MDP? Select all that apply.

- ☐ There exists a state s and some policy π such that $V^\pi(s) \leq V^*(s)$.
- ☐ There does not exist a state s such that for all policies π, $V^\pi(s) \leq V^*(s)$.
- ☐ For all states s and for all policies π, $V^\pi(s) \leq V^*(s)$.
- ◯ None of the above.

(d) [2 pts] Which of the following statements are guaranteed to be correct for Value Iteration? Select all that apply.

- ☐ At each iteration, and for all states, the value at the next iteration is $\geq$ the value at the current iteration.
- ☐ At each iteration, and for all states, the value at the next iteration is $>$ the value at the current iteration.
- ☐ At each iteration, the value function can be lower than the earlier values for some state.
- ☐ Once the value function is optimal at all states, value iteration will not change any value at any state.
- ◯ None of the above.