HOMEWORK 3
GRAPHICAL MODELS AND REINFORCEMENT LEARNING

[1] CMU 10-701: INTRODUCTION TO MACHINE LEARNING (FALL 2023)

https://machinelearningcmu.github.io/F23-10701/

OUT: Wednesday, Oct 4th, 2023

DUE: Wednesday, Oct 11th, 2023, 11:59pm

Instructions

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 2.1"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only. See the Academic Integrity Section in our course syllabus for more information: https://machinelearningcmu.github.io/F23-10701/#Syllabus.
- **Late Submission Policy:** See the late submission policy here: https://machinelearningcmu.github.io/F23-10701/#Syllabus.
- **Submitting your work:**
  - **Gradescope:** There will be one submission slot for this homework on Gradescope. For the written problems such as short answer, multiple choice, derivations, proofs, or plots, we will be using the written submission slot. Please use the provided template. The best way to format your homework is by using the Latex template released in the handout and writing your solutions in Latex. However submissions can be handwritten onto the template, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks. Each derivation/proof should be completed in the boxes provided below the question, **you should not move or change the sizes of these boxes** as Gradescope is expecting your solved homework PDF to match the template on Gradescope. If you find you need more space than the box provides you should consider cutting your solution down to its relevant parts, if you see no way to do this, please add an additional page at the end of the homework and guide us there with a 'See page xx for the rest of the solution'.
    Regrade requests can be made after the homework grades are released, however this gives the TA the opportunity to regrade your entire paper, meaning if additional mistakes are found then points will be deducted.

For multiple choice or select all that apply questions, shade in the box or circle in the template document corresponding to the correct answer(s) for each of the questions. For LaTeXusers, use ■ and ●for shaded boxes and circles, and don't change anything else. If an answer box is included for showing work, **you must show your work**!

[1]Compiled on Thursday 5th October, 2023 at 01:49

1 Graphical Models: Representation [14 points]

Refer to Figure 1 for Questions 1-6. Assume all variables (A, B, C, D, E) are **binary**.

$$\begin{tikzcd} A \arrow[r] & B \arrow[r] & C \\ D \arrow[ur] \arrow[r] & E & \end{tikzcd}$$

Figure 1: A Bayesian network.

1. **[2 points] True** or **False**: $E \perp C | B$. Please explain your answer.

2. **[2 points] True** or **False**: $A \perp E | C$. Please explain your answer.

3. **[2 points]** Write down the general form of the factorization of the joint probability $P(A, B, C, D, E)$, making **no** conditional independence assumptions.

4. **[1 point]** How many parameters do we need if we represent the distribution as in Question 3?
Note: since all variables are **binary**, we need only one parameter to represent both $P(A = 1)$ and $P(A = 0)$.

5. **[2 points]** Write down the factorization of the joint probability $P(A, B, C, D, E)$ assuming it follows the above graphical structure from Figure 1.

6. **[1 point]** How many parameters do we need if we represent the distribution as in Question 5?

7. **[4 points]** Given the following joint probability distribution, draw the corresponding graphical model representation:

$$P(x_1, x_2, x_3, x_4, x_5, x_6, x_7) =$$

$$P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2)P(x_5|x_3, x_2)P(x_6|x_3, x_5)P(x_7|x_4, x_6)$$

2 The Backdoor Adjustment [12 Points]

When estimating the causal effect of one variable (say X) on another (say Y), it's easy to get confused by the presence of confounding variables. With confounding at play, the conditional risk difference (the difference between $P(Y|X=1)$ and $P(Y|X=1)$) may not be the same as the causal risk difference, i.e., the average treatment effect (the difference between $P(Y|do(X=1))$ and $P(Y|(do(X=0)))$. When randomized controlled trials are off the table, we can still sometimes (given an adequate set of structural assumptions about how our variables are related) estimate treatment effects from observational data alone. One such case is when we observe a sufficient set of variables to block all backdoor paths (paths involving confounders) that connect X and Y. In such cases, we can estimate the average treatment effect by performing the backdoor adjustment. Recall that for two variables a and b, a set C satisfies the **backdoor criterion** if:

- Neither a nor b is a parent of any element in C
- C blocks every path from a to b that includes an arrow leading into a.

Consider the simple DAG shown in Figure 2, where X is some treatment, Y is some outcome, and Z is some confounding variable. Let's say that X, Y and Z are distributed as follows:

$P(Z=1)$	0.5

$P(X=1|Z=1)$	0.3
$P(X=1|Z=0)$	0.7

$P(Y=1|Z=1, X=1)$	0.9
$P(Y=1|Z=1, X=0)$	0.8
$P(Y=1|Z=0, X=1)$	0.5
$P(Y=1|Z=0, X=0)$	0.4

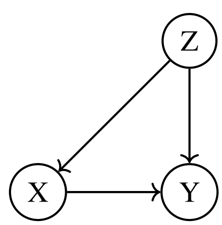


Figure 2

1. [**4 points**] For the graph in Figure 2, compute the conditional risk difference $P(Y=1|X=1) - P(Y=1|X=0)$:

2. [**4 points**] Now apply the backdoor correction to calculate $P(Y = 1|\text{do}(X = 1)) - P(Y = 1|\text{do}(X = 0))$:

3. [**2 points**] Explain the difference in the results from Questions 1 and 2.

4. [**2 points**] For the graph in Figure 3, give a minimal set S that satisfies the backdoor criterion for identifying the effect from A to C.

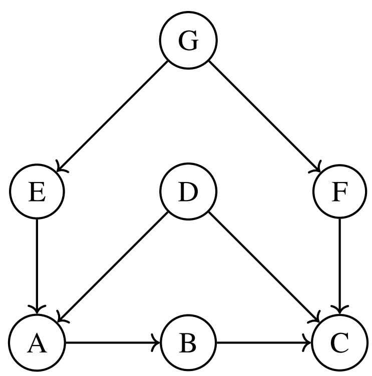


Figure 3

3 Graphical Models: Inference [16 points]

1. **[2 points]** Consider the graphical model below:

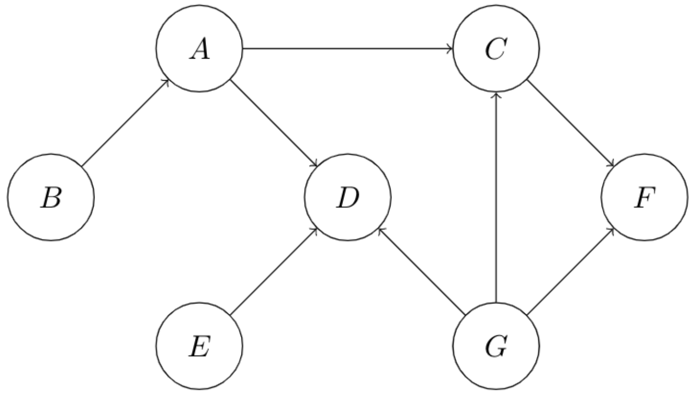
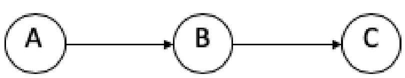


Figure 4: A Network

What variables are in the Markov blanket of G?

For the remainder of this section, consider the graphical model below, consisting of three binary random variables, and the associated training dataset:



5 training examples

A	B	C
1	1	0
1	0	1
1	1	1
0	1	0
1	1	1

Figure 5: A Network, Plus training data.

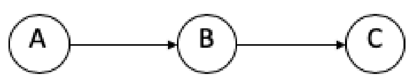
2. **[1 point]** How many parameters are needed to define this Bayes Net?

3. **[4 points]** Given the provided training dataset, what are the maximum likelihood estimates for each of the parameters of this Bayes Net?

4. **[2 points]** What probability will your trained network assign to $P(B = 1|A = 1, C = 1)$?

5. **[2 points]** Is your Bayes Net estimate of $P(B = 1|A = 1, C = 1)$ the same or different from what you would obtain if you calculated the MLE of $P(B = 1|A = 1, C = 1)$ directly from the training data? Briefly explain why or why not.

For the remainder of this section, consider using EM to train the same Bayes Net, using the previous dataset with an additional partially observed training example. **Note training example 6 has a missing value.**:



6 training examples

A	B	C
1	1	0
1	0	1
1	1	1
0	1	0
1	1	1
1	?	1

Figure 6: A Network, Plus partially observed training data.

6. **[1 point]** Consider using EM to train this network with this data. Furthermore, assume the network parameters are initialized not with random values, but instead using the network parameters you calculated in Question 3 above. During the first E step, precisely what quantities must be calculated? Please give specific probability term(s).

7. **[2 points]** What will be the value of this quantity/these quantities calculated during the E-step?

   The details are shown in 3.4.

8. **[2 points]** After completing the E-step, the M-step will re-estimate the entire set of network parameters. What will be the updated estimate for the parameter $P(B = 1|A = 1)$? Give your answer rounded to 2 decimal places.

4 Reinforcement Learning [14 points]

Consider an environment in which our agent requires caffeine to function (if it helps, you can think of the agent as a graduate student). Because caffeine is so important to our agent, we would like the agent to find a policy that will always take the shortest path to coffee.

In order to apply reinforcement learning techniques such as value iteration and policy iteration, we first need to model this scenario as a Markov decision process (MDP). Note that there may be many different ways to define each of the MDP components for a given problem. Recall that an MDP is defined as a tuple (S, A, P, R, γ) where:

S is the (finite) set of all possible states.

A is the (finite) set of all possible actions.

P is the transition function $P : S \times S \times A \to [0, 1]$, which maps (s', s, a) to $P(s'|s, a)$, i.e., the probability of transitioning to state $s' \in S$ when taking action $a \in A$ in state $s \in S$. Note that $\sum_{s' \in S} P(s'|s, a) = 1$ for all $s \in S, a \in A$.

R is the reward function $R : S \times A \times S \to \mathbb{R}$, which maps (s, a, s') to the real-value reward $R(s, a, s')$, i.e., the reward obtained when taking action $a \in A$ in state $s \in S$ and arriving at state $s' \in S$.

γ is the discount factor, which controls the relative importance of short-term and long-term rewards. We generally have $\gamma \in [0, 1)$ (i.e., $0 \leq \gamma < 1$), where smaller values mean greater discounting of future rewards.

For this problem, we represent the state of the agent at any time by the triple (x, y, o), where x and y are the horizontal and vertical coordinates of the agent's location, respectively, and o is the agent's orientation which is one of $\{N, E, W, S\}$ (North, East, West, South). The agent's current orientation is the only direction it can move. The agent is able to move forward, turn right, or turn left (deterministically). All actions are available in all states. More specifically, the action space is $|A| = \{F, R, L\}$ where:

- F (move Forward): The agent moves forward one cell along its current orientation, changing the agent's location but not its orientation.
- R (turn Right): The agent turns right, changing the agent's orientation but not its location.
- L (turn Left): The agent turns left, changing the agent's orientation but not its location.

Walls are represented by thick black lines. The agent cannot move through walls. If the agent attempts to move through a wall, it will remain in the same state.

When the agent reaches the coffee cup in any orientation, the episode ends. Another way to think of this is that every action in the coffee cup state keeps the agent in the coffee cup state.

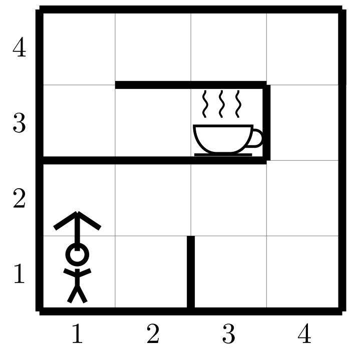


Figure 7: MDP for Problem 1, Part I. The goal is for the agent to have a policy that always leads it on the shortest path to the coffee at coordinate $(3, 3)$. In this instance, the agent's current state is $(1, 1, N)$

Consider the instance shown in Figure 7. The goal, displayed as a coffee cup, is located at $(3, 3)$. Using the above problem description answer the following questions:

1. [**1 point**] i.) How many states are in this MDP?

   [**1 point**] ii.) How many actions are in this MDP?

   [**1 point**] iii.) What is the total number of unique transition tuples?

2. [**2 points**] Fill in the following probabilities for the transition function P.

		s'			
s	a	(1,2,N)	(1,1,S)	(1,4,N)	(1,4,S)
(1,1,S)	F				
(1,1,N)	F				
(1,4,E)	R				

3. [**2 points**] i.) Propose a suitable reward function $R : S \times A \to \mathbb{R}$ for this MDP.

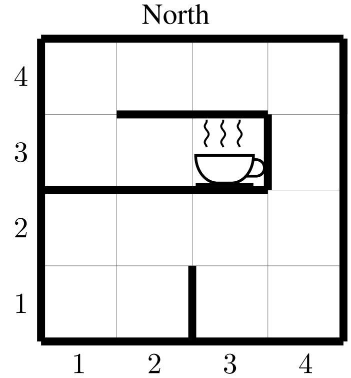
[**2 points**] ii.) Does the value of $\gamma \in (0, 1)$ affect the optimal policy in this setting? Briefly justify your answer.

4. [**1 point**] How many possible deterministic policies are there, including both optimal and non-optimal policies?

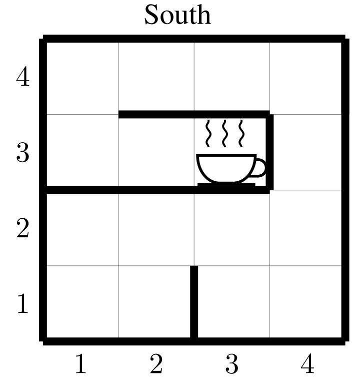
5. [**4 points**] Show one optimal deterministic policy for this MDP by filling in the 4 grids below (one for each orientation). You should label each cell with one of {F,R,L}, for Forward, turn Right and turn Left, respectively.

   (If there are multiple optimal policies, then give just one. The action chosen for the "coffee" state, $(3, 3, o)$ for $o \in \{N, E, W, S\}$, is arbitrary, so you can just keep the cup symbol there.)

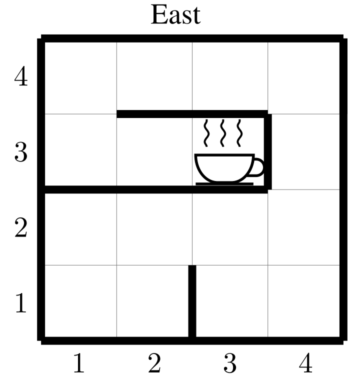
North



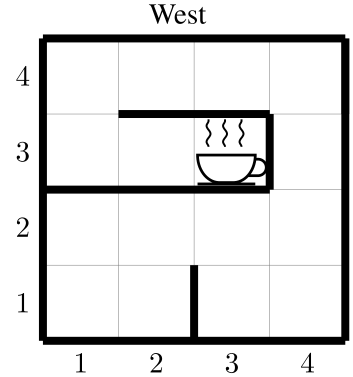
South



East



West



5 Collaboration Questions

1. (a) Did you receive any help whatsoever from anyone in solving this assignment?

   (b) If you answered 'yes', give full details (e.g. "Jane Doe explained to me what is asked in Question 3.4")

2. (a) Did you give any help whatsoever to anyone in solving this assignment?

   (b) If you answered 'yes', give full details (e.g. "I pointed Joe Smith to section 2.3 since he didn't know how to proceed with Question 2")

3. (a) Did you find or come across code that implements any part of this assignment?

   (b) If you answered 'yes', give full details (book & page, URL & location within the page, etc.).