CSE150B Spring 2024 Final Exam

Please, finish reading everything on this page before you start.

- Open book, web, and everything. No chatgpt. Do all the work **completely by yourself** without any discussion with others. Any form of communication related to the course before the submission deadline will be considered academic integrity violation.
- For clarification questions, ask me via **direct messages** on Slack (@Sean) and do **not** post anything in public channels. If any general correction/clarification is needed, I'll make announcements in the #general channel on Slack.
- Be assured that I will read every message, and if you do not see my reply within an hour or so (when my Slack status is green), it means my reply is "no comment" because the answer will be clear after re-reading the questions or going through class materials.
- Submit the follow

  1. Submit phot signment: "Final").
  2. Submit a vid at the following link: `https://www`

- **Make sure to le** ork is spotty. You need to show your face urn off camera and only show the answer s
- There is no strict ake it too long (20 min should be more th as for each question you answered. Think aminer about what you understand, while
- The grading will b sure academic integrity and also help get tly asks you to explain something, you ne rwise that question will be considered una
- Make sure the file I don't know if Dropbox sends you confirm of the page indicating that you have successfully uploaded your files. You can upload multiple times, but we do not guarantee that we will only check the last one. So try your best to send only a final version of the video.
- Zoom is highly recommended for doing your recording. If you recorded using your phone or in some other ways, try compressing the video into mp4 format (can use relatively low resolution) and avoid making the file too large. If you find yourself uploading gigabytes of video, you should change the file.
- Your overall letter grade for the course will only depend on the sum of all raw points you obtained in the class, and there is no notion of "percentage" in this class. You only need to get enough points for the threshold of the grade you want.

1 Classical Search

Question 1 (7 Points)**.** The undirected graph in Figure 1-(Left) has four nodes (A,B,C,D). The number on each edge indicates the cost of going through the edge – e.g., going from A to B has a cost of 5, same as going from B to A. Set node A as the start, and D the goal node. Answer the following questions and explain your reasoning.

1. In the correct algorithm for Dijkstra, if we visit a node that is already in the frontier, we need to compare the cost of this node from the current path with its cost in the frontier (the last two lines in the pseudocode of Slide 36 of this chapter 1 Classical Search.pdf). Now, on this given graph, if we do not do that comparison and replacement (i.e., remove the last line in the pseudocode), then what is the path from A to D that this wrong modification of Dijkstra algorithm will return? Explain the steps of how you obtained it.

2. What is the path that will be returned by the correct Dijkstra algorithm instead?

3. In your PA1, we agreed that it was ok to not do this frontier check and replacement, and still always get the optimal pa xplain why it was ok in the graphs in PA1 nation, you can use the fact that the Dijk sing path cost.

4. Define the followi $1, h(C) = 7, h(D) = 0$. Is h a consistent h

5. What is the path euristic h?

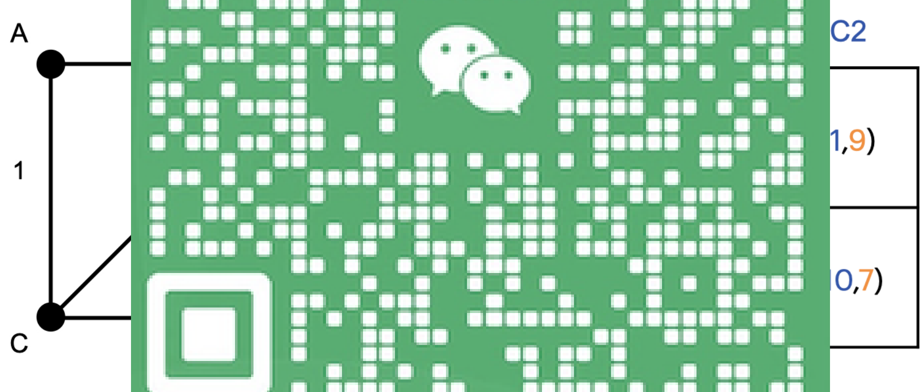


Figure 1: (Left) Graph for the classical search question. (Right) Game for adversarial search question.

2 Adversarial Search

Question 2 (3 Points)**.** Consider the game in Figure 1-(Right). There are two players: the column player and the row player. The column player goes first, and chooses either the first column (C1) or the second (C2), and after that, the row player chooses either the first row (R1) or the second (R2). After these two steps, the game ends and the pair of two numbers located by these two choices is the outcome of the game. In each pair of numbers, the first number is the reward that the column player will receive, and the second number is what the row player will receive. For instance, if the column player chooses C1, and then the row player chooses R1, then the outcome is (10,2), which means the column player will receive a reward of 10 and the row player a reward of 2.

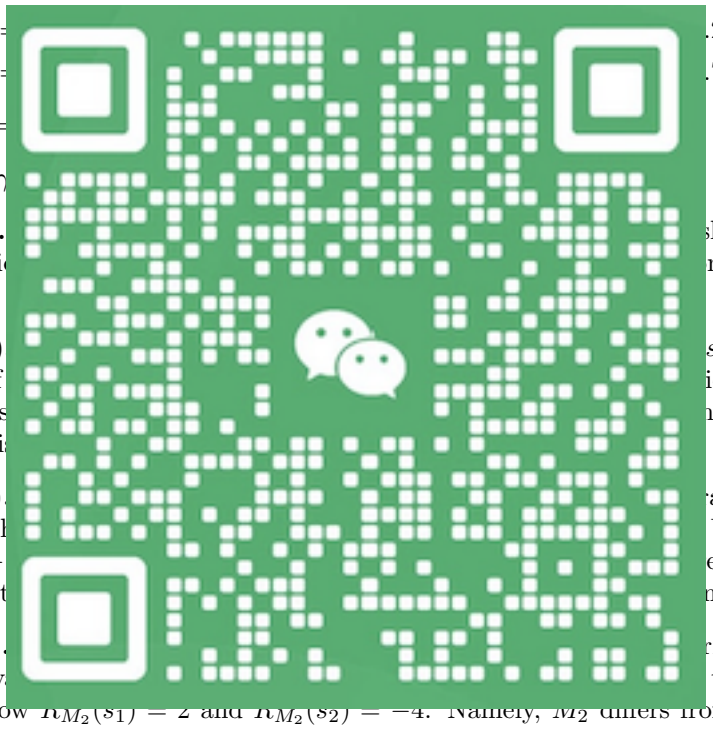
Assume that both players are rational players and want to maximize their own reward (no minimizers), what should each of their action be, and what is the final outcome of the game under those actions?

In your answer, draw the game tree that is similar to the minimax tree, just that now the two players are no longer the max and min players (just annotate their nodes as column and row players). Explain your reasoning on the tree. I hope you realize this is how the algorithms we talked about in minimax can be applied to games where players are not directly adversarial to each other (not zero-sum).

3 Markov Decision Processes and Reinforcement Learning

Consider the following MDP:

- State space: $S = \{s_1, s_2\}$
- Actions: $\mathrm{A}(s_1) = \{a_1, a_2\}$ and $\mathrm{A}(s_2) = \{a_1, a_2\}$.
- Transition model:
  - $P(s_1|s_1, a_1) =$ .2
  - $P(s_1|s_2, a_1) =$ .7
- Rewards: $R(s_1) =$
- Discount factor: γ



Question 3 (2 Points). show the states, actions, Q-states (i.e., state-acti n the appropriate edges in the graph.

Question 4 (3 Points) $s_1) = V_0(s_2) = 0$. That is, perform two steps of ides). Note that each V is a vector of the values ow how to compute V_1 and V_2, where each V_i i

Question 5 (2 Points). ration to converge to be within 0.1 error from th $V^* = (V^*(s_1), V^*(s_2))$, and we want $|V_k(s_1) -$ e does k need to be to converge to under this t mentation involved.

Question 6 (2 Points). rite $V^*_{M_1}(s_1), V^*_{M_1}(s_2)$ to represent the optimal v the same as M_1 except that the rewards are now $R_{M_2}(s_1) = 2$ and $R_{M_2}(s_2) = -4$. Namely, M_2 differs from M_1 only in that the reward on each state is twice as large. We write the optimal values of M_2 as $V^*_{M_2}(s_1)$ and $V^*_{M_2}(s_2)$.

Now, suppose you know what $V^*_{M_1}(s_1)$ and $V^*_{M_1}(s_2)$ are (which you don't, and you do not need to implement it to find out), then what should $V^*_{M_2}(s_1)$ and $V^*_{M_2}(s_2)$ be? That is, write down the values of $V^*_{M_2}(s_1)$ and $V^*_{M_2}(s_2)$ using the values of $V^*_{M_1}(s_1)$ and $V^*_{M_1}(s_2)$. Explain your reasoning.

Question 7 (4 Points). We now perform Q-learning in the MDP defined in the beginning, i.e., M_1, without accessing the transition probabilities. We initialize all Q-values to be zero $Q(s_i, a_i) = 0$. For simplicity we will use a fixed learning rate parameter $\alpha = 0.1$ in all temporal-difference updates (so it does not change with the number of visits). We start from the initial state s_2, and suppose we perform/observe the following three steps in the learning process in the MDP:

- (Step 1) At the initial state s_2, we take the action a_1, and then observe that we are transitioned back to s_2 by the MDP.

- (Step 2) We then take the action a_1 again from s_2, and observe that we are transitioned now to s_1 by the MDP.
- (Step 3) Now at the state s_1, we take the action of a_1 yet again, and then observe that we are transitioned to state s_2 by the MDP.

After each step of these three steps, we perform temporal-difference updates for the Q-values on the relevant state-action pairs. Now answer the following questions.

1. What values do we have for $Q(s_1, a_1)$ and $Q(s_2, a_1)$ now, after these three steps of updates? Write down how you obtained them.
2. Suppose from here we will use the ε-greedy strategy with $\varepsilon = 0.3$, which means that with ε probability we will use an arbitrary action (each of the two actions will be chosen equally likely in this case), and with $1 - \varepsilon$ probability we will choose the best action according to the current Q-values. Now that we are in s_2 after Step 3, what is the probability of seeing the transition (s_2, a_1, s_1) in the next step? That is, calculate the probability of the event "according to the ε-greedy policy, we obtained the action a_1 in the current state s_2, and after applying this action, the MDP puts us in s_1 as the next state."
3. If instead of ε-gre action that maximizes Q-values in each s next step?

4 Monte Carl

Question 8 (2 Points) d coins as shown in the slides for this chapter. uniformly randomly at each round is the same ng. You should use the definition of regret over

Question 9 (Extra 2 P and it can be confusing if you are not familiar w empt it.

In MCTS we use the hat balance exploration and exploitation. We ex two-coin betting game, the UCB strategy achie can draw i.i.d. samples for each coin, because t ne.

Give a concrete exa setting. That is, design a minimax tree, such th de change over time, as the MCTS algorithm ex de is what we are using roll-outs from that nod

The mathematical c aining what the MCTS algorithm does on the tree that you design, so that it is clear that the win rate distribution for the options at the root node may shift over time.

5 Constraint Solving and Propositional Reasoning

Question 10 (3 Points)**.** Consider the following constraint system: The variables are $X = \{x_1, x_2\}$ with $x_1 \in D_1 = [0, 1]$ and $x_2 \in D_2 = [0, 1]$, both are continuous intervals of real numbers. The constraints are

$$C_1 : \ x_1 = x_2 + 0.2$$
$$C_2 : \ x_1 = x_2^2$$

Perform propagation on the domain $D = D_1 \times D_2$ using the constraints $\{C_1, C_2\}$ by enforcing arc-consistency. Show how the domains on each variable will be updated in the first several propagation steps, and then show what end results you will obtain by iterating such propagation, and explain why.

Question 11 (2 Points)**.** Put the following formula into CNF:

$$\varphi : \neg\Big((p_2 \to (p_3 \to \neg p_1)) \to p_1\Big)$$

You only need to expand the definition of the logical connective "$\to$" and then using De Morgan laws should be enough (i.e., no need to go through the general procedures of introducing new variables, etc.).