Midterm

Artificial Intelligence
Fall 2021 CS47100-AI

Student name: ____________________ Student PUID: ________________
Note: You are free to use your intuition to find the steps in the proof. But, make sure you do not use your intuition to justify steps in your proofs. Total: **85 points**.

---

1. Solve the following problems on Probability.
   (a) Select all expressions that are equivalent to $P(A,B|C)$, given no independence assumptions: [10]
   A. $\frac{P(C|A)P(A|B)P(B)}{P(C)}$ B. $\frac{P(B,C|A)P(A)}{P(B,C)}$ C. $\frac{P(A|C)P(C|B)P(B)}{P(B,C)}$ D. $\frac{P(A|C)P(B,C)}{P(C)}$
   E. $\frac{P(C|A,B)P(B|A)P(A)}{P(B|C)P(C)}$ F. $P(A|B,C)P(B|C)$ G. None of the Above
   (b) Select all expressions that are equivalent to $P(A|B,C)$, given no independence assumptions: [10]
   A. $\frac{P(C|A)P(A|B)P(B)}{P(C)}$ B. $\frac{P(B,C|A)P(A)}{P(B,C)}$ C. $\frac{P(A|C)P(C|B)P(B)}{P(B,C)}$ D. $\frac{P(A|C)P(B,C)}{P(C)}$
   E. $\frac{P(C|A,B)P(B|A)P(A)}{P(B|C)P(C)}$ F. $P(A|B,C)P(B|C)$ G. None of the Above

2. A fair six sided dice is rolled repeatedly and you observe outcomes sequentially. Formally, dice roll outcomes are independently and uniformly sampled from the set $\{1,2,3,4,5,6\}$. At every time step before the h^{th} roll you can choose between two actions:

   - Stop: stop and receive a reward equal to the number shown on the dice or,
   - Roll: roll again and receive no immediate reward.

   If not having stopped before then, at time step h (which would be reached after $h-1$ rolls) you are forced to take the action Stop, you receive the corresponding reward and the game ends.

   We will model the game as a finite horizon MDP with six states and two actions. The state at time step k corresponds to the number shown on the dice at the k^{th} roll. Assume that the discount factor, γ, is 1.

   (a) The value function at time step h, when it is no longer possible to roll the dice again, is $V^h(1) = 1$, $V^h(2) = 2$, $\ldots$, $V^h(6) = 6$. Compute the value function at time step $h-1$. [5]
   (b) Express the value function at time step $k-1$, with $2 < k \leq h$ recursively in terms of the value function at roll k, so in terms of $V^k(1)$, $V^k(2)$, $\ldots$, $V^k(6)$: [5]
   Solution:

   The Q function at time step k for action "Roll" does not depend on the state since the number shown by the dice is irrelevant once you decided to roll. We use the shorthand notation $q(k) = Q^k(\text{state,"Roll"})$ since the only dependence is on k.

   (c) Compute $q(h-1)$. [10]
   (d) Express $q(k-1)$ recursively as a function of $q(k)$, with $2 < k \leq h$. [5]
   (e) What is the optimal policy $\pi^k(s)$ at roll k as a decision rule based on the current state s and $q(k)$?(i.e. $\pi^k(s) =$ Roll if ________________, stop otherwise) [5]

3. Consider the grid-world given below and Pacman who is trying to learn the optimal policy. If an action results in landing into one of the shaded states the corresponding reward is awarded during that transition. All shaded states are terminal states, i.e., the MDP terminates once arrived in a shaded state. The other states have the North, East, South, West actions available, which deterministically move Pacman to the corresponding neighboring state (or have Pacman stay in place if the action tries to move out of the grad). Assume the discount factor $\gamma = 0.5$ and the Q-learning rate $\alpha = 0.5$ for all calculations. Pacman starts in state $(1, 3)$.

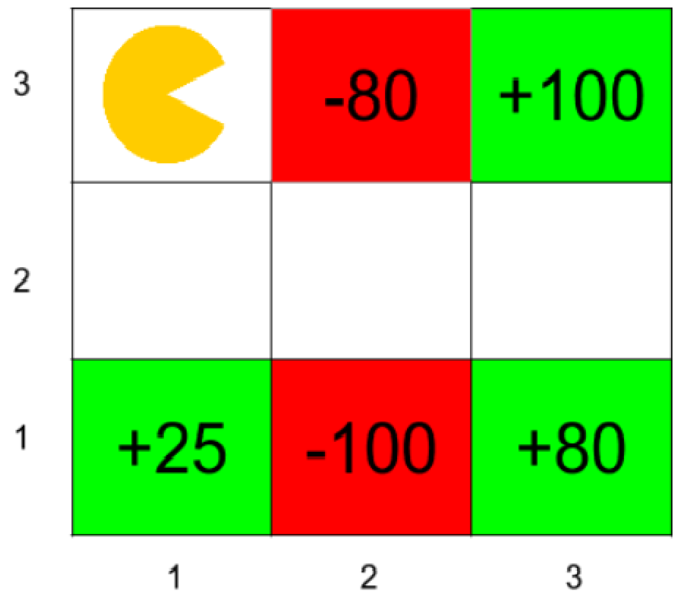

Figure 1: Grid world

(a) What is the value of the optimal value function V^* at the following states: $V^*(3,2)$, $V^*(2,2)$, $V^*(1,3)$ [5]

(b) The agent starts from the top left corner and you are given the following episodes from runs of the agent through this grid-world. Each line in an Episode is a tuple containing (s, a, s', r). [5]

Episode 1	Episode 2	Episode 3
(1,3), S, (1,2), 0	(1,3), S, (1,2), 0	(1,3), S, (1,2), 0
(1,2), E, (2,2), 0	(1,2), E, (2,2), 0	(1,2), E, (2,2), 0
(2,2), S, (2,1), -100	(2,2), E, (3,2), 0	(2,2), E, (3,2), 0
	(3,2), N, (3,3), +100	(3,2), S, (3,1), +80

Using Q-Learning updates, what are the following Q-values after the above three episodes: $Q((3,2), N)$, $Q((1,2), S)$, $Q((2,2), E)$.

(c) Consider a feature based representation of the Q-value function: [10]

$$Q_f(s,a) = w_1 f_1(s) + w_2 f_2(s) + w_3 f_3(a)$$

$f_1(s)$:The x coordinate of the state $\qquad$ $f_2(s)$:The y coordinate of the state

$$f_3(N) = 1,\ f_3(S) = 2,\ f_3(E) = 3,\ f_3(W) = 4$$

i. Given that all w_i are initially 0, what are their values after the first episode: w_1,w_2,w_3

ii. Assume the weight vector w is equal to $(1, 1, 1)$. What is the action prescribed by the Q-function in state $(2, 2)$?

4. Suppose you want to go from Lawson Building (LWSN) to Purdue Memorial Union (PMU) using the following graph to determine the route you can take.

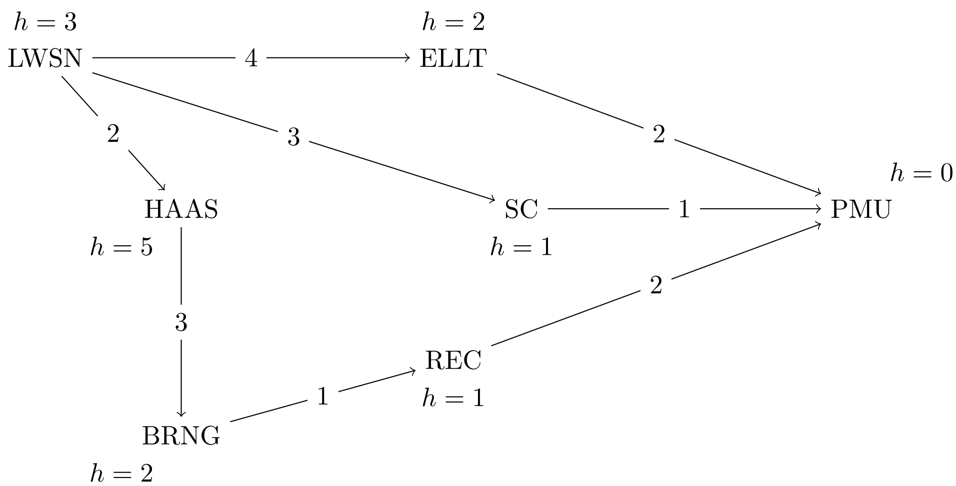

Figure 2: Graph for problem 1

(a) Apply each of the following algorithms to this problem and write down the order in which the nodes of the graph are visited. If when expanding a node there are more than one option **use alphabetical order for tie breaking**. If the algorithm uses a heuristic, use the h value above each node.

i. **Example:** Depth-First Search: LWSN ELLT PMU

ii. Breadth-First Search [3]

iii. Uniform Cost Search [3]

iv. A^* [3]

(b) Is the heuristic shown in the graph admissible? Is it consistent? Why or why not? [6]