AI 533 **Intelligent Agents and Decision Making** Spring 2023

OREGON STATE UNIVERSITY

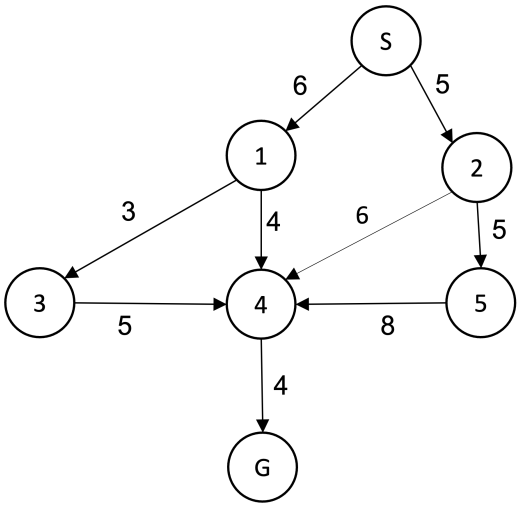
School of Electrical Engineering and Computer Science

Instructor: Sandhya Saisubramanian

Total points: 100 Assignment 1 Due date: April 25, 2023

Instructions: This homework assignment consists of a written portion and a programming portion. Collaboration is not allowed on any part of this assignment. Solutions must be typed (hand written and scanned submissions will not be accepted) and saved as a .pdf file. You will submit a single .zip file that contains the the code base and solutions as a .pdf file.

1. **(15 points)** Calculate the shortest path from S to G in Figure 1 using A^* algorithm with the heuristic values provided in the table. The g values are the edge costs in Figure 1. Write the shortest path, its solution cost, and the order in which the nodes are expanded.



Node	h value
S	7
1	6
2	5
3	4
4	2
5	3

Figure 1: Graph for informed search

2. **(15 points)** Consider a campus food delivery robot that is responsible for collecting orders from the a pickup station and delivering it to Kelley Engineering Center. Write an action schema in STRIPS for order pickup for the following scenarios. (i) The robot can take only one item in a trip. (ii) Suppose you want the robot to carry two items, one in each of its in-built compartment, how would you modify your answer to the previous question?

   Note: think about the conditions in the world that the robot has to meet in order to collect an item. Example: ensure the robot is in the right location for order pickup.

3. **(15 points)** Given an MDP $M = (S, A, T, R, \gamma)$ with a fixed state state s_0 and a fixed policy π, the probability that the action at time $t = 0$ is $a \in A$ is:

$$\Pr(A_0 = a) = \pi(s_0, a).$$

Similarly, the probability that the state at time $t = 1$ is $s \in S$ is:

$$\Pr(S_1 = s) = \sum_{a_0 \in A} \pi(s_0, a_0) T(s_0, a_0, s).$$

Write a similar expression (using only S, A, T, R, γ, π and Bayes' theorem) for the following:

(i) The expected reward at time $t = 6$ given that the action at time $t = 3$ is $a \in A$ and the state at time $t = 5$ is $s \in S$. Use $R(s, a)$ for reward notation.

(ii) The probability that the action at time $t = 16$ is $a' \in A$ given that the action at time $t = 15$ is $a \in A$ and the state at time $t = 14$ is $s \in S$.

4. **(5 points)**Write the Bellman equation for state value of a finite, discrete MDP with discount factor γ and reward function denoted by $R(s, a, s')$.

5. **(50 points)** Consider the gridworld in Figure 2, with two states covered in water (s9, s10 and two states with wildfire (s1, s2). The agent can move in all four directions. The agent succeeds with probability 0.8 and may slide to the neighboring cells with probability 0.1. Illustration of the transition probability for actions 'up' and 'right' are shown in Figure 2. The agent receives a reward of $+100$ when it reaches the goal state. The agent receives a reward of -5 in the water states, -10 in widlfire states, and a reward of -1 in all other states. The process terminates when the agent reaches the goal state. The agent's objective is to maximize the expected reward it can obtain.

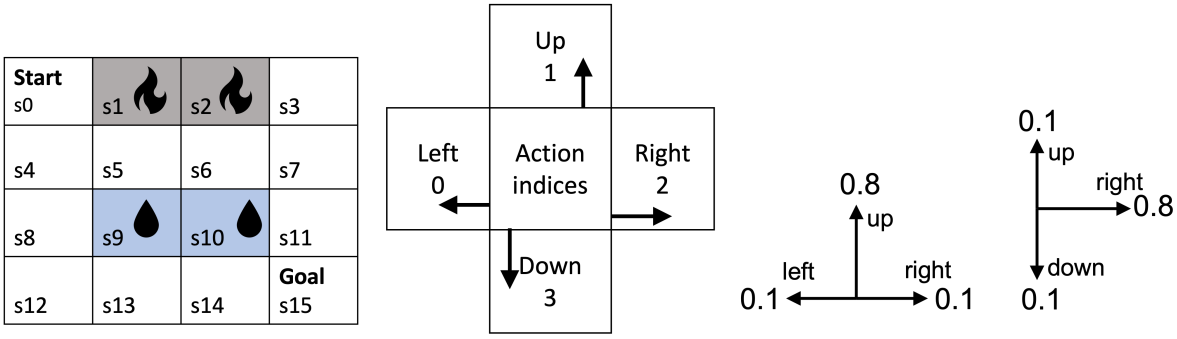


Figure 2:

- **(10 points)** Setup (code up) the environment for solving using value iteration and policy iteration.
- **(15 points)** Implement value iteration for this problem with $\gamma = 0.3$. Print the policy and the objective value. State clearly if your implementation will cause VI iteration to terminate when it reaches the goal state.
- **(10 points)** Implement value iteration for this problem with $\gamma = 0.95$. Print the policy and the objective value. How does the policy in states leading to wildfire and water change, as the discount factor is increased?
- **(15 points)** Implement policy iteration for this problem with $\gamma = 0.95$. Print the policy and the objective value. Is the policy and the objective value same as that of value iteration? State clearly if your implementation causes each iteration to terminate when it reaches the goal state.