

# Homework 5 Primer

Due Tuesday, April 12<sup>th</sup> at 09:00pm ET

*You are encouraged to discuss the assignment in general with your classmates, and may optionally collaborate with one other student. If you choose to do so, you must indicate with whom you worked. Multiple teams (or non-partnered students) submitting the same solutions will be considered plagiarism.*

The goal of this assignment is to deepen your understanding of some of the topics covered in the lectures and readings. We will grade your answers based on whether they demonstrate an understanding of the concepts in each question. We will award partial credit for answers that demonstrate partial understanding, so show your work!

## What to Submit

You should submit a file named `homework5primer.pdf`, containing your answers to the questions. You can record your answers on this document (preferred) or create your own.

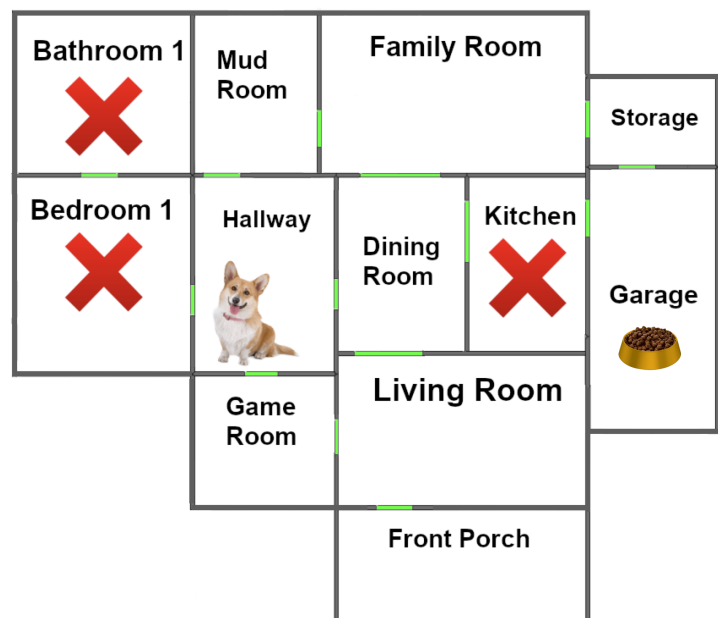
## 1. Training Dogs (MD<sub>og</sub>Ps) (25 points)

The diagram below depicts a map of our house. We'd like to train our dog, Jeff, to go to the garage to get his dinner without entering any of the prohibited rooms (marked with a red X).

Doors between rooms are colored green. When Jeff moves toward a wall with a door, he enters the next room; if the wall has no door, he stays in the current room.

Additionally, Jeff is easily scared by vacuums, so anytime he tries to go through a door, there is a 0.1 probability that he thinks he hears a vacuum and goes in the *opposite* direction (for example, when moving from the **Hallway** to the **Mud Room**, he may actually end up in the **Game Room**).

Jeff hates baths, so whenever he goes into a *prohibited room*, we give him a bath on the spot (without changing rooms), making him unhappy. Additionally, given that Jeff is a hungry dog, he incurs a penalty to his happiness every time he makes a move but does not get to his dinner



We decide to structure Jeff's training as an MDP, with Jeff as the agent, the house as the environment (each room represents a different state  $s$ ), and Jeff's happiness as the reward function. Let Jeff's starting position be the **front porch**, with the **garage** as the sole terminal state. Assume your MDP is undiscounted (that is,  $\gamma = 1$ ). Let the reward Jeff gets when entering a room be  $R(s) = R_{\text{room}}(s) + R_{\text{hunger}}(s)$ , i.e., the reward for room  $s$  is the reward for going into that room itself plus the reward for hunger in that room, with the following rewards:

$R_{\text{room}}(\text{prohibited room}) = -2$

$R_{\text{room}}(\text{allowed room}) = 0$

$R_{\text{room}}(\text{garage}) = 10$

$R_{\text{hunger}}(\text{any room}) = -2$


Using the information above, answer the following questions:

- a. Define a set of actions  $A$  that would allow Jeff to travel throughout the house. Give a brief qualitative description of the transition function  $P(s' | s, a)$  when  $s = \text{Dining Room}$  for each action  $a \in A$ .

**Actions ( $A$ ):**

**$P(s' | s = \text{dining room}, a)$  for each action  $a$  in Actions:**

- b. Imagine an optimal policy that sends Jeff to the **family room** when he is in the **mud room** or **dining room** on the way to the **garage**. Why might this same policy send him from the **hallway** to the **mud room** rather than to the **dining room**?

For questions 1c-1d let's say that we give Jeff a pair of earplugs () and he no longer fears the vacuum. That is, there is now a 0.0 probability that he thinks he hears a vacuum and goes in the opposite direction.

- c. In this revised scenario, is there more than one optimal policy that sends Jeff from the **front porch** to the **garage**?

**Answer** (select one):

**Explanation:**

- d. How can you change *either*  $R_{\text{hunger}}(\text{any room})$  or  $R_{\text{room}}(\text{prohibited room})$  such that there is only one optimal policy and it sends Jeff from the **front porch** up to the **family room** and

then through the **storage room** into the **garage**?

For the remaining question, let Jeff lose his earplugs and he returns to a 0.1 probability that he thinks he hears a vacuum and goes in the opposite direction.

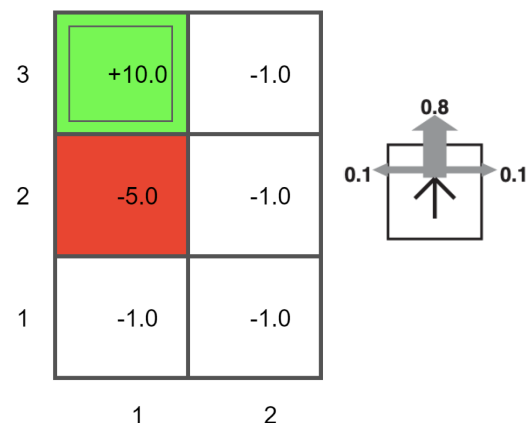
- e. Say Jeff turns into a cyborg and no longer feels hunger (no penalty for hunger) and wants to wander freely while avoiding prohibited rooms. What needs to be changed about  $R_{\text{room}}(s)$  and  $R_{\text{hunger}}(s)$  to let him wander instead of heading to the **garage** (no longer a terminal state)?



## 2. Value Iteration (25 points)

Consider the gridworld MDP shown below right. The terminal state (1,3) has a reward of +10 and the non-terminal state below it has a reward of -5. Rewards are -1 for all other states.

The agent makes its intended move (up, down, left, or right) with a probability 0.8, and moves in a perpendicular direction with probability 0.1 for each side (e.g., if intending to go right, the agent can move up or down with a probability of 0.1 each). If the agent runs into a wall, it stays in the same place.



Calculate the utilities of the following states for the next two iterations of the value iteration algorithm using a discount factor of  $\gamma = 0.9$ . Write your answer in the table below, where columns are states and rows are iterations. Note, the initial iteration is provided and the next iteration is partially provided. **Show your work.**

s	(1,1)	(2,1)	(1,2)	(2,2)	(2,3)
$U_0(s)$	-1	-1	-5	-1	-1
$U_1(s)$	-1.9	-1.9		-1.9	
$U_2(s)$					

**YOUR WORK BELOW**