



<u>Challenges</u>







### ◀ Back to RLIITM-1



:rowd 😂

## nent 1

# **preement Learning**

Leaderboard

## nanth



**Overview** 

<u>Notebooks</u>

1680

**Discussion** 

**2** 78

**\*** 0

<u>Insights</u>

**305** 

Resources

**Follow** 

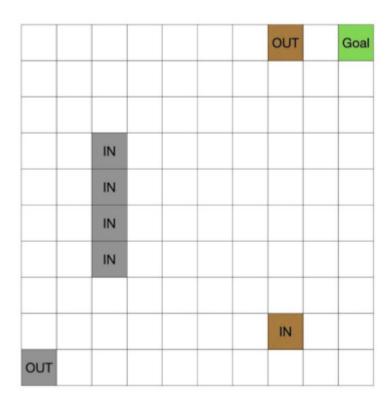
<u>Submissions</u>

**Create Submission** 

## Introduction

In this problem you will be implementing value iteration on a  $10 \times 10$  gridworld based on the actions, rewards and the state space. Consider a 10 × 10 gridworld (GridWorld-1) as shown in the Figure 1:

### FIGURE 1: GRIDWORLD-1



- State space: Gridworld has 100 distinct states. There is a special absorbing state called **Goal**. There are two wormholes labeled as **IN** in color Grey and Brown, any action taken in those states will teleport you to state labeled "OUT" in color Grey and Brown respectively. States labeled **OUT** are just a normal state.
- Action space: There are 4 actions  $A = \{North, East, South, West\}$ , which moves you one cell in the corresponding direction.
- Transition model: The Gridworld is stochastic. In this model, an action X  $\in$  {North, East, South, West} moves you one cell in the X direction of your current position with probability 0.8, but with probabilities 0.1 and 0.1 moves you one cell at angels of  $90^{\circ}$  and  $-90^{\circ}$  to the direction X,

#### **NOTEBOOKS**

Rules

See all

0

0 [Starter Notebook] RL -**Value Iteration** 

By ashivani 5 months ago

respectively (Refer Table 1 for more details). For example, if the selected action is North, it will transition you one cell to the North of your current position with probability 0.8, one cell to the East of your current position with probability 0.1 and one cell to the West of your current position with probability 0.1. Transitions that take you off the grid will not result in any state change. There are no transitions available once you reach the **Goal** state

#### TABLE 1

Action	Probabilities			
X	North	East	South	West
North	0.8	0.1	0.0	0.1
East	0.1	0.8	0.1	0.0
South	0.0	0.1	0.8	0.1
West	0.1	0.0	0.1	0.8

• Rewards: You will receive a reward of -1 for all transitions (including the one that take you off the grid) except the transition to the **Goal** state.

Any transition to the **Goal** state gives you a reward of +100.

#### **Instructions**

### 1.Implement:

1.1. value iteration. Find the pseudocode below. Let **S** be the state space and **A** the action space:

Initialize 
$$J_0(s) = 0, \forall s \in S$$
;  
For  $i = 0, 1, 2, ...$ 

• 
$$J_{i+1}(s) = \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} P_{ss'}(a) [r(s, a, s') + J_i(s')], \forall s \in \mathcal{S}.$$

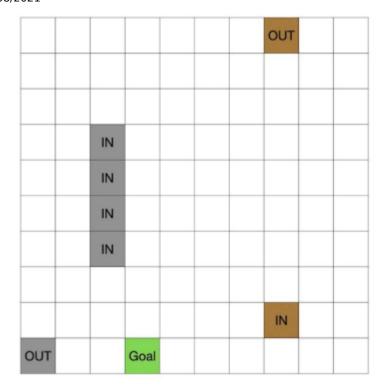
### 1.2. a greedy policy w.r.t $J_i$ as,

$$\pi_i(s) = \underset{a \in \mathcal{A}}{\operatorname{argmax}} \sum_{s' \in \mathcal{S}} P_{ss'}(a) [r(s, a, s') + J_i(s')], \forall s \in \mathcal{S}.$$

- 2. The value iteration loop goes to infinity (refer the pseudocode given above), so when would you stop your value iteration?
- 3. Plot a graph of max  $s \in S |J_{i+1}(s) J_i(s)|$  vs iterations.
- 4. Tabulate the values of J(s) and greedy policy  $\pi(s)$ ,  $\forall s \in S$ , after 10 iterations, 25 iterations, and after you stop the value iteration.
- 5. Consider a new gridworld (GridWorld-2) as shown Figure 2. Compare and contrast the

behavior of J and greedy policy  $\pi$  for GridWorld-1 and GridWorld-2.

FIGURE 2: GRIDWORLD-2



You will be writing your solutions & making a submission through a notebook. You can follow the instructions in the <u>starter code</u>.

## Files

Under the Resources section you will find data files that contains parameters for the environment for this problem.

## Submission

Submissions will be made through a notebook following the instructions in the starter code.

## Contact

• RL TAs

Alcrowd 😂	PLATFORM	CONNECT	LEGAL
	<u>Challenges</u>	Blog	<u>Cookie consent</u>
	<u>Organize a Challenge</u>	<u>Contact</u>	<u>Terms of Use</u>
	<u>FAQ</u>	<u>GitHub</u> ☑	<u>Participation Terms</u>
	<u>Our Team</u>	<u>GitLab</u> ☑	<u>Privacy</u>
	We are Hiring!	<u>Twitter</u> ☑	
		<u>Discourse</u> ☑	© 2021 Alcrowd. All rights reserved.
			reserveu.