Ayça Avcı

53961

COMP 341

Artificial Intelligence

Homework 5 Report

Question 1:

Reflex agent selects actions based on the agent's current perception of the world and not based on past perceptions. Our value iteration agent does not actually learn from experience. Rather, it thinks its MDP model to arrive at a complete policy before ever interacting with a real environment. When it does interact with the environment, it simply follows the precomputed policy, it becomes a reflex agent and our value iteration becomes “offline planning”.

Question 2:

I changed the noise parameter only. Since noise refers to how often an agent ends up in an unintended successor state when they perform an action, I reduced it to 1e-3 from 0.2. Since noise is too low compared to its initial value, our agent can arrive its terminal state.

Question 3:

a) Prefer the close exit (+1), risking the cliff (-10): I set discount to 0.1, noise to 1e-3 and reward to 0.3. Since noise is low enough, probability of going unintended states are low. Since discount is low enough, agent will prefer going from short path. Since reward is fair enough, it will not be stay in its current state.
b) Prefer the close exit (+1), but avoiding the cliff (-10): I set discount to 0.1, noise to 0.1 and reward to 0.3. I only increased the noise parameter compared to part a. Since noise is high enough and discount is also high enough, there is a risk that agent might go unintended state, but, it will prefer the short path and try to avoid cliff.
c) Prefer the distant exit (+1), risking the cliff (-10): I set discount to 0.9, noise to 1e-3 and reward to 0.3. Since discount high enough and noise is low enough, it will take the long path and since going to an unintended state probability is low because of low noise value, it will risk the cliff.
d) Prefer the distant exit (+1), avoiding the cliff (-10): I set discount to 0.9, noise to 0.1 and reward to 0.3. Since discount high enough and noise is high enough, it will take

the long path and since going to an unintended state probability is high because of high noise value, it will avoid the cliff.

e) Avoid both exits and the cliff (so an episode should never terminate): I set discount to 0.9, noise to 0.1 and reward to 100. Since reward of being its current state is too high, it will not take any action and stay its current state.

Question 4:

Q-learning agent learns by trial and error from interactions with the environment through its update (state, action, nextState, reward) method. In the other hand, in value iteration, agent thinks its MDP model to arrive at a complete policy before ever interacting with a real environment. When it does interact with the environment, it simply follows the precomputed policy. In Q-learning, ties between values are broken randomly for better behavior. In a particular state, actions that the agent has not seen before still have a Q-value, specifically a Q-value of zero, and if all the actions that the agent has seen before have a negative Q-value, an unseen action may be optimal.

Question 5:

When I run the "python gridworld.py -a q -k 50 -n 0 -g BridgeGrid -e 1" on the terminal, agent could not find the optimal policy. After changing epsilon value to 0 from 1, since exploration probability became 0, agent did not want to explore any states and just partially explored one state except its initial state. I tried different epsilons and learning rates, but none of them find the optimal policy after 50 iterations. So, I answered the question as 'NOT POSSIBLE'. Low number of iteration (in this case 50) might not be enough to find optimal policy might be the answer for this question.

Question 6:

In practice, if the state space is extremely large, it is impractical to tabularize Q-values of all state-action pairs. Hence, tabularized Q-learning does not work for larger grids. Additionally, 5000 iterations might be enough to explore small grids, but not the larger ones. Larger ones might require much more iterations to be explored.

Question 7:

If the ghost was so close (one step away from the agent), agent changed its direction and tried to move away from the ghost. Agent considers taking the closest food if there are no ghosts nearby. If there was a food and the ghost at the same state, agent did not go to that state.