Ex6 | (67842) AI To Introduction

8 באוגוסט 2024

Question 8

- The value iteration computational complexity depends on three parameters.:
  - i = the number of iterations.
  - s = the number of states.
  - a = the number of actions.
- In each iteration we compute all the values for this iteration for all the combinations between states and actions.
- Therefore, the computational complexity would be $O(i \cdot s \cdot a)$.

Question 9

- In question 3 there are 3 paraeters we changed:
  - The discount factor:
    - This factor is responsible of how much weight we give the future states.
    - This means that the state values are more spread out on the grid.
    - If we would like to take a risk free path, we would lower the discount in order to get the policy according to the immediate values, and not according to the future states.
    - For example, in 3(a) and 3(b) the only difference we made is the discount.
    - In (a) we want to take the riskier path, hence we chose a larger discount, and in (b) we chose a lower discount on order to take the less risky path.
  - The noise factor:
    - This factor represents the percent that the action given will not happen.
    - This means that if we want to take a risky path, the factor should be low, and higher when we want to take a risk free path.
    - This is in order that the agent will not accidently slip into the bad state.

    - For example, in (c), we want to take the risk of walking close to the clif. If we ignore the clif (set the factor to 0), the agent will choose this route.
    - When we dont want to take the risky route, we assign a number larger than 0 to this factor in order that the agent will know that there is a risk to slip.
  - The living reward factor:
    - This factor represents the reward the agent gets as long as he lives.
    - If we want the agent to live longer, the reward will be higher, otherwise lower.
    - For example, in (e) we want the agent to stay alive as long as possible. In order to do that we set the living reward to 20, which is higher than both of the terminal states, which means he wont have any desire to get to those states.
    - In the other cases we want the agent to find the fastest path. In order to do that we set the reward to be lower than the goal states in order to get those higher rewards as fast as possible.

Question 9

- We can select ations based on the ditribution probability from the Q values. We set a value to be the temperature. The higher this value is, the more we explore other actions.
- In order to choose action a in state s:

$$P\left(a|s\right) = \frac{e^{Q(s,a)/temperature}}{\sum_{action} e^{Q(s,action)/temperature}}$$

- This will choose a action according to the distribution of the Q values.