Complete individually or in pairs. If in pairs, only one person in the group needs to submit the answers, but make sure to include group members' names.

MDP: Mini-blackjack problem

In mini-blackjack, you repeatedly draw a card (with replacement) that is equally likely to be a 2, 3, or 4. You can either Draw or Stop if the total score of the cards you have drawn is less than 6. Otherwise, you must Stop. When you Stop, your utility is equal to your total score (up to 5), or zero if you get a total of 6 or higher. When you Draw, you receive no utility. There is no discount (gamma = 1).

**Q1. What are the states and the actions for this MDP?**

The state∈(0, 2, 3, 4, 5, done)

The action ∈(draw, done)

**Q2. What is the transition function and the reward function for this MDP?**

The transition function is:

T(s, stop, done) = 1

T(s, draw, s’) if s and s’ ∈(2 , 3, 4), T= 1/3,

If s = 2 and s’ = done, T= 1/3

If s =3 and s’ = done, T = 2/3

If s ∈(4, 5) and s’ = done, then T = 1

T(s, a, s’) = 0

The reward functions is:

R(s, stop, done) = s, s <= 5

R(s, a, s’) = 0

**Q3. Give the optimal policy for this MDP.**

By using value iteration:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **V** | **0** | **2** | **3** | **4** | **5** | **Done** |
| **V0** | **0** | **0** | **0** | **0** | **0** | **0** |
| **V1** | **0** | **2** | **3** | **4** | **5** | **0** |
| **V2** | **3** | **3** | **3** | **4** | **5** | **0** |
| **V3** | **3.3333** | **3** | **3** | **4** | **5** | **0** |

By using value iteration for each step, we could know that the optimal policy is draw if s <= 2, and if s is greater than 2, we stop.

**Q4. What is the smallest number of rounds (k) of value iteration for which this MDP will have its exact values (if value iteration will never converge exactly, state so).**

The smallest number of rounds(k) of value iteration for which this MDP will have its exact value is 3.