Q3: MDPs and the Bellman Equation

CS 695, Prof. Stein

Due: Wednesday October 18, 2022

1 Relationships between MDP quantities

In class, we discussed the Bellman equation. Here is one form of that equation:

$$V^*(s) = \max_a \mathbb{E}\left[R(s,a,s') + \gamma V^*(s')\right] \tag{1}$$

and, equivalently,

$$V^*(s) = \max_a \sum_{s' \in S} T(s,a,s') \left[R(s,a,s') + \gamma V^*(s')\right] \tag{2}$$

where $T(s,a,s')$ is the probability of ending up in state s' after executing action a from state s and R is the instantaneous reward (for the same states). Remember, the value function $V^*(s)$ tells you the (optimal) value of being in a certain state s. The state-action value function $Q^*(s,a)$ is the value of being in state s, taking action a and then behaving optimally thereafter.

In this question you will be asked to produce the relationship between a few different terms we defined in class. The answer to some of these questions can be found in the slides, while for others you may also want to consult Sutton & Barto. In your answers, you may use the reward $R(s,a,s')$ and the state transition model $T(s,a,s')$.

1. Define the optimal value function $V^*(s)$ in terms of the optimal state-action value function $Q^*(s,a)$.

2. Imagine you have access only to the optimal state-action value function $Q^*(s,a)$. Write the optimal policy $\pi^*(s)$ in terms of Q^*.

3. Now we'll go the other way: imagine you have V^*, but you want Q^*. Using the MDP model (including the reward and the state transition

model), write Q^* in terms of V^*. *Hint: look at the Bellman equation and think about what it represents and what each of these terms represent.*

4. Given the optimal value function $V^*(s)$, write a definition of $\pi^*(s)$ (without using Q^*). You may use your answer to part 3 in your response.

2 Describing an MDP for Pac-Man

Problem inspired by a similar assignment from Oregon State.

In this problem you will describe the various quantities that make up an MDP for a simplified version of the classic Pac-Man video game. (If you're unfamiliar with it, here's it's Wikipedia page.) In this problem, I ask you to describe Pac-Man as an MDP.

You do not need to write any equations for this problem, though if you find them helpful for your description, feel free to use equations and/or set theory notation to define key terms. Part of the state space will include the board itself, which is formally a grid with places that you are not allowed to go; it is sufficient to simply say so: a formal mathematical definition is unnecessary. I just want to know that you can think about how you might translate a more complex example into an MDP. For simplicity, you may assume that the moves executed by each ghost (enemy) at every step will be random.

1. Describe the state space of Pac-Man. Be sure to include *everything*, including the board, the pellets, the ghosts, etc.

2. What actions are available to the agent? (This defines your action space).

3. Describe the transition model. Is it determinstic (no randomness) or stochastic (with randomness)? *Note: it's not as simple as the motion of the agent! At each time step, multiple things can change; make sure you discuss what those are.*

4. Define the reward function. What are the inputs to the reward function (i.e., does it depend on the starting state? the action? the state you end up in)? *Note: you may define the instantaneous reward given for accomplishing various tasks, like eating pellets.*