Name: Akassh Shah

Course: Deep Learning for Games, Fall 2019

Date: 30th of September, 2019

Assignment 2: Reinforcement Q-Learning from Scratch in Python with OpenAI Gym

Assignment 2 is designed to reinforce reinforcement learning (RL) concepts, and introduce the OpenAI Gym framework. This assignment is based on the tutorial you can access at <https://www.learndatasci.com/tutorials/reinforcement-q-learning-scratch-python-openai-gym/>. It describes the Taxi-V2 environment, whose source code is available at <https://github.com/openai/gym/blob/master/gym/envs/toy_text/taxi.py>

For full credit, please read the questions below and answer them completely.

1. In the described reward function, for which action and state combination does the agent receive the highest positive reward?
   1. The agent will receive the most reward for a successful drop-off. Therefore, the action must be 6 (dropoff) and the state must be where the passenger wanted to be dropped off.
2. In the described reward function, for which actions and state combinations are the agent penalized for?
   1. The agent will receive the most punishment for dropping of the passenger at a wrong location. Therefore the action must be dropoff, and location will be not R, G, Y or B.
3. How many states are possible in the taxi problem? That is, what is the size of the state space?
   1. 500
4. How many actions are possible in the taxi problem? That is, what is the size of the action space?
   1. 6
5. In ascending order of their numerical value provided in the tutorial, list the word-based names of actions that can move a taxi from one square to another.
   1. 0: south
   2. 1: north
   3. 2: east
   4. 3: west
6. What are the other two actions in the action space that are not listed in the previous question? Name them and describe what they do.
   1. Pickup: picks up the passenger from that location
   2. Dropoff: drops off a passenger to that location
7. What is the name of the Gym environment that we are discussing in this tutorial?
   1. "Taxi-v2"
8. A lot of how we will work with the Gym environment is through the class that describes the environment. What is the name of the class method for Env that resets the environment? What does this method return?
   1. env.reset() -> this function resets the environment and returns a random initial state
9. The step function is also called on an Env object. It takes an action in the environment (i.e., a step in the environment) by executing the action, which likely changes the environment and results in a new observation (i.e., game state). What are the four return values of the step function?
   1. observation: Observations of the environment
   2. reward: Value that lets us know if the action was good or bad
   3. done: If that episode is over i.e. if we successfully picked up and dropped of a passenger
10. How does the tutorial penalize running into walls?
    1. -1 penalty
11. In this taxi environment, which function returns the size of an action space?
    1. env.action\_space
12. Which function returns the size of a state space?
    1. env.observation\_space
13. Which function can we call on env to draw a picture of a state to the screen?
    1. env.render()
14. Suppose we have a taxi at row 4, column 3, and a passenger in location 1 with a destination of location 2. Encode this state and print it to the screen. Which state number does this set of numbers represent?
    1. state = env.encode(4, 3, 1, 2)
15. Based on the example code in the tutorial, as opposed to the text in a sentence, how do we assign the state in in the env? Please write the full assignment operation provided to you in the code example.
    1. env.s = state # where state can be state = env.encode(4, 3, 1, 2)
16. When our Taxi-V2 environment is created, it also creates a reward table for us that is accessible through a dictionary with what name?
    1. env.P[i] # where i is the state for which you want the reward table
17. In the reward table dictionary, what types of values do the keys represent? What are the values stored in the dictionary?
    1. action: [(probability, nextstate, reward, done)]
18. How is an episode defined for this environment?
    1. An episode is considered complete if we successfully picked up and dropped of a passenger
19. Name the three ways that we are measuring our agent’s performance.
    1. Average rewards per move
    2. Average timesteps per episode
    3. Average penalties per episode
20. Pretend you have found a terrible agent via a poorly trained Q-table. Consider all three metrics from the previous question and describe the type of behavior you would expect from this agent.
    1. I would expect this agent to not pick up the passenger. This will cause less damage than picking up the passenger and riding around with said passenger.