Khoa Tran
02/19/2021
CSE 415

Part B Report

For part b, I implemented three functions: handle_transition, choose_next_action, and extract_policy. In the handle_transition, the user can control the agent through the problem state by selecting actions. In this method, I looped through all the possible actions and calculated and updated the max q value and the dictionary. In the choose_next_action function, epsilon is implemented as it determines exploration (choose random action from state) or exploitation (choose optimal action w highest the q value). The epsilon-greedy selects the action with the highest estimated reward most of the time. Having a fixed epsilon value closer to 1 will force the algorithm to take more random actions and not use past knowledge. However, with custom epsilon, the algorithm with exploit the current knowledge, making it more balance. I did not implement an exploration function.