Name: Berke Derin Berktay

Number: 72968

Question 1:

The agents that are trained in this homework are called “Reflex Agents” because these agents’ actions depend only on predetermined policies that depend on the current state. Or in other words, the action, or reflex of thewse agents are only dependent on such policies. The procedure of value iteration is called “offline planning” is because the iterated actions are offline meaning they are predetermined already based only on the currently implemented policy and nothing else. This is not true for reinforcement learning. Hence the training phase is very crucial for the case when we are doing offline planning.

Question 2:

For the programming Question 2, did you change the discount factor or the noise parameter and why? Discount 0.9 noise 0.2

For this specific question, I have changed the noise to 0 since changing this noise to 0 means that if the agent decides to go east for instance there is a zero percent chance of the agent accidentally going south. Therefore the agent does not consider when crossing the eastward path the possibility of ending up at -10 square since the possibility of the next state not being the desired one is zero. This is way more effective in this case than altering the discount since discount effects then decision regarding whether or not to wait around longer to end up in a terminal square, which would not be very effective when compared to the discount factor in this case.

Question 3:

1. Prefer the close exit (+1), risking the cliff (-10)
2. Prefer the close exit (+1), but avoiding the cliff (-10)
3. Prefer the distant exit (+10), risking the cliff (-10)
4. Prefer the distant exit (+10), avoiding the cliff (-10)
5. Avoid both exits and the cliff (so an episode should never terminate)

Parameters chosen:

3a. 1,0,-5

3b. 0.2,0.2,0

3c. 1,0,-1

3d. 1,0.3,-1

3e. 1,0,10

For a, I made the noise 0 in order to go theough the rishier bath towards east, the reasoning is explained in the previous question. By making the living reward -5, we forced it to choose the closer exit with the less reqrd of 1 over the more distant exit of -10 since for every state when we are in the game we lose the 5 points and therefore we by doing it like this we force it to terminate the game as soon as possible by taking the closest exit. For b, by making the noise nonzero we ensure it goes north first , so we take don’t take the riskier path and by making discoint 0.2, we ensure that the

farther reewards are worth way less over a short period of time and therefore we choose the closest of 1 over 10. I have explained these tradeoffs for all 3 of these parameters, and for the remaining subquestions I have chosen the numbers based on the logic behind these 3 parameter tradeoffs that I hav explained. However, the last case needs explaining since it is a bit of a different case. For this case the idea was to make the reward for living at each step so high that simply living's reward over time would be more than the reward at even the best terminating state.

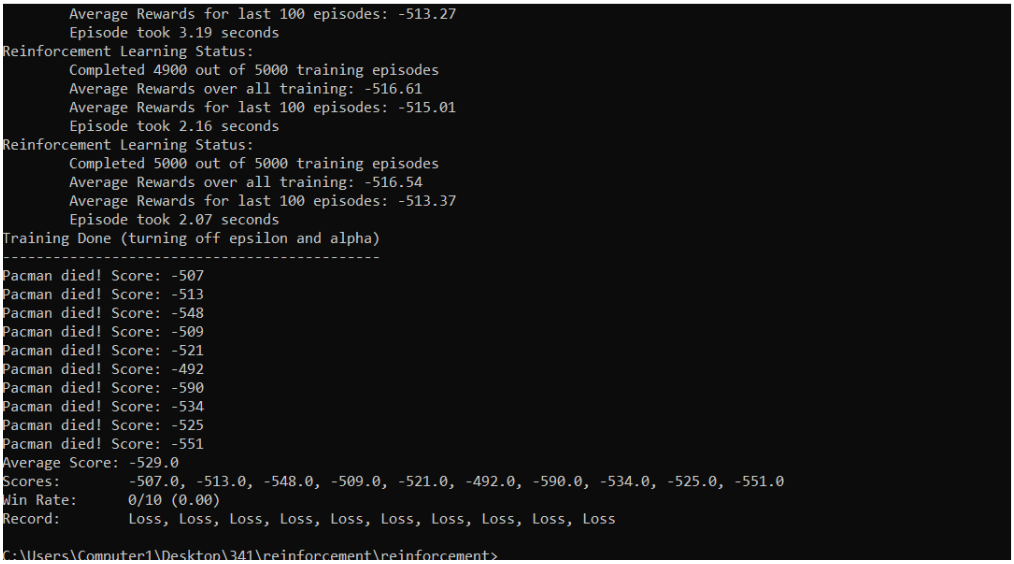
Question 4:

The results have differed drastically for both changes with the same amount of iterations. Worse paths have been achieved for the q-learning version and this might be because of the epsilon value that also considers the randomness in some cases when its value is high.

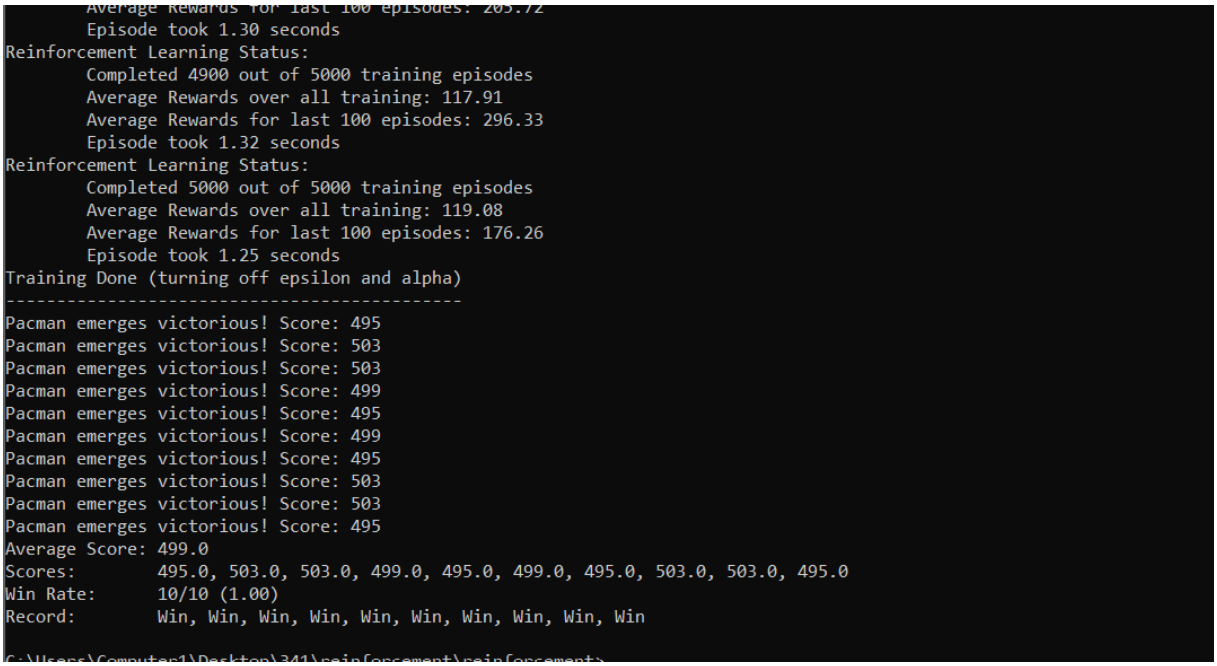
Question 5:

It is not possible. This is because when the epsilon value is 1, we simply move randomly, however for it to be 99% we need to make sure that we have explored the whole map and unfortunately 50 iterations is nowhere near close enough to the amount that will be sufficient enough for the agent to explore the whole map even if it completely moves randomly.

Question 6:



Versus



It works for smaller grids but not the medium and larger grids as it can also be observed from my results above. This is because number of state spaces that the agent must explore during the training period is way more dramatically with a larger gridspace when compared with the amount when the program is run in a small gridspace. Hence the agent cannot establish encountering a ghost as being something poor for all of the states