TDDC17-Lab5

Tasks for part II:

7. Question 1 :

a). In Task 2, we have chosen the states by making use of the *discretize* function provided and passing values such as the angle and we provided the nrValues as 8 so that we can divide the angles to a set of nrValues-1 set of values. We have provided the min and max values as -pi/4 and +pi/4 so that the values will be more discrete. After calculating the values we are creating the new state and updating the state. Thus we can see that the state is returned.

We have selected the reward functions by carefully observing the values of angles and we have seen that the most optimum value for angle when we discretize is 4 and hence in our reward function we give the maximum reward when the value is 4 . And for all the other discretized values we are providing the rewards according the closeness of the values to the most optimum value which is 4. Thus we define our reward function.

b). We have updated the different components in the Q-learning update. In the Q- update we have a variable to store the previous state, named as oldvalue from the Q-table. We need the alpha which is the learning rate, which act as an average over all possible future states of the actions. We have GAMMA\_DISCOUNT\_FACTOR which makes the agent very greedy and short sighted when the value of GAMMA is very near to one. Previous rewards is another factor which is needed to get the new updated value. Qvalues can be actually said as the sum of the rewards. Q-value for the optimal policy is the sum of the immediate reward, plus the discounted best Q-value in the next state, which should be done again and again. After getting the new value we should update the table with the updated new value. So that the Q learning table gets updated and hence the agent learns which is the best state based on the rewards.

8.

When we turn off exploration from start before learning, the Qvalue table won’t be updated and hence there wont be any learning by the agent. Since the values are not getting updated. The agent tends to move in any angle without trying to keep the upward position since it doesn’t know that the upward position will be optimum and will give the highest reward. This is the reason why tha gent moves in random direction if the exploration is turned OF.

Tasks for part III:

In task 3 we have used the discretize2 function and we are taking into consideration the velocities vx and vy along with the value of angle as in the previous case. Thus the agent takes into consideration velocities and hence try to keep the velocities to zero along with the angle as upward as possible. Thus we can see that in this task the states keeps track of all the three factors like the angle, vx and vy. Thus it learns how to hover by taking the most optimal values of all these three values using rewards.

We have attached the code as files.