

REPORT Lab 5

Group 2, Subgroup 7 (prigu857 and fabcr549)

Part II

1a. We chose 10 as the number of states and 1 as the base reward. The agent will get it when it is in state 5, which is the goal state in our case. The other states reward is defined as following: the closer the state is to the goal, the closer the reward will be to 1.

1b. The Main components used in QLearning are Qtable, NTable and Qvalue. Qtable will be used to store the Qvalue for the particular state-action pair. Ntable keep track of the number of times the same state-action pair has been executed. Qvalue is defined as the discounted reward the agent will get after applying all the combinations of state-action pair on a state.

There are some other components which are used to determine the Qvalue are GAMMA_DISCOUNT_FACTOR, LEARNING_RATE_CONSTATNT. GAMMA_DISCOUNT_FACTOR is the discount factor applied while calculating the reward for each state-action pair, so to prefer pairs that will get a higher near-term reward. The value of GAMMA_DISCOUNT_FACTOR lies between 0-1, in our case it is 0.9. LEARNING_RATE_CONSTATNT is used to compute the learning rate parament alpha based on the number of times each state-action pair has been tested. The alpha parameter is a weight that is applied to every Qvalue to reduce loss and converge to a solution.

2a. The QLearning algorithm will still try to maximize the reward. Given that the QTable has not been updated with the knowledge gained from exploring possible combinations of additional states, the algorithm will pick the initial action the simulation starts with, and it will try to stay in it.

Part III

We defined 10 states for the angle, 4 states for the horizontal velocity and 6 for the vertical velocity. The goal state for this part will be "an4vx3vy2".

The reward function is defined as a combination of the rewards coming from the multiple variables considered. Specifically, the closer the state is to the goal, the closer the reward will be to 1, as was happening in part 1. Moreover, we gave a higher priority to the vertical velocity contribution as suggested in the lab description. This was done by using constant weights while calculating the final reward.