COMP 417 Assignment 3

Due: Nov 7, 2022 at 11:59pm

1 Objectives

In this assignment, you are going to work with the same inverted pendulum benchmark as you did in the previous assignment by PID controllers. The main purpose is to stabilize the cart pole through a reinforcement learning approach. More specifically, we transfer our inverted pendulum problem into a grid world form where we want to find state and Q values to control the system. To obtain the mentioned grid world, it is necessary to discretize the spaces of states and actions. Following this approach, each cell in the resulted grid world represents a pair of state and action, while its value shows the corresponding Q value of being in that state and taking that action. To stabilize the pole angle, we take $(theta, theta-dot)$ as the two states required by the controller. Furthermore, considering a discretized action space, we assume that we have three options at each state: go to the left ($action = 0$), to the right ($action = 2$), or do nothing ($action = 1$).

2 States are accessible

In the base code, there are 40 steps considered in discretization of the state space for both of the mentioned states. This results in a grid world with shape of $(40, 40, 3)$ where the first two elements stand for the states and "3" corresponds to the number of possible actions at each state. Based on this:

(A) Initialize all Q values with 0, and include a random action which is applied when ever: $np.random.rand() > 0.8$. Then, stabilize the pole angle using the $argmax$ of Q values at each state. Note, you can use the random action in the beginning and for a finite time. Otherwise, this random action may play the role of a disturbance. (30%)

(B) Change 0.8 to 0.9, and discuss the difference and the role of the random action. (20%)

(C) As you may have seen, the cart most of the times runs out of the screen, what is the reason? Suggest a solution to solve this. (No need to code for this part, just explain your idea.) (15%)

(D) Following step 1, find the state values. Report the final 40×40 shaped state value matrix. (15%)

(E) Investigate the role of the learning rate and the discounting factor by changing them and conducting simulations to see various results. (20%)

3 Things to submit

In summary, for part A, report the final simulation where the pole angle is stabilized (theta vs time plot please).
For part B, try to stabilize the pole angle (does not necessarily work); then, report the best theta vs time plot from the most recent outcomes and explain the role of the random action.
In part C, just explain your suggested solution.
For part D, report the final state values.
For part E, perform different simulations with various learning rates to understand the role of this parameter in convergence and stability of the results; then, do the same for the discounting factor. You should explain what you got, no need for reporting the exact Q or state values.
You will need to submit both your code and a report(in pdf format please) zipped together. Please be concise and report interesting things like challenges.