DRL Homework 01

Group 8

April 22, 2022

Contents

1 Task 1

Set of states: $P = \{0, PW, LW, BW, RW, QW, KW, PB, BB, KB, RB, QB, KB\}$,
$S \subseteq P^{64}$

Set of actions: $A = \{(PW, F1), (PW, F2), (PW, CR), (PW, CL), ..., \}$

Probabilistic state dynamic: $p(s'|s, a) = 1$ if the move is legal , $p(s'|s, a) = 0$ otherwise

Reward: $R = \{1000, -1, \}$ Reward is 1000 for checkmate, -1 for making any move

Policy: $\pi(a|s)$: Probability distribution over all possible moves in a specific state
The opponent's moves are not known in advance, so we can address this by thinking about observations rather than states.

good formalization! Maybe you could add for the policy that it returns an action for each state.

2 Task 2

our main source: https://github.com/openai/gym/blob/master/gym/envs/box2d/lunar_lander.py

$S = \{x, y, v_1, v_2, angle, angular\_velocity, contact\_left\_leg, contact\_right\_leg\}$

Set of actions $A = \{left, right, bottom, nothing\}$

Reward: Crashes: -100 rest: 100 Each leg ground contact: 10 Fire main engine: -3 Solved: 200

Policy: $\pi(a|s)$: Probability distribution over firing left, right, bottom engine or doing nothing

again: good formalization! Same as in Task 1: You could add that the policy in the end maps an action to each state. Probabilistic state dynamics are missing.

3 Task 3

State transition function: Probability of ending in a state when making an action in another
Example: Robot wants to move forward but there is chance that the motors dont work properly and it doesnt move
Reward: Reward of performing this action and ending in a certain state.
Example: If we have a self driving car it has to be punished for taking too long because fuel runs out.
State dynamics are not always known and often have to be guessed or approximated. But this is almost impossible sometimes, making RL not practically for every problem.

We like the description of state transition function and examples given! The final sentences are also good, maybe you could give one example to make it even more clear.

Summary:
We really liked that you specifically formalized the MDPs. Maybe you could revise the defintion of a policy: We thought of it as a function that takes one state as input and returns a specific action, more than a probability distribution.
All in all: Good submission!