Homework Assignment

Course: Deep Reinforcement Learning

Sophie Haas, Johanna Linkemeyer, Tjorven Weber (Group 9)

April 30, 2022

1 Review for Group 12

Overall, nice homework! Still, we have some remarks :)

1.1 Task 1

- the state space in chess is actually unknown, but most definitely much larger than 768 (estimated complexity of 10^{46})

- the state transition function does not assess the validity of an action (to our knowledge), but, given an action and a state, returns a new state based on a probability distribution; in chess though, the state transition dynamic is deterministic

- action space: might need to consider that the movement of each piece (even same-typed pieces) are distinct actions and cannot not be summarized into only one action $\rightarrow$ that would result in a much larger action space

- rewards: The reinforcement book somewhere mentions, that it is actually better in the case of chess to not assign a positive reward for taking a piece (and negative reward for losing one) $\rightarrow$ that might actually lead to the policy trying to take out as many pieces as possible instead of focusing on winning

1.2 Task 2

- Very nice! The rewards could have been at least copy pasted though :p

1.3 Task 3

- Again nice answers! The examples would be more interesting though when being different from task 1 and 2

- The explanation on how actions are chosen based on the reward could be a little more extensive (as they are not directly based on the reward of the concurrent state, but on the expected overall return)

2 Review for Group 16

2.1 Task 1

- nice that you added additional sources

- detailed and well explained (mentioning that set of legal moves differs for each color)

- nice that you also explained your thoughts and gave reasons but sometimes no clear answer is given because of this (e.g. reward)

- policy: joint categorical distribution (should be stated concretely)

2.2 Task 2

- The set of states is simply the pixel space to our understanding.

- Your policy is basically the goal state, not a policy.

2.3 Task 3

- Your examples are not really helpful in understanding your definitions of the reward function and state transition function, it would be nice if you would have formulated them out a little bit more precisely.

- otherwise fine

3 Review for Group 20

3.1 Task 1

- could be more precise (e.g. board is 8x8, 32 pieces, different allowed moves for each type of piece etc.)

- probabilistic state dynamics: uncertainty not relevant, as moves are still deterministic

- reward: don't need to actually take the king; what about moves where no piece is taken?

- policy: joint categorical distribution (joint probability distribution over all of the categorical distributions for each chess figure)

3.2   Task 2

- probabilistic state dynamics: maybe also gravity or wind (generally just some physics simulation like with the self-driving car)

- reward: high reward for landing on the right place with low enough velocity, penalty for not landing on landing pad

- policy: good thinking that multiple actions can be taken at once

3.3   Task 3

- minor notation error for environment dynamics

- otherwise nice