Reinforcement Learning
DD2380

Marc Delgado Sánchez
Andrea Salinetti

KTH Royal Institute of Technology
School of Electrical Engineering and Computer Science

October 15, 2024

1. **Define the following properties of the *FishingDerbyRL MDP*:**

   (a) **State Space $\mathcal{S}$**: The total number of states of the environment (which is represented by a grid world)

   (b) **Action Space $\mathcal{A}$**: All possible actions carried by the diver.

   The State Space $\mathcal{S}$ is composed of every explorable cell of the *sea* in which the diver, the Jelly Fish and the King Fish are located.
   The Action Space $\mathcal{A}$ is composed of the possible movements that the diver can execute, mainly *up, down, left, right*.

2. **Define and test at least one interval for each of them accordingly, in order to achieve the following desirable policies:**

   (a) Not improving/learning.

   (b) High variance but fast learning.

   (c) Low variance and high long-term return.

   (d) High variance and high long-term return.

   For the model to not improve or learn anything, the learning rate α must be set to 0, so that the resulting policy is exactly equal to the initial one.
   To achieve high variance but fast learning, the learning rate α must take values close to 1.
   Moreover, to achieve high long-term return we must set the discount factor γ to high values, close to 1.
   Also, if we want to decrease the variance while learning, we must set the learning rate α to medium values like 0.5.

3. **If the reward structure of an MDP is simple enough, the optimal policy degenerates in a simple heuristic. Given the *3_2_3.yml* reward structure and initial position of jelly fish/king fish/diver, what is the value of the long term return of the optimal policy?**
   The model converges to the solution where the optimal policy consists of following the shortest path from the initial position of the diver to the King Fish's avoiding the Jelly Fish.