9Lab
Reinforcement Learning
from MDP to RL and Q-Learning, Deep Q-Learning, and MARL

Mirko Viroli
mirko.viroli@unibo.it

C.D.L. Magistrale in Ingegneria e Scienze Informatiche
Alma Mater Studiorum—Università di Bologna, Cesena

a.a. 2023/2024

One slide sum-up on Reinforcement Learning

MDP and RL

- MDP is essentially DTMC plus actions/rewards, and the goal of maximising cumulative reward
- optimal policy can be found by Bellman equations, which badly scale in practice, though
- large state-space: addressable by Dynamic Programming, i.e. iterative algorithms
- unknwon model or non-Markovianity: addressable by Monte Carlo approaches, i.e. simulation

Q-learning and beyond

- Q-learning: mixed DP and MC, with an off-policy learning that keeps improving the Q-table
- Deep Q-learning: represents Q-table by a DNN, with additionals ingredients to properly converge
- MARL: many agents call for a wider model and specific algorithms

Starting point and goals

References

- 09-repo: from virtuale (`https://github.com/mviroli/asmd23-public-models`)
- MARL and DQL repo: `https://github.com/cric96/advanced-reinforcement-learning-asmd-code`

General goals

- play with basic RL and Q-learning
- play with Deep Q-Learning
- play with MARL

Tasks

BASIC-Q-LEARNING

- Get acquainted with the basic tool of Q-learning, focussing on `examples/TryQLearningMatrix`
- check how variation of key parameters (ϵ, γ, α, episode length) affects learning
- check how learning gets more difficult as the grid size increases

DESIGN-BY-Q-LEARNING

- changing enviroment, state, rewards (and adding jumps, holes, items, enemies, walls, moving obstacles) you can really make your "robot" learn virtually anything, e.g.:
  - define an environment as a sort of corridor with obstacles: your robot has to "zigzag walk" to avoid them
  - "program by learning" a robot to collect items one at a time and go back to inital position
  - "program by learning" a robot to move obstacles so as to hide from enemies

MARL and Deep Q-Learning

- check MARL and DQL repo and its readme: it has tasks on playing with MARL, with DQL, and with high-level modelling of learning problems