AI Lab - Reinforcement Learning (RL)

Alessandro Farinelli

Thanks to Alberto Castellini, Davide Corsi, Luca Marzari and Celeste Veronese for help with slides and code

University of Verona
Department of Computer Science

May 2024

Start Your Working Environment

Update your repository to download the new lesson

Important: do a backup copy of your working directory to make sure you avoid any issue

```
> cd AI_Lab
> git commit -a -m "a message describing the commit"
> git pull
> conda activate ai-lab
> jupyter notebook
```

- Your assignments for this lesson are at: *RL/RL_4_problem.ipynb*. You will be required to implement Q-Learning and SARSA algorithms
- In the following you can find the pseudocode

Q-Learning

```
Input: $environment\ [A, S], problem, episodes, \alpha, \gamma, expl\_func, expl\_param$
Output: $policy, rewards, lengths$
$\forall a \in A, \forall s \in S$ initialize $Q(s, a)$ arbitrarily
$rewards, lengths \leftarrow [0, ..., 0]$        ▷ Null vectors of length $episodes$
for $i \leftarrow 0$ to $episodes$ do
    Initialize $s$
    repeat
        $a \leftarrow \text{EXPL\_FUNC}(Q, s, expl\_param)$
        $s', r \leftarrow$ take action $a$ from state $s$        ▷ Act and observe
        $Q(s, a) \leftarrow Q(s, a) + \alpha(R + \gamma \max_{a' \in A_s} Q(s', a') - Q(s, a))$        ▷ TD
        $s \leftarrow s'$
    until $s$ is terminal
    Update $rewards, lengths$
$\pi \leftarrow [0, ..., 0]$        ▷ Null vector of length $|S|$
for each $s$ in $S$ do        ▷ Extract policy
    $\pi_s \leftarrow \text{argmax}_{a \in A_s} Q(s, a)$
return $\pi, rewards, lengths$
```

SARSA

```
Input: $environment\ [A, S], problem, episodes, \alpha, \gamma, expl\_func, expl\_param$
Output: $policy, rewards, lengths$
$\forall a \in A, \forall s \in S$ initialize $Q(s, a)$ arbitrarily
$rewards, lengths \leftarrow [0, ..., 0]$        ▷ Null vectors of length $episodes$
for $i \leftarrow 0$ to $episodes$ do
    Initialize $s$
    $a \leftarrow \text{EXPL\_FUNC}(Q, s, expl\_param)$
    repeat
        $s', r \leftarrow$ take action $a$ from state $s$        ▷ Act and observe
        $a' \leftarrow \text{EXPL\_FUNC}(Q, s', expl\_param)$
        $Q(s, a) \leftarrow Q(s, a) + \alpha(R + \gamma Q(s', a') - Q(s, a))$        ▷ TD
        $s \leftarrow s'$
        $a \leftarrow a'$
    until $s$ is terminal
    Update $rewards, lengths$
$\pi \leftarrow [0, ..., 0]$        ▷ Null vector of length $|S|$
for each $s$ in $S$ do        ▷ Extract policy
    $\pi_s \leftarrow \operatorname*{argmax}_{a \in A_s} Q(s, a)$
return $\pi, rewards, lengths$
```