AI Lab - Lesson 5

Reinforcement Learning

Davide Corsi
Alessandro Farinelli

University of Verona
Department of Computer Science

January 13th 2022

UNIVERSITÀ
di VERONA
Dipartimento
di INFORMATICA

Start Your Working Environment

Start the previously installed (Session 1) conda environment *ai-lab*

Listing 1: Update Environment

```
cd AI-Lab
git stash (NB: remember to backup the previous lessons before this step!)
git pull
git stash pop
conda activate ai-lab
jupyter notebook
```

Listing 2: Open Lesson

To open the tutorial navigate with your browser to:
lesson_4/lesson_4_problem.ipynb

Assignments

- Your assignments for this session are at: *lesson_4/lesson_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
    $\pi_s \leftarrow \operatorname{argmax}_{a \in A_s} Q(s, a)$
**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
    $\pi_s \leftarrow \underset{a \in A_s}{\operatorname{argmax}}\ Q(s, a)$
return $\pi, rewards, lengths$
```