

Assignment 3

Reinforcement Learning - A.Y. 2025/2026

Assigned: December 2nd, 2025

Deadline: December 17th, 2025

Rules

The assignment is due on December 17th, 2025. Students may discuss assignments, but **each student must code up and write up their solutions independently**. Students must also indicate on each homework **the names of the colleagues they collaborated with** and what online resources they used.

The theory solutions must be submitted in a PDF file named “XXXXXXX.pdf”, where XXXXXXX is your matricula (write your matricula also inside the PDF). We encourage you to type the equations in an editor rather than uploading a scanned written solution. **In the PDF** you have to hand over **the answers to the theory questions (not just the numerical results, but also the derivations)** and a **small report of the practice exercises**.

The practice exercises must be uploaded in a zip file named “XXXXXXX.zip”, where XXXXXXX is your matricula. **The zip file must have the same structure of the assignment.zip** that you find in the attachments, but with the correct solution. You are only allowed to type your code in the files named “student.py”. Any modification to the other files will result in penalization. You are not allowed to use any other python library that is not present in python or in the “requirements.txt”, if you want to use an additional library, you must ask for permission. You can use as many functions as you need inside the “student.py” file, however you **cannot** modify the signature or return of any function already provided to you in the file, unless strictly stated by us. The zip file you will submit **must have** the same structure as the assignment.zip

All the questions must be asked in the Classroom platform but it is forbidden to share the solutions on every forum or on Classroom.

Theory: REINFORCE

Suppose you have an environment with 2 possible actions and a 2-d state representation ($x(s) \in \mathbb{R}^2$). Consider the REINFORCE Algorithm with the following Linear Function Approximator (LFA) policy (Logistic Regression) such that:

$$\pi(a = 1 \mid s) = \sigma(w^\top x(s)) \quad (1)$$

$$a = \mathbb{1}_{\pi(a=1|s)>0.5} \quad (2)$$

where $w = [0.8, 1]^\top$ are the weights and $y = \sigma(t) = \frac{1}{1+e^{-t}}$ is the sigmoid function.

Suppose you are doing an iteration of the REINFORCE algorithm and you have just run an episode getting the following trajectory:

$$x(s_0) = [1, 0]^\top, \quad a_0 = 0, \quad r_1 = 0 \quad (3)$$

$$x(s_1) = [1, 0]^\top, \quad a_1 = 1, \quad r_2 = 1 \quad (4)$$

$$x(s_2) = [0, 1]^\top \quad (5)$$

Show the weights w update according to the REINFORCE algorithm ($\alpha = 0.1, \gamma = 0.9$).

Practice

Solve the CarRacing-v2 gym environment using one of the following algorithms:

- Double DQN with Prioritized Experience Replay (PER) (<https://arxiv.org/pdf/1511.05952.pdf>)
- World Models (<https://arxiv.org/pdf/1803.10122.pdf>)
- Advantage Actor-Critic (A2C) (<https://arxiv.org/pdf/2205.09123.pdf>)
- TRPO (<https://arxiv.org/pdf/1502.05477.pdf>)
- PPO (<https://arxiv.org/pdf/1707.06347.pdf>)

If you choose to implement World Models, you are allowed to use the CMA library <https://pypi.org/project/cma/>.

In the folder “car racing” you find three files:

- “main.py” that contains the main script to evaluate your solution. Don’t modify this file!
- “student.py” is the file you have to modify, by implementing the algorithm you have chosen.
- “requirements.txt” contains the name of the libraries needed for this part of the assignment.

For this exercise, in addition to the python source code, you must also submit a trained model. Make sure the model can be loaded correctly using the load method in student.py. You can freely modify the save and load methods as long as their signature remains the same as the template. Do **not** also include additional datasets if they are not strictly required to execute the policy (e.g. for World Models), especially if they have a very large size. To train the model you submit, you can use any strategy that makes use of the algorithm you chose to implement.

You may add some requirements, but they need to be authorized on Classroom. In order to request them, place a private comment under the assignment post with the **name and version** of the libraries you want to add. **Stable-baselines and any other library that already implements the algorithm are banned.** You need to use **pytorch** as the deep learning framework.

For this exercise, **3 bonus points** (valid for the final exam grade) will be awarded to the 3 students delivering the 3 best performing models.

Note: the algorithms among which you can choose have different implementation complexities. In ascending order (from less to most complex): A2C, PPO, DDQN+PER, TRPO, World Models. Overall, we will assign grades by considering both the performance of the final model, the complexity of the algorithm that you chose and any improvements/additions that you made (implementation details, reward shaping, observation preprocessing, etc.). For example, even if your model is performing poorly, you can still receive a high grade if you correctly implemented a complex algorithm or you implemented well-motivated improvements/additions to it.