Assignment 3: DQN, Model Based and Policy Gradient

Reinforcment Learning - A.Y. 2022/2023

November 18th, 2022

Rules

The assignment is due on December 4th, 2022. 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. We encourage you to type the equations on 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" file. You can use as many functions you need inside the "student.py" file. The zip file must have the same structure of 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

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:

$$\pi(a = 1|s) = \sigma(w^T x(s)) \tag{1}$$

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

where $w = [0.8, 1]$ 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:

$$x(s_0) = [1, 0]^T, \quad a_0 = 0, \quad r_1 = 0 \tag{3}$$

$$x(s_1) = [1, 0]^T, \quad a_1 = 1, \quad r_2 = 1 \tag{4}$$

$$x(s_2) = [0, 1]^T \tag{5}$$

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 proportional prioritization (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)

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.
- "requirements.txt" contains the name of the libraries needed for this part of the assignment.

You may add some requirements (include the version) but they need to be authorized on Classroom. In order to request them place a comment under the assignment post. (Stable-baselines and other libraries that already implement the algorithm are banned. You need to use py-torch as deep learning framework).

The grade will be assigned basing on the correctness of the code. 3 additional points will be awarded to the 3 best students according to the following criteria:

- agent performance
- algorithm/implementation complexity