Introduction

The objective of this assignment is to gain experience with Deep Q learning and to produce a paper which looks like a research paper.

Task: Given a fixed neural network, you will train an agent which will be able to play cartpole V1 on different pole lengths, e.g. make a generalist.

The assignment is done in groups of 3 and you will write a report around your research question and findings (explanation follows).

The deadline – not negotiable – is as follows:

Friday October 17, 23:59

Each day of delay in the submission will result in a deduction of 1 point from your assignment grade.

Task:

You will train an agent given the ‘standard’ neural network (provided in the test script) using Deep Q learning. During training you are only allowed to train on different pole lengths. You can adjust the pole length of the cartpole using the following command:

env.unwrapped.length = put_length_here.

You can visualize the cartpole by adding the render_mode = “human”.

env = gym.make("CartPole-v1", render_mode="human")

You can train as many episodes as you want.

It is not allowed to change any other gym dynamics of the environment (car size, force magnitude, etc)!

After training you should test your agent on the test script provided. The goal is to maximize performance on all pole lengths (sizes = np.linspace(0.4, 1.8, 30)). Meaning, we test 30 linear divided pole lengths running from 0.4 up to 1.8. It is not allowed to use the outcome of the test script as a loss for the neural network.

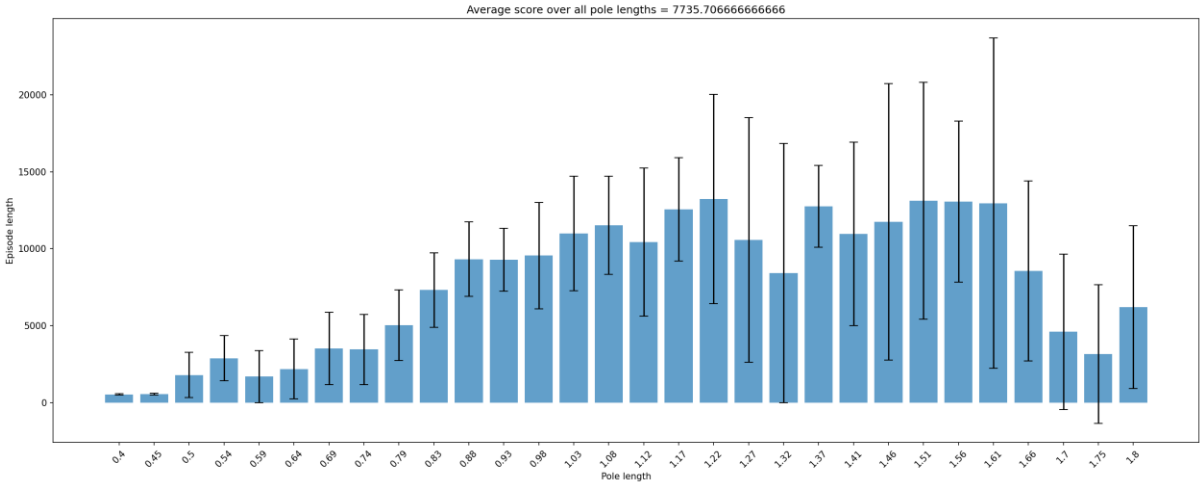
The test script simulate as follows:

1. Phase 1: The first 500 steps it simulates the environment normal given the pole length
2. Phase 2: From 500 to 1000 it throws in “wind attacks” every 25 steps. Meaning that the car will get a bump to either left or right to bring the pole out of balance
3. Phase 3: From 1000 onwards until the end, there will be more force (push left or right) applied to the car than usual. This will linearly increase using the following formula:

   env.unwrapped.force_mag = 25 + (0.01 * total_reward)

   total reward is increased +1 every time step during an episode

As there is stochasticity in the simulation, every pole will be tested 10 times and its result will be averaged. In your report please provide a bar plot showing the avg + std results per approach (see below). In the title of the bar, mention the overall average over all pole lengths. Example code for plotting is provided in the test script.



3 training strategies

We expect 3 different training strategies where you compare those. Example ideas:

- Different training order of the pole lengths
- Experiments around the replay buffer
- Modification of the reward function
- Different hyperparameter (or schedules of parameters)
- Creativity is rewarded here.

For all 3 training strategies report the test results.

Note: most of your grade is not determined by the performance of your experiments but from your motivation, explanation, etc. So, make sure you take time to set up a well motivated experiment. You will build a research question around your strategies, which is the main motivation of your paper.

The report:

Your report should have a maximum of 3 pages (containing everything, e.g., text, figures, etc).

You need to include the following info at the BEGINNING of the assignment: Name of the course; Number (and possibly name) of the team (e.g., Team 12, Team Jacob); Name and student ID nr. of the team members; Date. It is allowed to use an extra page for that, but this cover page should not contain anything beyond this identification info. It is also allowed to add one page as appendix, which can only contain figures.

The pages describing the actual content (3 pages only) should be as follows:

- Introduction (make sure to define a clear research question or goal). State here the motivation for your research.
- Methods: explain your different approaches and its motivation, parameter settings, experimental setup , etc. Make sure everything is reproducible with the information presented.
- Results and discussion: discuss the differences between the results of your approaches ; do they outperform each other? Comment on a possible explanation for that; Discuss the differences between your results, are they better/equal/worse? If possible, add significance scores. Comment on a possible explanation for that.
- Literature list / bibliography: cite all works you are using in your project

Please make your report in overleaf, you have to use the template provided on Canvas.

Tips:

- Do not blindfold yourself on performance. The most impact for your grade can be achieved by doing well thought research. Do you have a good motivation? A clear explanation? A good analysis? A good research question?
- Try to come up with interesting training approaches. What do you want to investigate?
- The goal for you is to produce a research paper. Make sure to also include references.

TA session:

- During the TA session the TA's will have slides to explain Deep Q Learning, and also to help you with brainstorming and giving feedback around the 3 different training strategies.
- There will also be TA sessions in the week before the exam. Oct 13-oct 17

Code:

- Please put a link to your github repository with your code.
- Put the trained 'standard' neural networks from your 3 strategies in your repository.
- Please view the grading rubric below for all pointers.

Contribution:

We expected an equal contribution from all group members. If some members contribute less they could be penalized or excluded.

Grading rubric:

Problem definition	**10%**	Research question clearly defined Goes beyond only performance
Performance	**15%**	Average performance over all poll lengths of best performing algorithm
Algorithm	**25%**	1. Motivation for the algorithms 2. Three algorithms/variants of mechanisms are tested. 3. Complete and conscise description of algorithms. 4. Explained and justified choice of parameters
Methodology	**30%**	1. Clear description of experiment procedure 2. The research is motivated with related work, e.g. other papers. 3. Research reproducibility: report everything required to reproduce the results. 4. Have appropriate plots and statistical tests for the expected metric. Have repeated the experiments due to stochasticity. 5. Discussion of comparing the three algorithms. 6. Discussion of comparing the results with a baseline.
Structure	**10%**	1. Use of correct grammar and academic language 2. Clear separation and sense of structure. 3. Supporting points are presented in logical progression 4. Sections have high degree of coherence. 5. Argumentation structure is excellent. 6. Algorithm details are not mixed with coding details. 7. Structure of the paper follows the research paper format provided
Code	**10%**	1. Well organized code and readable. 2. Code comments, e.g. explaining the key parts of the code. 3. Modual code for potential adjustments or extensions. 4. Trained neural networks in the repository.