

RL benchmark for the reaching task

1 Introduction and methods

A2C with Pendulum-V0 and 100 000 timesteps + 10 seeds check 1 to 8 env check normalise or not

2 Results

2.1 Learning curves

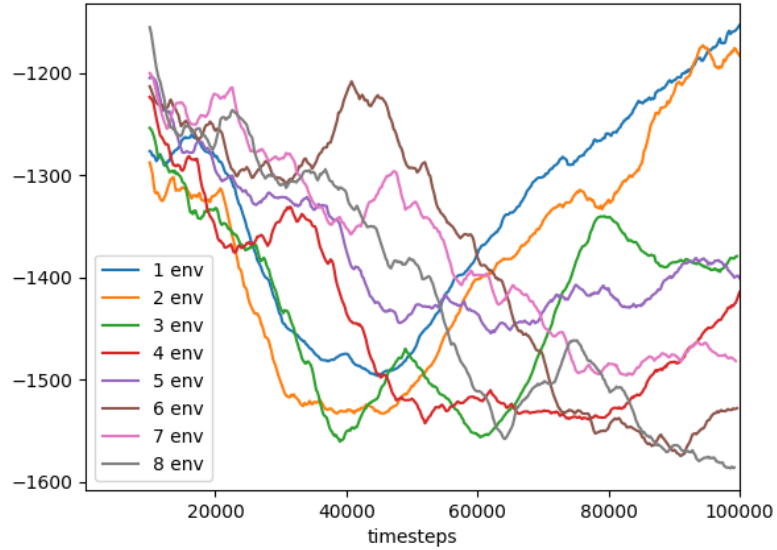


Figure 1: Learning curves by nb of env.

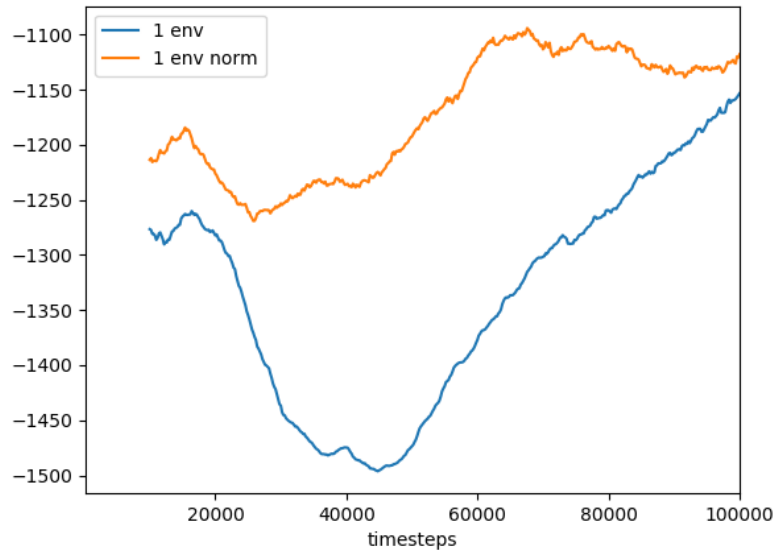


Figure 2: Learning curves when normalising or not with 1 env.

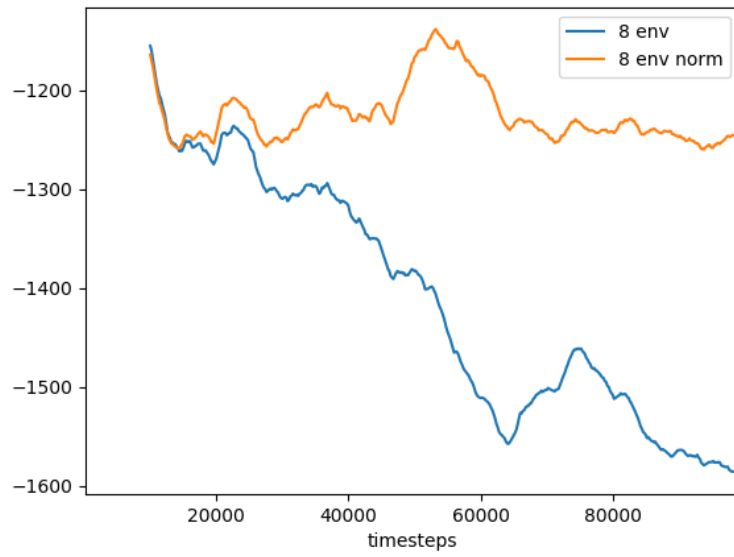


Figure 3: Learning curves when normalising or not with 8 env.

2.2 Evaluation

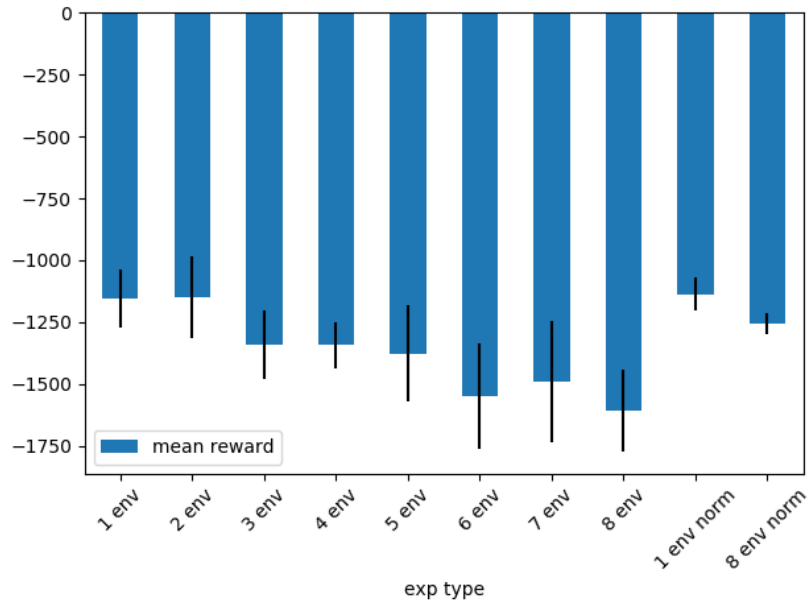


Figure 4: Reward vs env type.

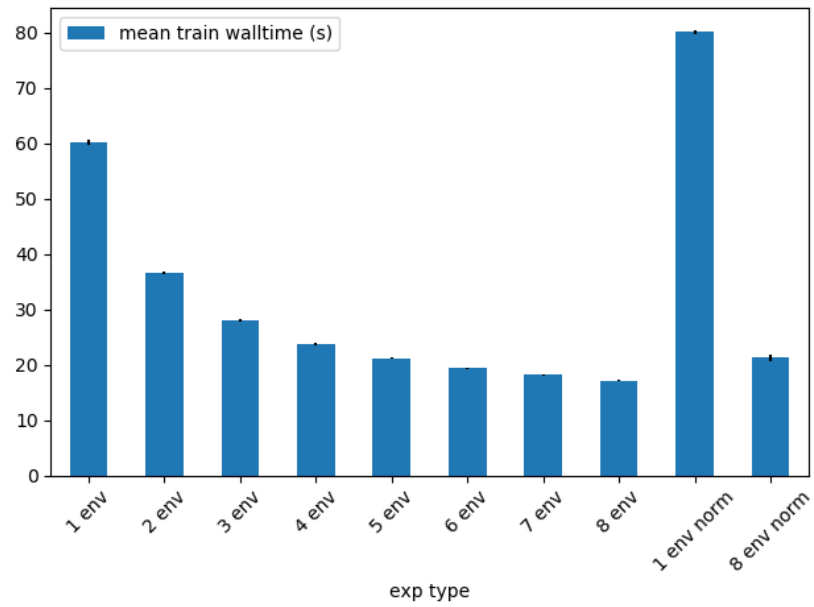


Figure 5: Train time vs env type.

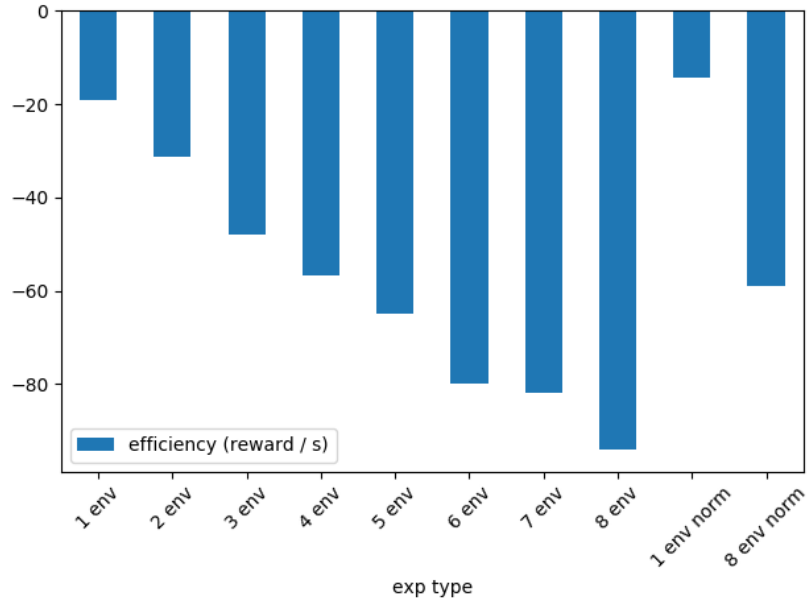


Figure 6: Efficiency vs env type.

3 Findings summary

1. Increasing the nb of env tend to harm the performance.
2. Normalising helps.
3. Normalising increases the training time.
4. The most efficient is to use one env and to normalise
5. I NEED TO RE RUN THIS WITH MORE TIME STEPS.