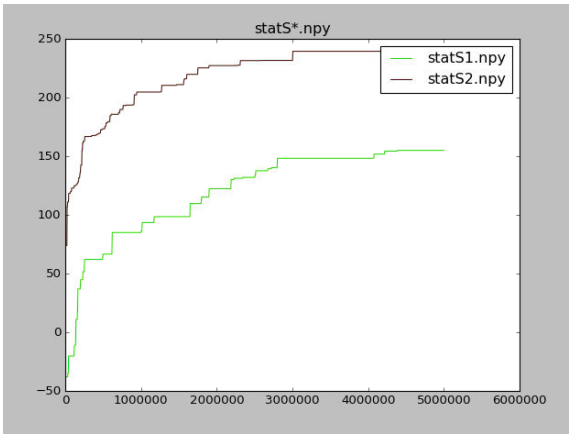
Exercise 7

Task description

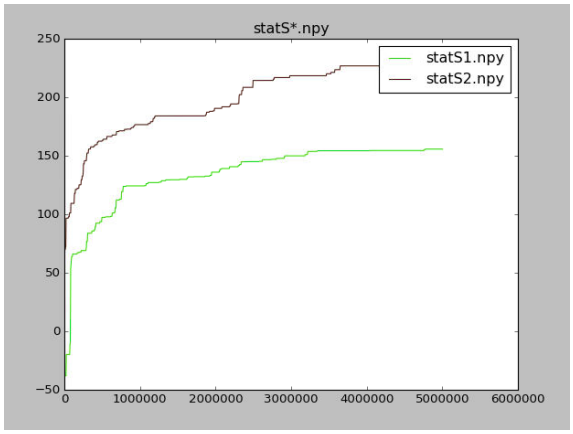
Train the policy for a pybullet locomotor problem with the PPO algorithm (do not forget to restore the original pybullet files with the reward functions suitable for reinforcement learning, as described in Section 3.3). Compare the result with those obtained with evolutionary strategies.

Results

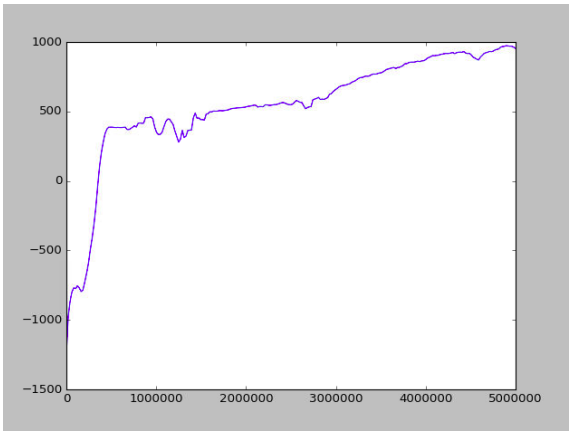
On Fig. 1 you can see the results of training two models on the original and custom rewards. Seed 1 was intended for halfsheetah model, the seed 2 was for hopper model.



(a) Feed forward NN with Original reward



(b) Feed forward NN with Custom reward



(c) MLP NN with Custom reward, train algorithm: PPO2

Figure 1: Results of exercise 4 and exercise 7

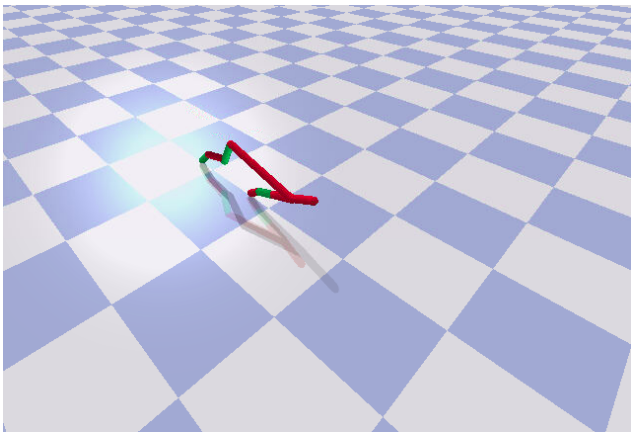


Figure 2: Halfsheetah

The halfsheetah model you can see on Fig. 2. The task for the agent to move as far as possible in one jump. As you can see when using the standard algorithm, the maximum policy is about 150 with custom rewarding. In the case of the algorithm PPO 2 - this value tends to 1000. Which means that this algorithm allows you to quickly converge to the maximum solution.