Final task (DCEM, DQNSoft, PPO, SAC). Отчет

Содержание

1. В финальном домашнем задании необходимо ответить на вопрос, какой же алгоритм все таки лучше решает ту или иную задачу. А именно, нужно рассмотреть одну из следующих задач:
   - CartPole
   - ~~Pendulum~~
   - ~~LunarLender с дискретным пространством действий (по умолчанию)~~
   - ~~LunarLender с непрерывным пространством действий (нужно положитьcontinuous=True, см. пояснения здесь Lunar Lander)~~
   - ~~BipedalWalker~~

   Нужно сравнить следующие алгоритмы CEM, DQN Soft / Hard Target Network, PPO, SAC.

Выбранная игра: **CartPole-v1.**
Результаты можно посмотреть в ClearML

Общие гиперпараметры алгоритмов:

- **gamma:** 0.99
- **learning rate:** 0.0001
- **batch size:** 64
- **tau:** 1e-4
- **episode n:** 100
- **trajectory n:** 20
- **t max:** 500

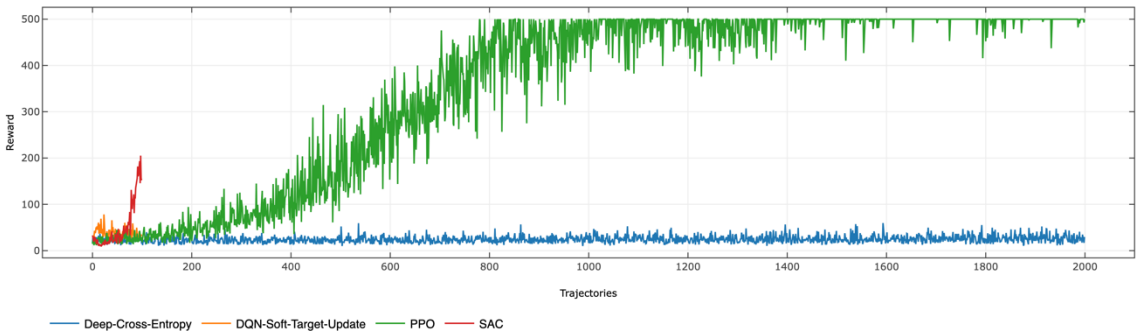
Размер сетей в алгоритмах: 3 линейных слоя 64 нейрона в скрытом слое с функцией активацией ReLU.

Гиперпараметры алгоритмов:

- **Deep Cross Entropy:** q_param = 0.8
- **DQN Soft Target Update:** epsilon_decrease = 0.01, epsilon_min=0.01
- **PPO:** epsilon = 0.2, epoch_n = 100, v_lr = 5e-2
- **SAC:** alpha = 1e-3, q_lr = 5e-4, temperature = 1

Каждый алгоритм запускался 3 раза и результаты усреднялись.

Reward на итерациях обучения:



Вывод:
При выбранных общих гиперпараметрах и размера НС лучше всего задачу CartPole решает алгоритм PPO.
Алгоритму SAC не хватает количество итераций обучения, алгоритмы Deep Cross Entropy и DQN Soft Target Network не может выучиться и показывает случайное поведение.