

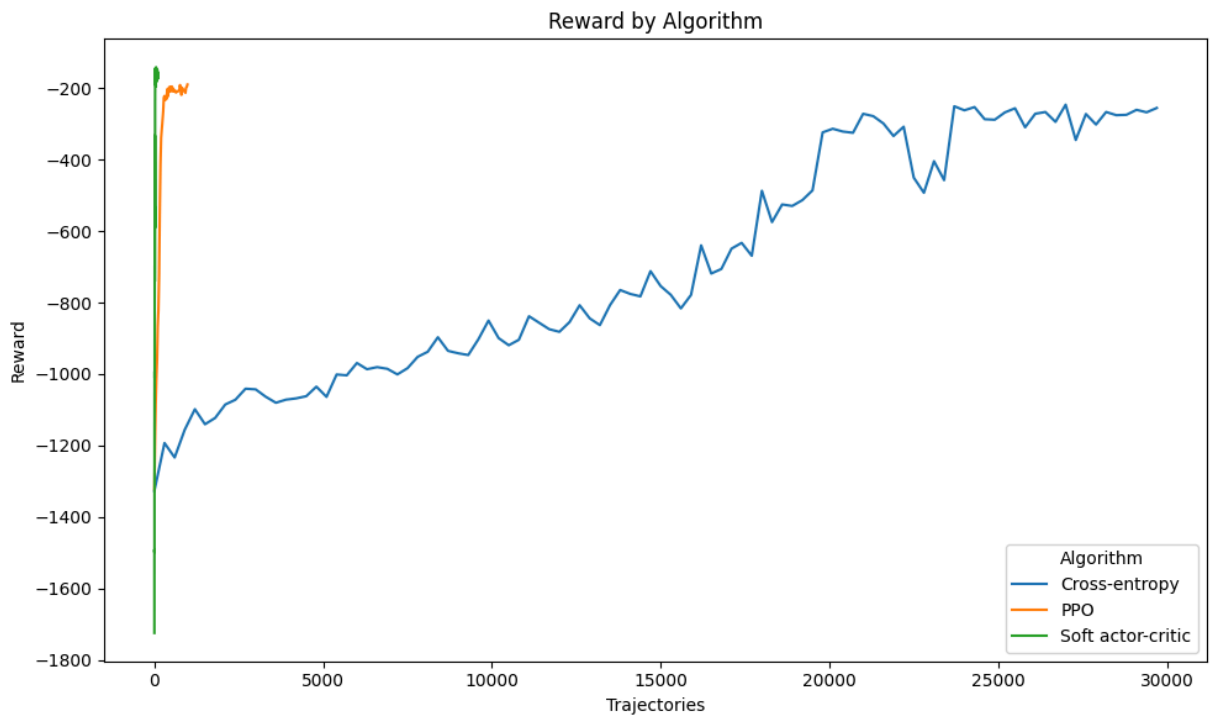
Задание 1

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:

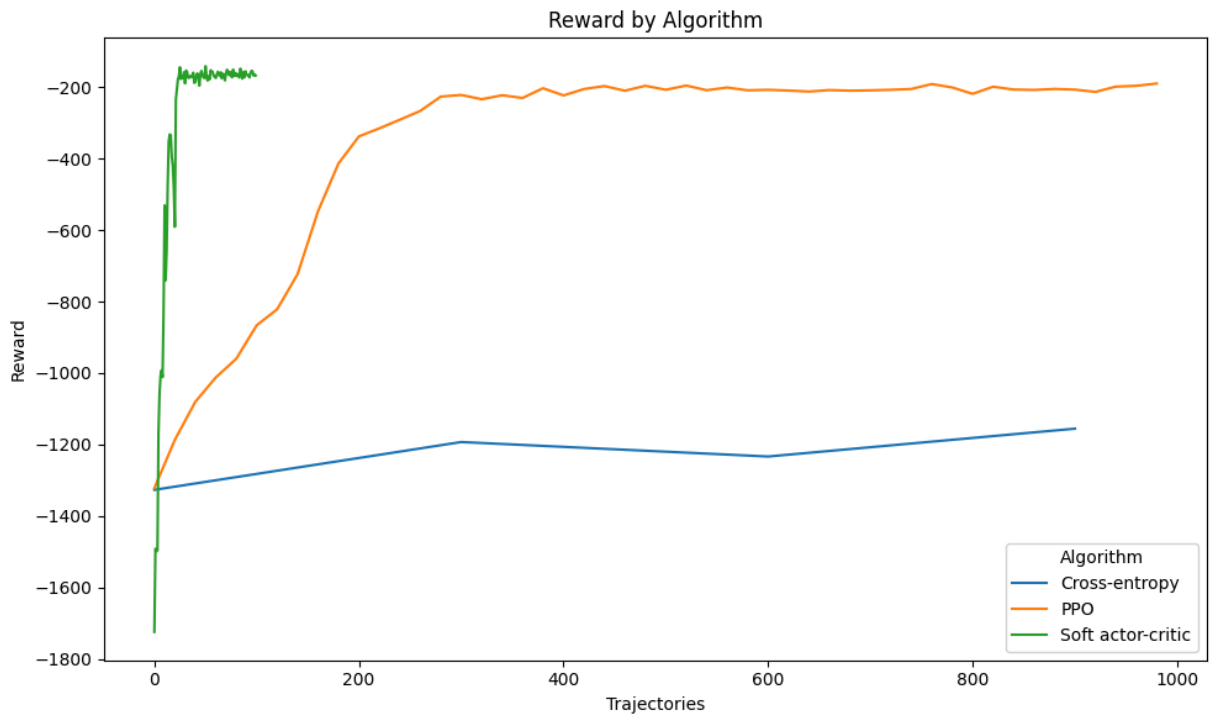
- CartPole
- Pendulum
- LunarLander с дискретным пространством действий (по умолчанию)
- LunarLander с непрерывным пространством действий (нужно положить `continuous=True`, см. пояснения здесь Lunar Lander)
- BipedalWalker

В данной работе рассматривается задача **Pendulum**

In [28]:



In [32]:



В данной работе сравнивается SAC, PPO и Cross entropy. DQN не рассматривается. Второй график приближает начало обучения, потому что скорость алгоритмов слишком разная. SAC обучается намного быстрее, PPO немного медленнее, кросс энтропии нужно огромное количество обращений к среде, чтоб приблизиться к стабильно высоким наградам.

In []: