Тагиров Али, ИТМО

Задание 7

В последнем задании рассмотрел задачу **Pendulum.**

При сравнении алгоритмов использовал следующие гиперпараметры:

- learning rate = 0.001
- batch size = 64
- gamma = 0.99
- tau = 0.01 (DQN)
- nn hidden size = 128

Так как в задаче Pendulum пространство действий непрерывное, для метода DQN Soft Target его следует дискретизировать:

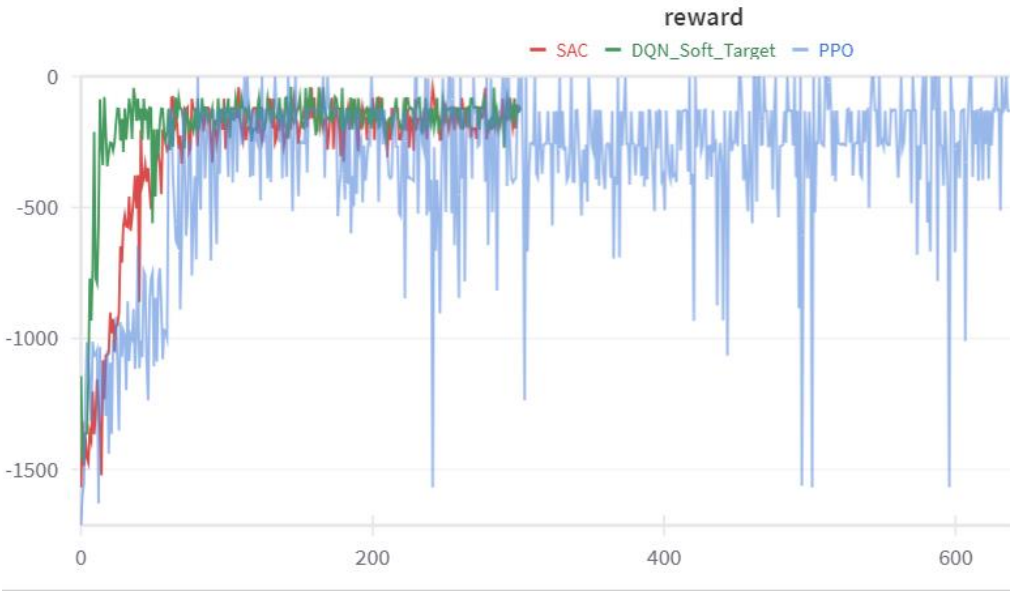
```
action_dim = 5
action_space = np.linspace(-2, 2, num=action_dim)
```

```
action_index = agent.get_action(state)
action = action_space[action_index]
next_state, reward, done, _ = env.step([action])

agent.fit(state, action_index, reward, done, next_state, t)
```

Не смог обучить CEM, попробовал дискретизировать пространство действий, как и для DQN, также пробовал, как во втором задании использовать в качестве функции активации гиперболический тангенс и MSE в качестве функции потерь.

Усредненные по трем запускам кривые обучения алгоритмов:



При указанных гиперпараметрах лучше всего себя проявила модель DQN Soft Target, модель быстрее всех обучилась, а также имеет наименьший разброс, при том, что пространство действий пришлось дискретизировать. Хуже всех себя показала модель PPO, которая имеет сильный разброс по сравнению с конкурентами.