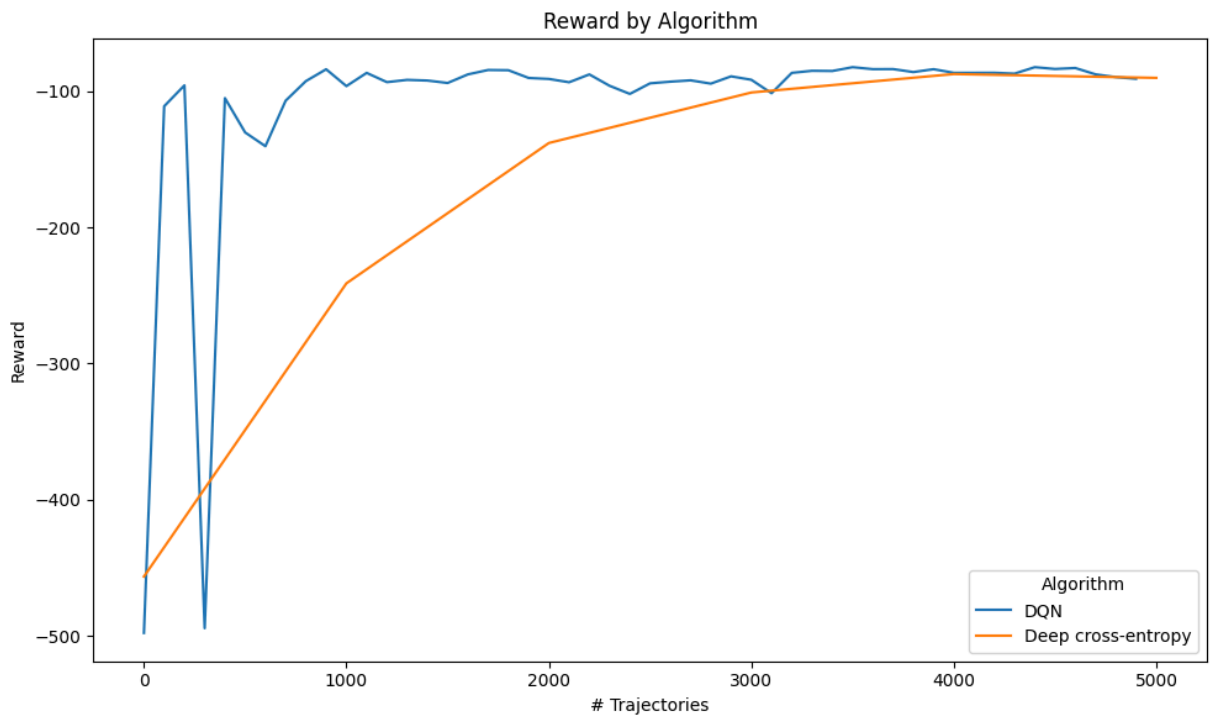


Задание 1

Обучить Агента решать Acrobot-v1, MountainCar-v0, или LunarLander-v2 (одну на выбор) методом DQN. Найти оптимальные гиперпараметры. Сравнить с алгоритмом Deep Cross-Entropy на графиках. Помним про корректность сравнения.

В данной работе рассматривается Acrobot-v1.

In [58]:

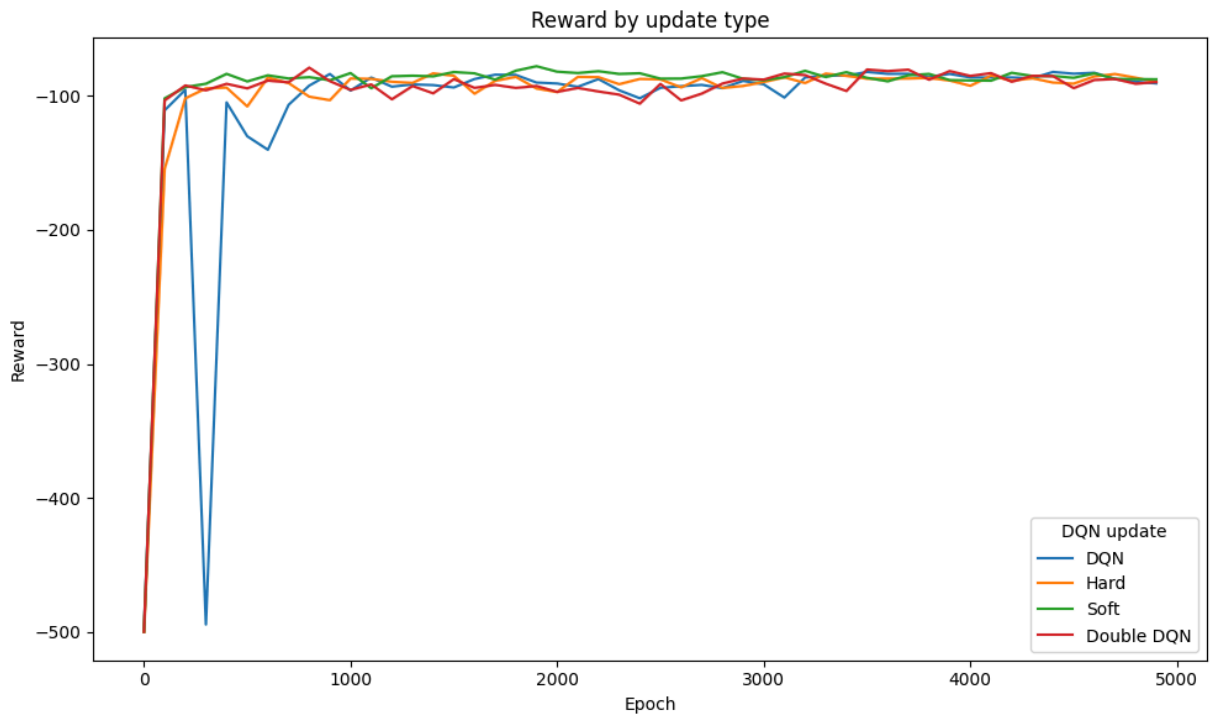


Для корректности сравнения графики построены в зависимости от количества траекторий, поэтому в методе кросс энтропии точки для каждой 1000 траекторий (размер популяции). Сети для обучение обоих алгоритмов имеют одинаковую архитектуру и параметры обучения. Из графика видно, что алгоритм DQN сходится к оптимальному значению существенно быстрее, и после 1000 итераций уже стабильно решает задачу. Для кросс энтропии 1000 траекторий это только одна итерация и решения среды ещё нет.

Результат:

- DQN сходится намного быстрее метода кросс энтропии.

In [63]:



Сравнивая разные подходы направленные на устранение автокорреляции, можно однозначно заключить, что все они работают намного лучше, чем оригинальный DQN. Конкретно для данной среды Double DQN и Soft update показывают чуть лучшие результаты, чем Hard update, но кажется, что эти различия лежат в пределах погрешности.

Результат:

- Все методы устранения автокорреляции улучшают DQN.
- Double DQN и Soft update показывают лучшие результаты.

In []: