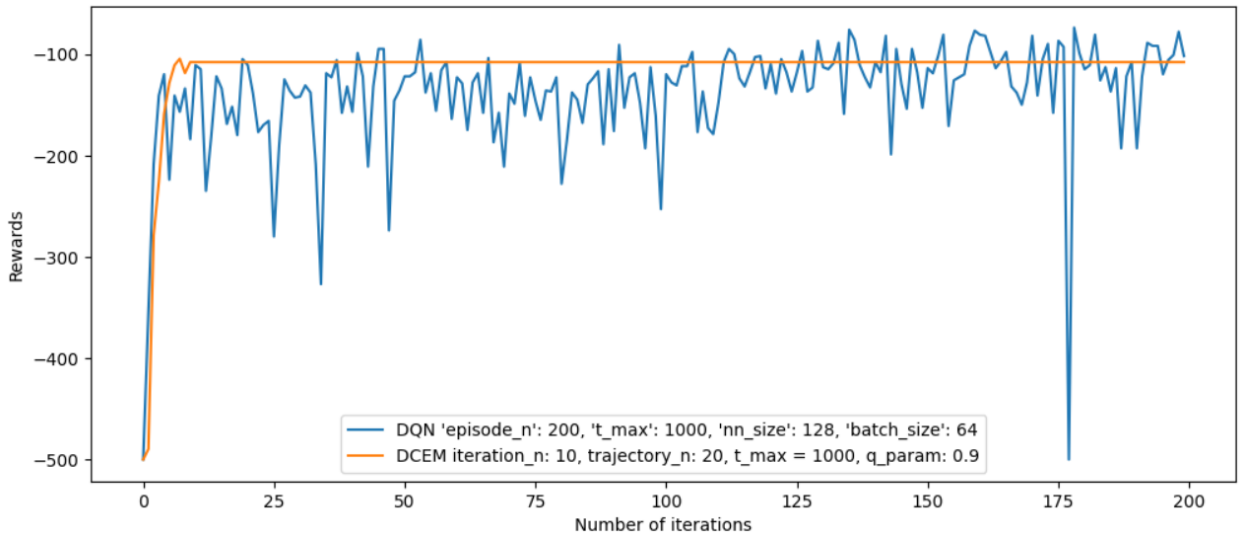
Задание 1. Обучить Агента решать Acrobot-v1. Сравнить с алгоритмом Deep Cross-Entropy на графиках.

Эксперимент: Сравнение наград DQN и DCEM



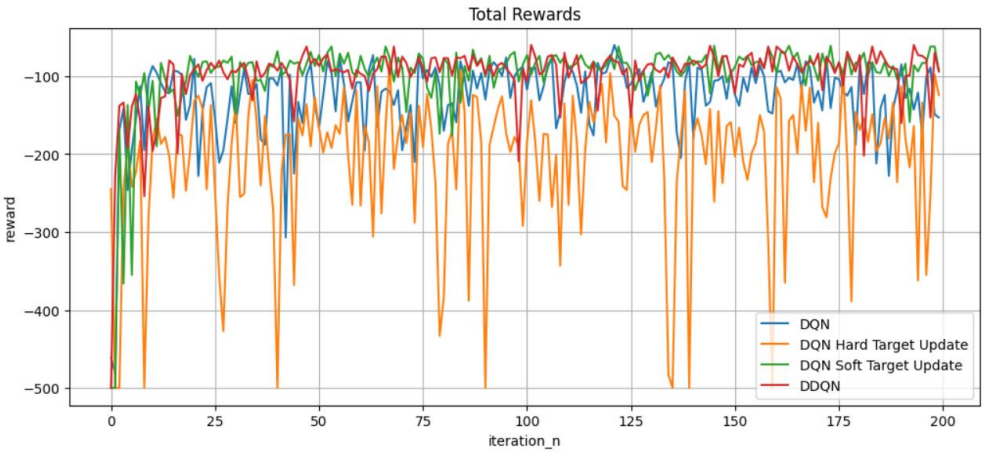
Вывод.

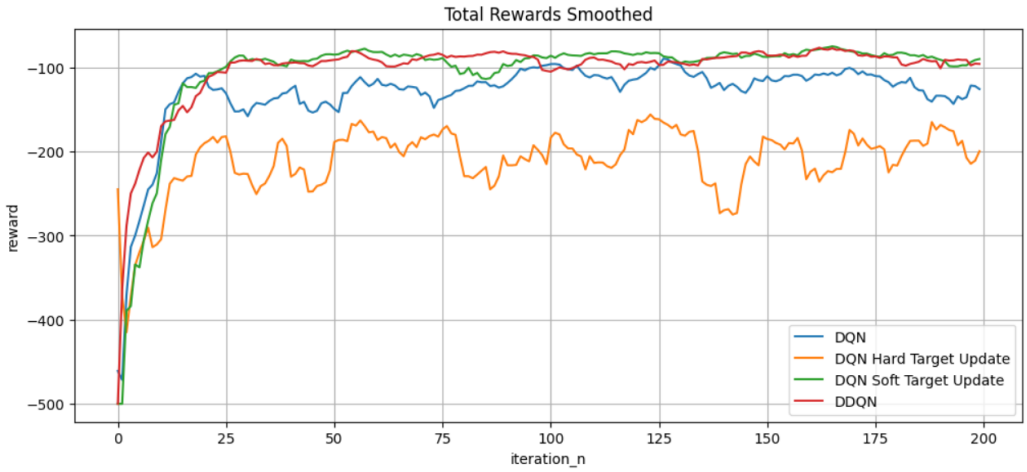
После подбора оптимальные параметры для алгоритма DQN получились следующие {'episode_n': 200, 't_max': 1000, 'nn_size': 128, 'batch_size': 64}. Для корректного сравнения я использовал количество обращений к среде, поскольку кросс энтропии для нормального обучения необходимы 1000 траекторий и 20 итераций внутреннего цикла, количество итераций внешнего цикла пришлось сократить до 10 приравняв количество траекторий к количеству DQN. Траектории CEM 10 * 20 * 1000 == траектории DQN 200 * 1000.

По результатам видно что алгоритм кросс энтропии быстро сходится и достигает сопоставимого качества с DQN алгоритмом, который в свою очередь является не совсем стабильным, возможно из-за проблемы с автокоррелаяцией.

Задание 2. Реализовать и сравнить с DQN следующие модификации, DQN с Hard Target Update, DQN с Soft Target Update и Double DQN

Эксперимент: Сравнение наград DQN, DQN Hard update, DQN Soft update, DDQN





Вывод.

По результатам soft сглаживание и DDQN работаю лучше обычного DQN и DQN с hard сглаживанием. Применяя soft сглаживание, алгоритм становиться более стабильным с более высокой наградой, более наглядно это видно на графике со скользящей средней наградой.