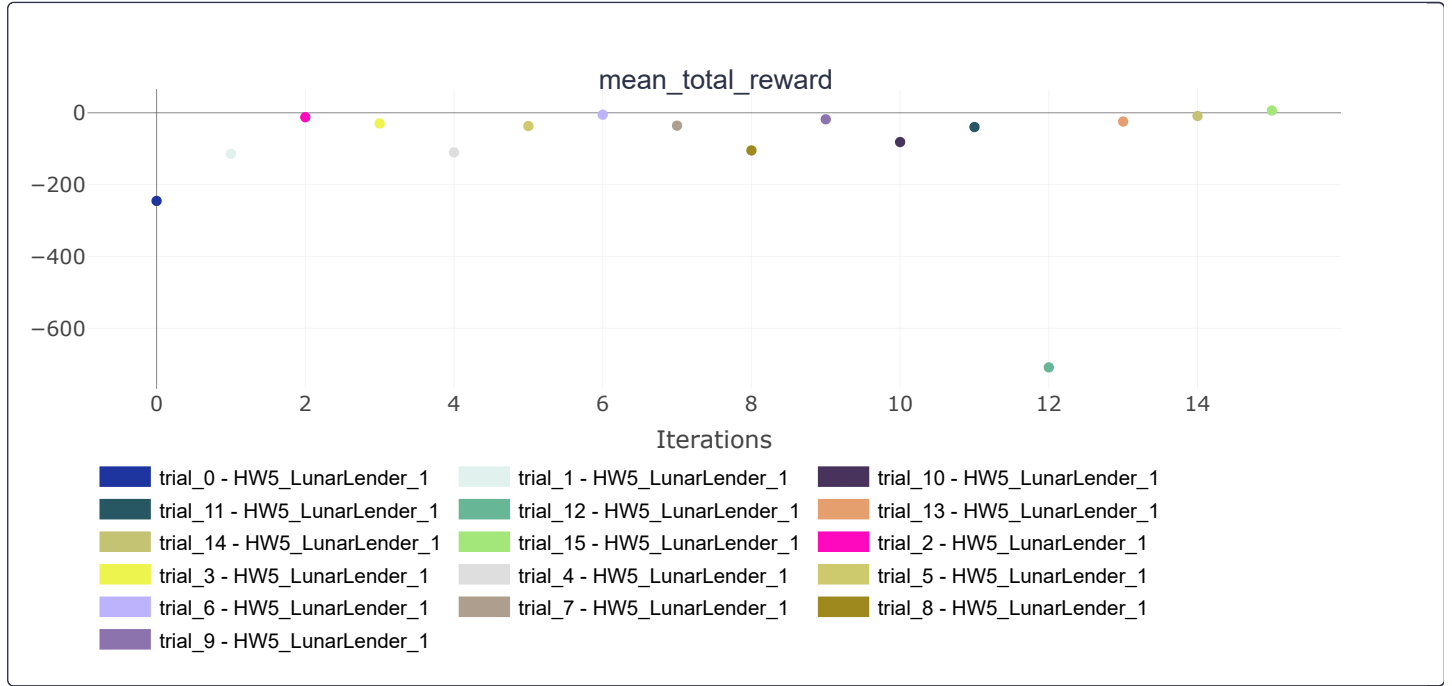


Домашнее задание 5

Репозиторий

1. Обучение агента DQN LunarLander, подбор гиперпараметров и сравнение с алгоритмом Deep Cross-Entropy

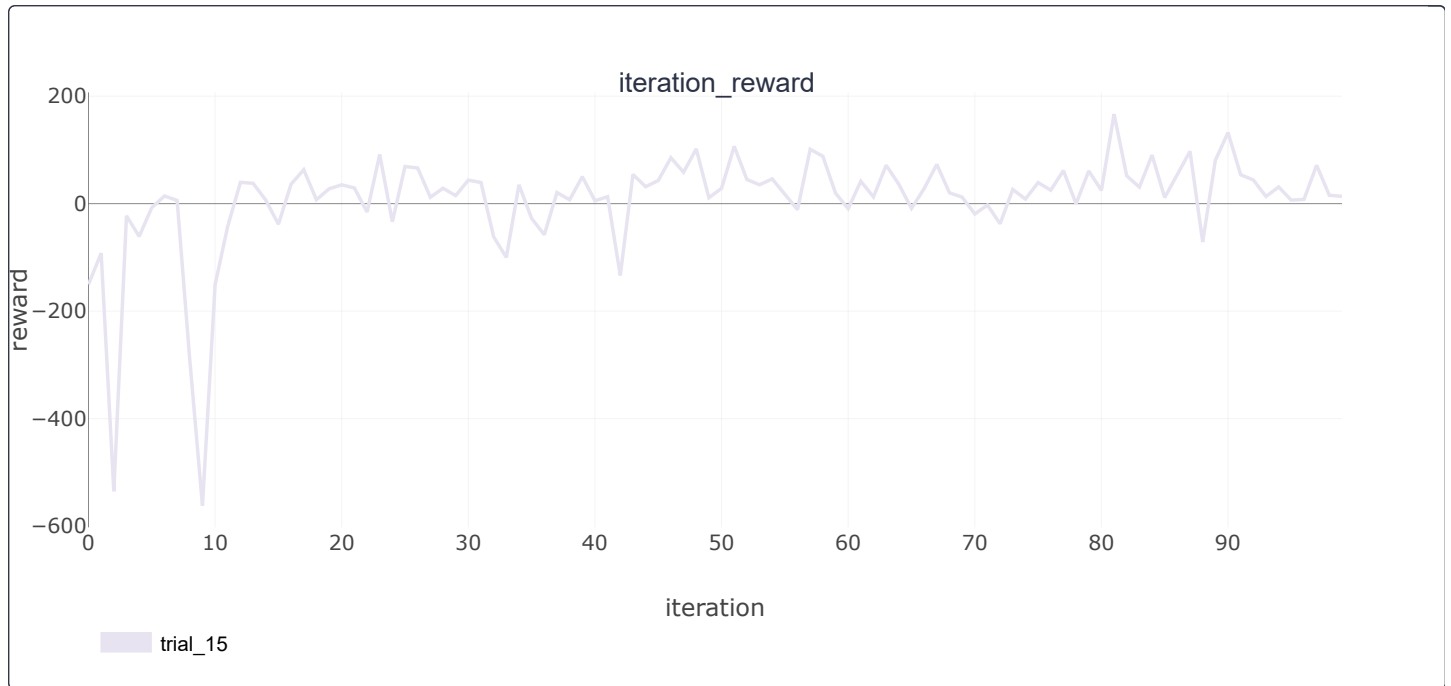
В данном задании был обучен алгоритм DQN на задаче LunarLander. Были подобраны гиперпараметры агента с помощью библиотеки Optuna. На графике ниже показаны 15 прогонов для подбора гиперпараметров. По оси x средний скор по 100 итерациям.



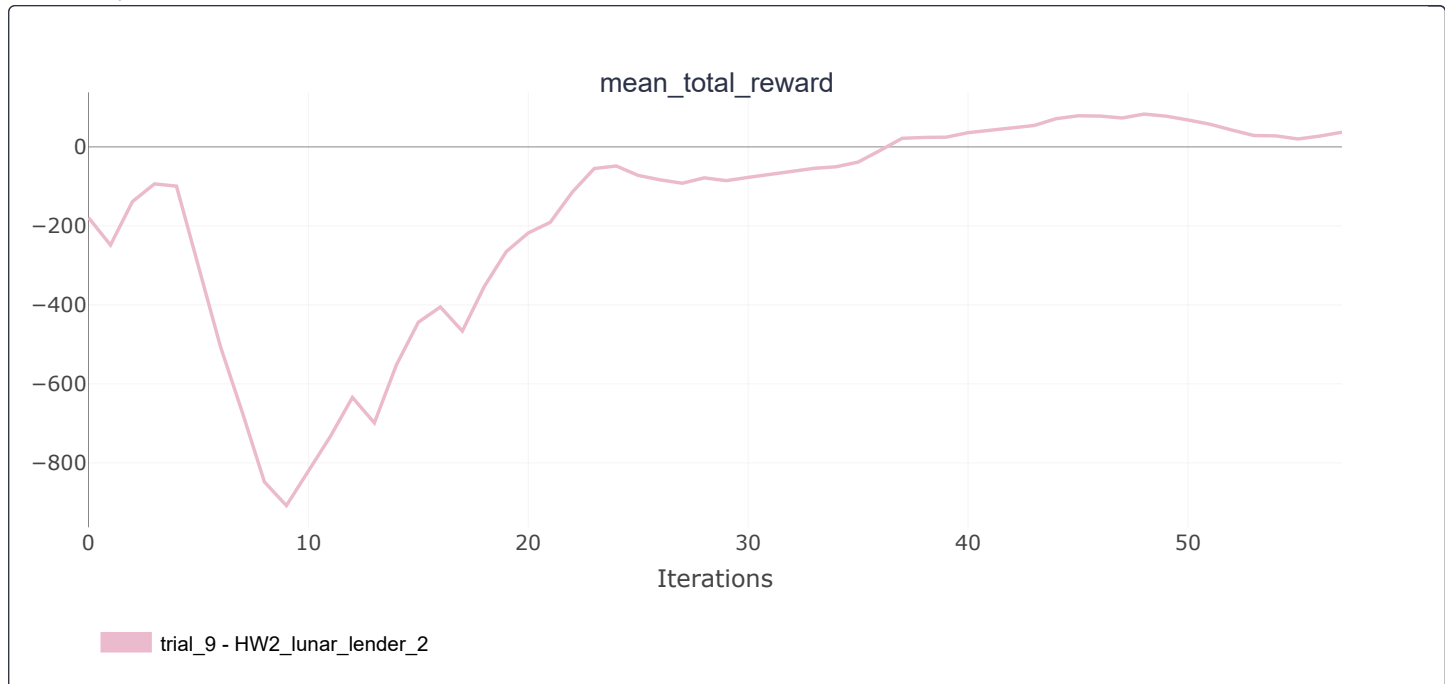
Лучшим оказался прогон 15 со следующими гиперпараметрами:

```
"batch_size": 42,
"epilon_min": 0.006916178956329845,
"epsilon_decrease": 0.006533780874234401,
"gamma": 0.9675863131457069,
"lr": 0.007383209469817978
```

Следующий график показывает обучение агента DQN с лучшими гиперпараметрами. По оси x эпизод, по оси y награда за эпизод.



Следующий график показывает обучение агента Deep Cross-Entropy с лучшими гиперпараметрами. По оси x эпизод, по оси y средняя сумма наград за эпизод.



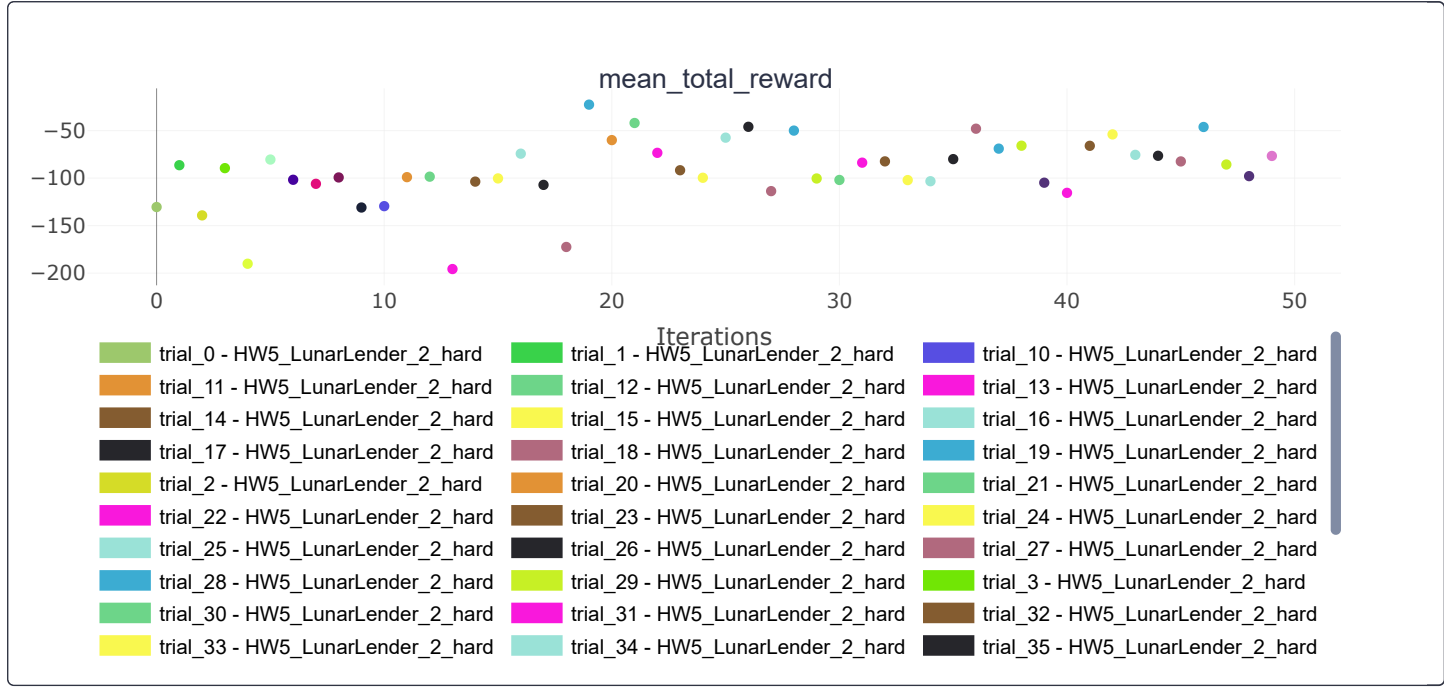
Как видно на графиках Deep Cross-Entropy алгоритм более гладкий (за счет усреднения), однако DQN показывает более хорошие значения наград по большим количествам эпох.

2. Сравнение DQN с Hard Target Update, DQN с Soft Target Update, Double DQN

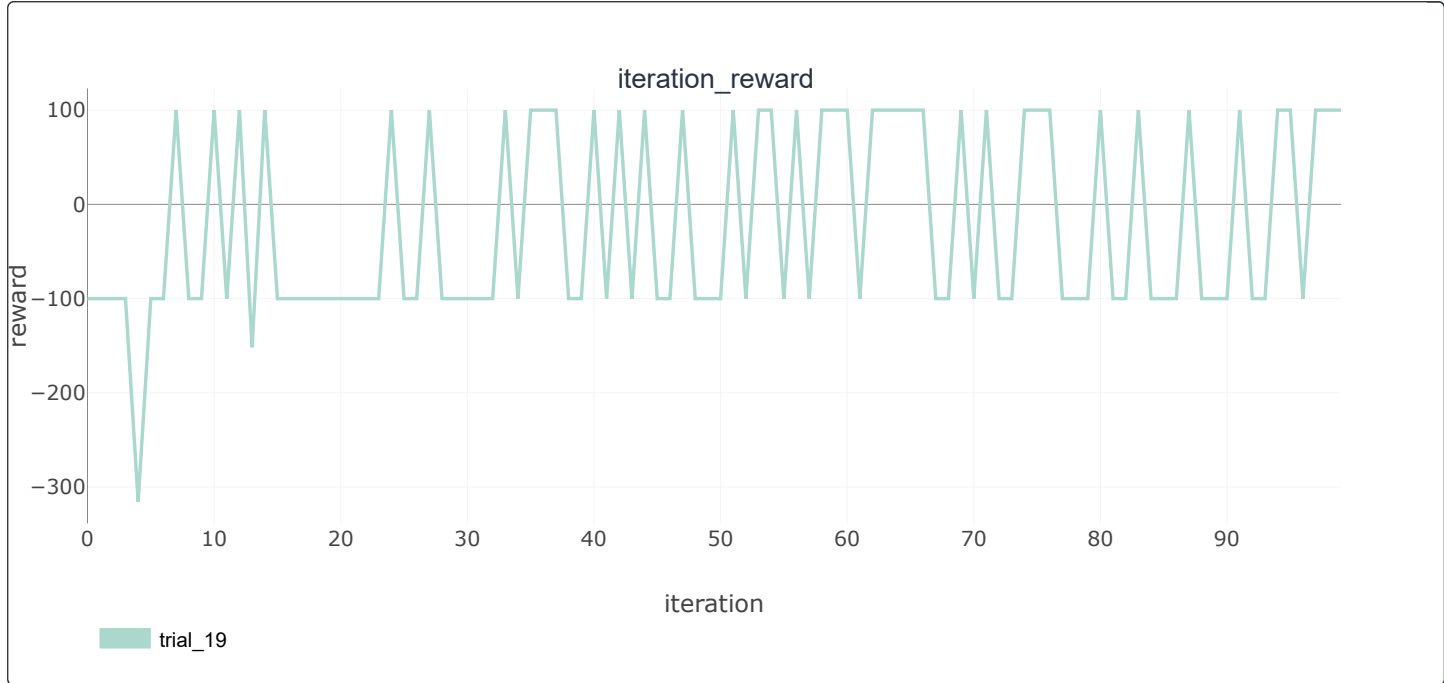
Ниже представлены графики с подбором гиперпараметров для DQN с Hard Target Update, DQN с Soft Target Update, Double DQN. Почему-то средняя награда по эпохам перестала подниматься больше 100.

DQN с Hard Target Update

Подбор гиперпараметров

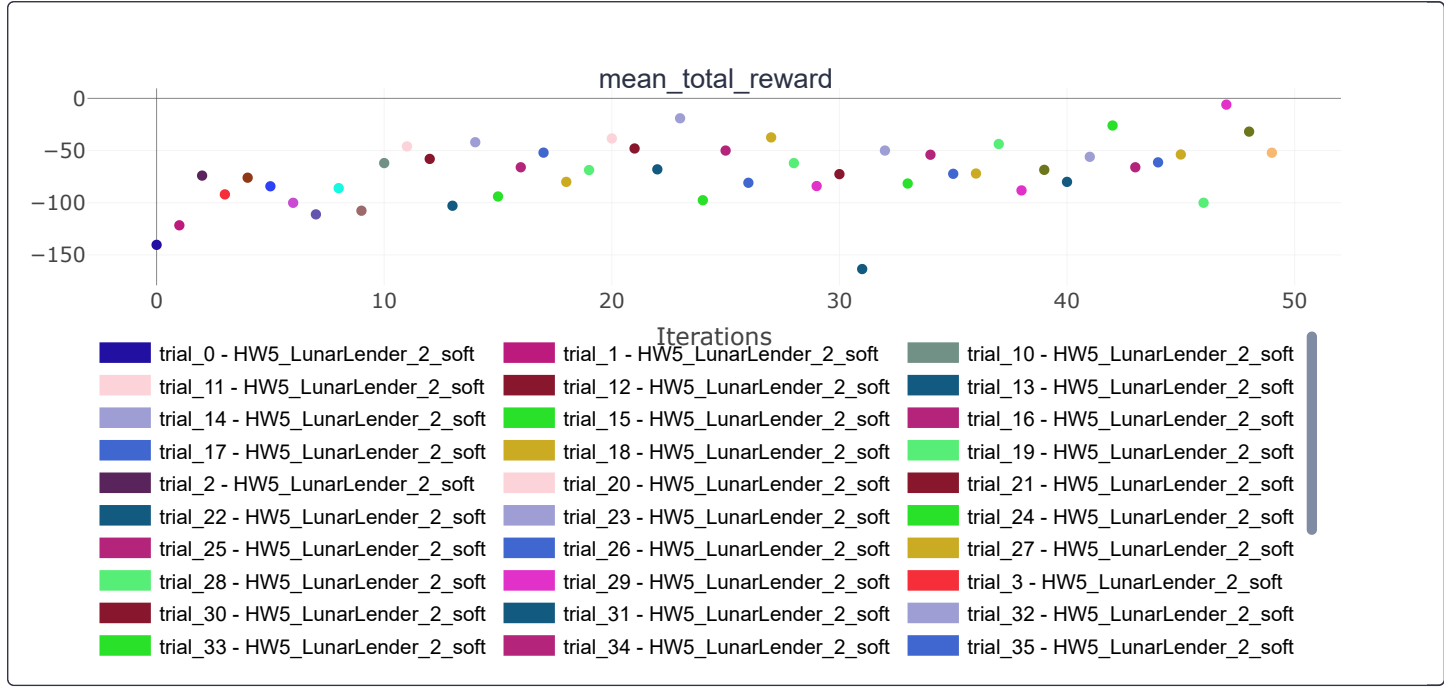


Лучшим оказался прогон 19 со средним -22

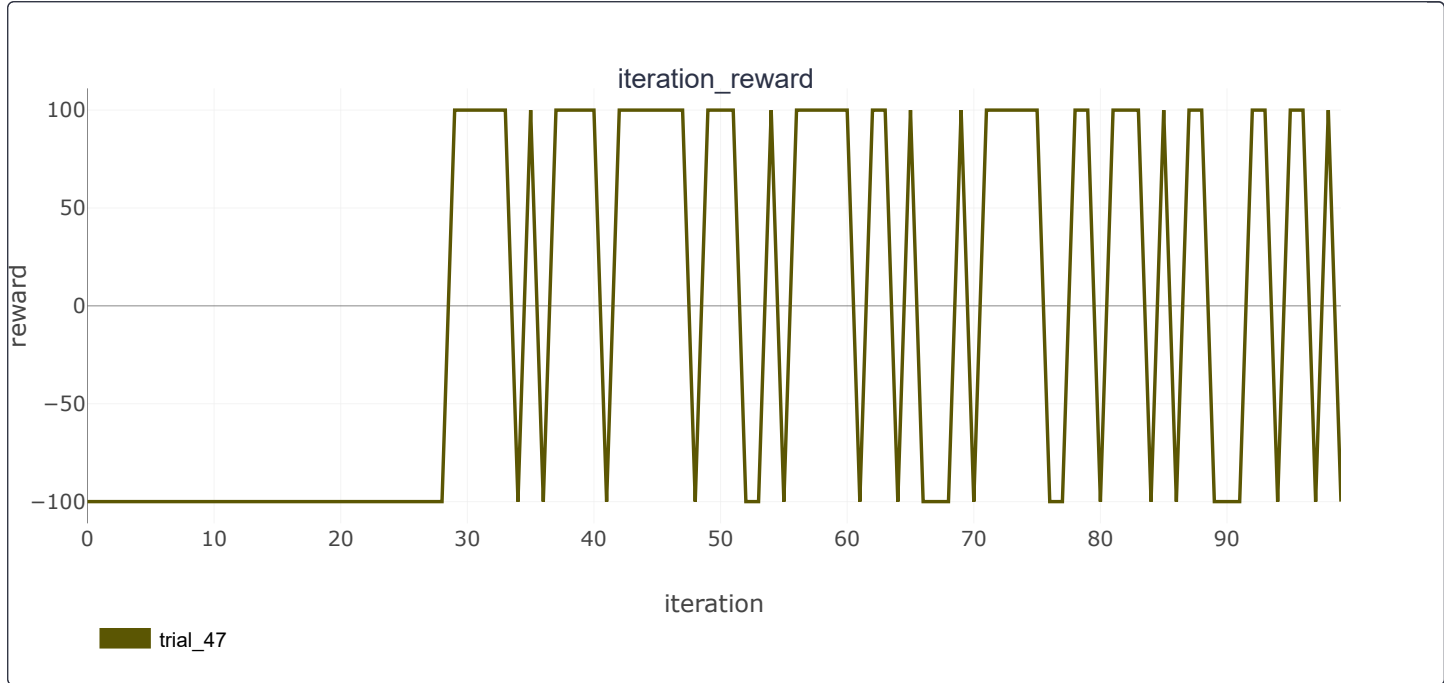


DQN c Soft Target Update

Подбор гиперпараметров

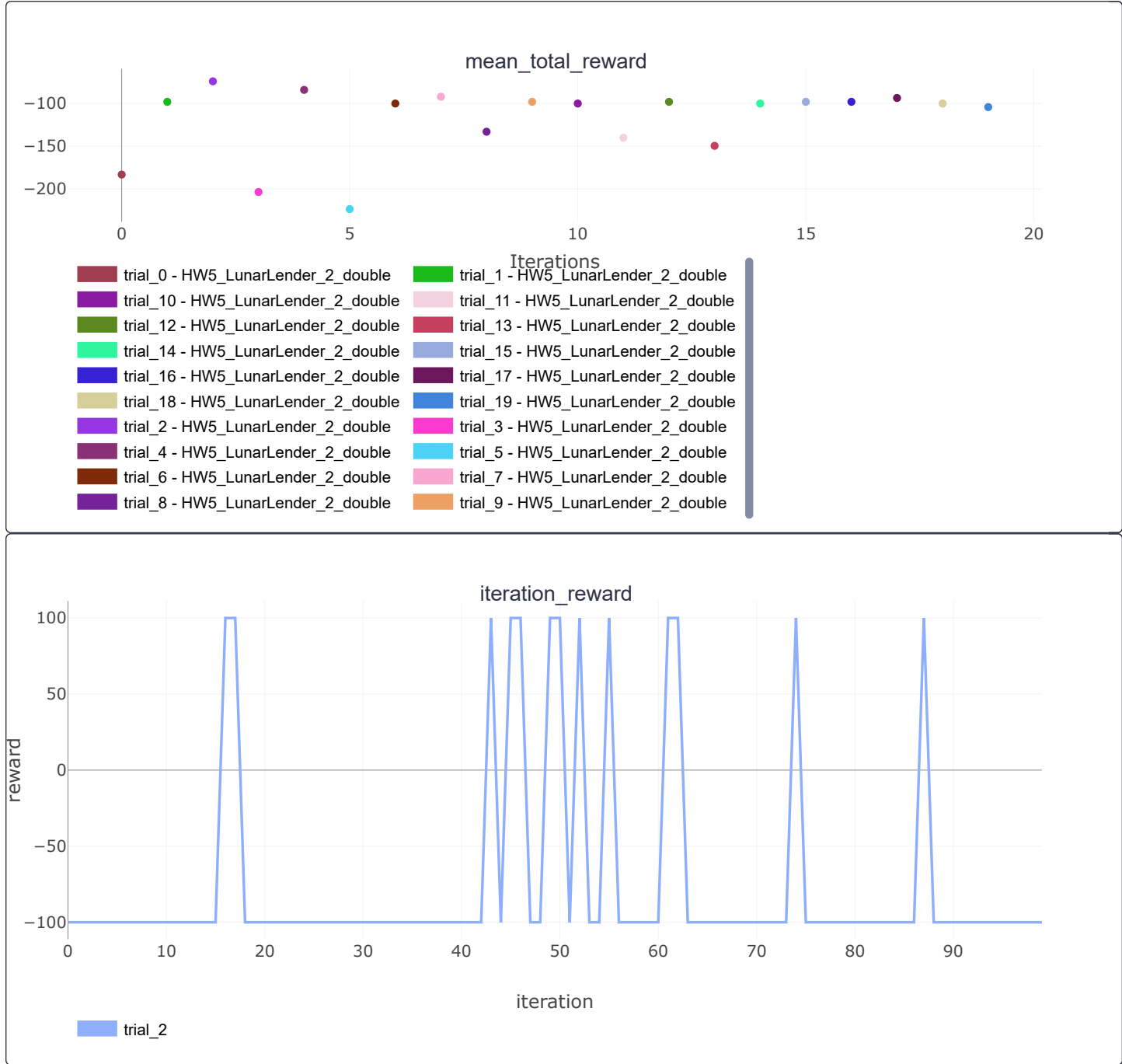


Лучшим оказался прогон 47 со средним -6



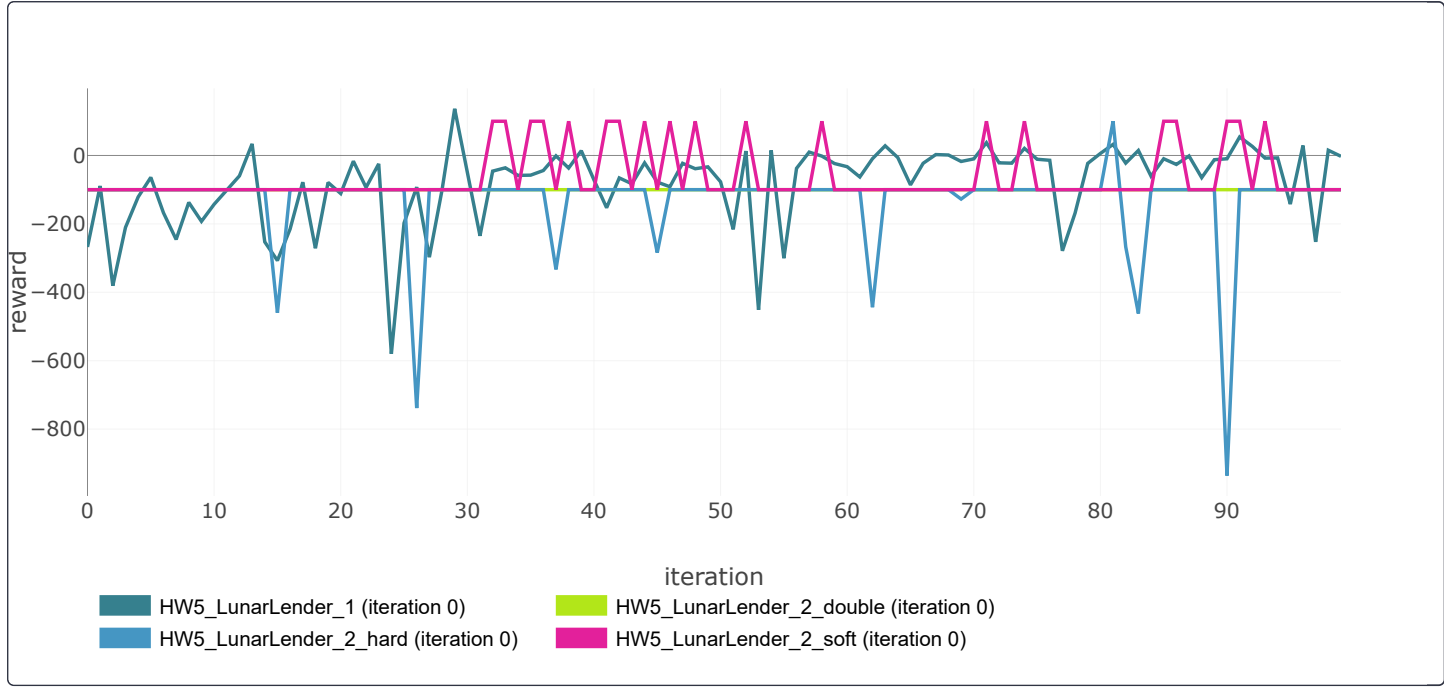
Double DQN

Подбор гиперпараметров



Сравнение

Далее представлен график сравнения алгоритмов



Можно сделать вывод, что алгоритм DQN с Soft Target Update является наиболее стабильным из всех представленных