

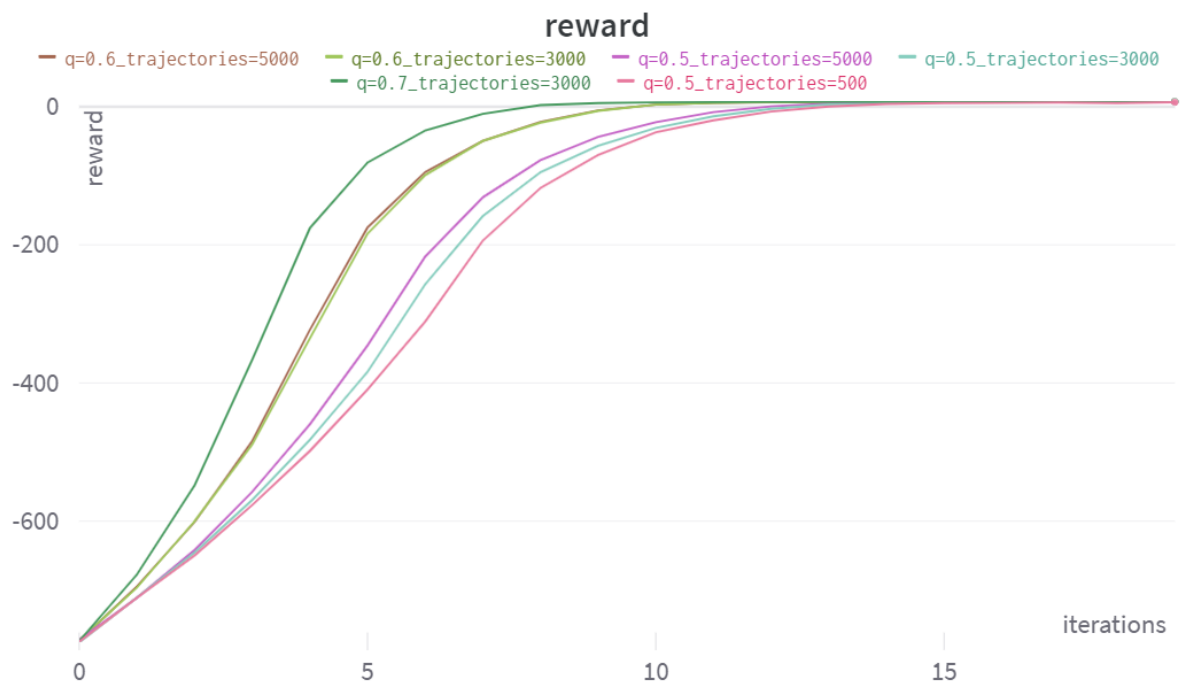
Задание 1

Были исследованы результаты со следующими гиперпараметрами:

- q : [0.5, 0.6, 0.7, 0.8, 0.9]
- количество траекторий: [100, 300, 500, 1000, 3000, 5000]
- количество итераций: 20

Количество действий в одной траектории = 1000, не стал подбирать другие значения, потому что уже после первых итераций обучения для решения задачи требуется меньше действий, и к 20 итерации алгоритм сходится

q_param	trajectories	reward ▼
0.6	5000	7.848
0.6	3000	7.68
0.5	5000	7.668
0.5	3000	7.625
0.7	3000	7.461



Наилучший reward у алгоритма с параметрами $q=0.6$ и $trajectories = 5000$, также на графике видно, что быстрее всех сходится алгоритм с $q=0.7$, т.е с повышением “качества” выбранных траекторий скорость сходимости алгоритма возрастает

Задание 2

2.1 Laplace smoothing

Были исследованы результаты со следующими гиперпараметрами:

- q : [0.6, 0.7, 0.8, 0.9]
- количество траекторий: [1000, 5000]
- λ : [0.01, 0.05, 0.1, 0.3, 0.7, 0.9]
- количество итераций: 50

λ	q_{para}	trajectories	reward ▼
0.01	0.6	5000	7.785
0.05	0.6	5000	7.566
0.1	0.6	5000	7.4
0.3	0.6	5000	6.891
0.7	0.6	5000	6.319





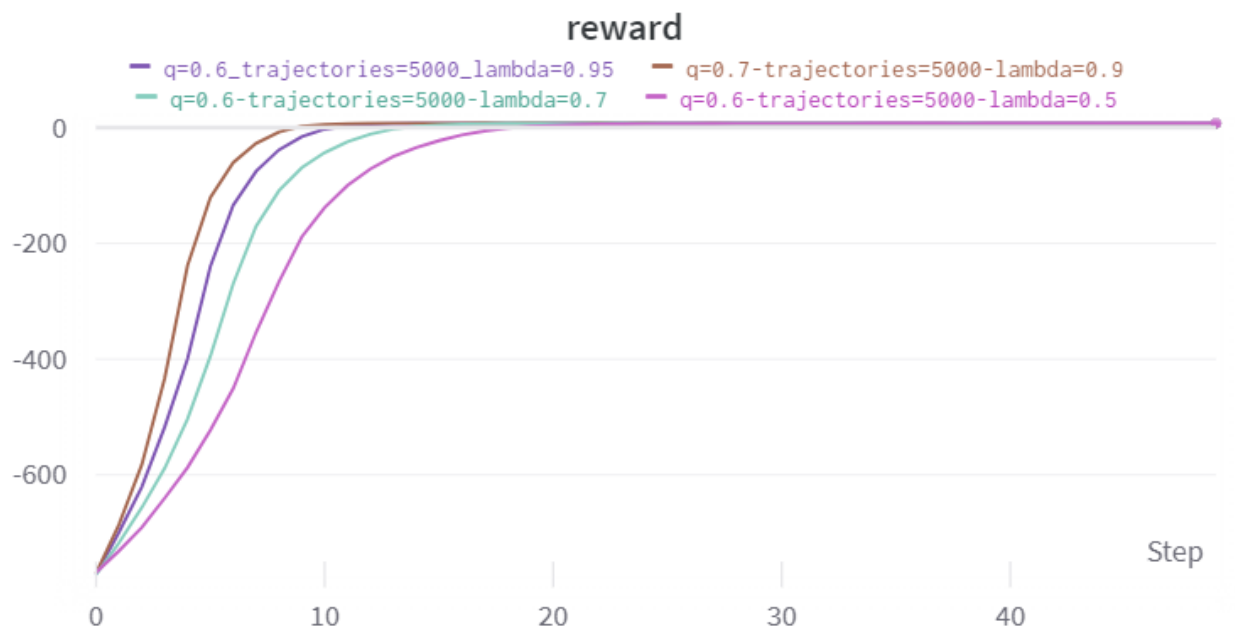
Лучшие результаты были получены с параметрами $q = 0.6$, $\text{trajectories} = 5000$, $\lambda = 0.01$, при уменьшении λ растёт итоговый reward, но при этом скорость обучения не сильно.

2.2 Police smoothing

Были исследованы результаты со следующими гиперпараметрами:

- q : [0.6, 0.7, 0.8, 0.9]
- количество траекторий: [1000, 5000]
- λ : [0.5, 0.6, 0.7, 0.8, 0.9, 0.95]
- количество итераций: 50

lambda	q_param	trajectories	reward ▼
0.95	0.6	5000	7.885
0.9	0.6	5000	7.851
0.9	0.7	5000	7.824
0.7	0.6	5000	7.81
0.5	0.6	5000	7.798



Лучше всего себя показал алгоритм с параметрами $q = 0.6$, $trajectories = 5000$, $lambda = 0.95$, немного превзойдя алгоритм без сглаживания с теми же параметрами q и $trajectories$. Быстрее всего сходятся алгоритмы с более высокой отсечкой по q и $lambda$ близкой к единице. Небольшое значение $lambda$ замедляет обучение, так как модель больше ориентируется на значения с предыдущего действия.