Домашняя работа #1 RL

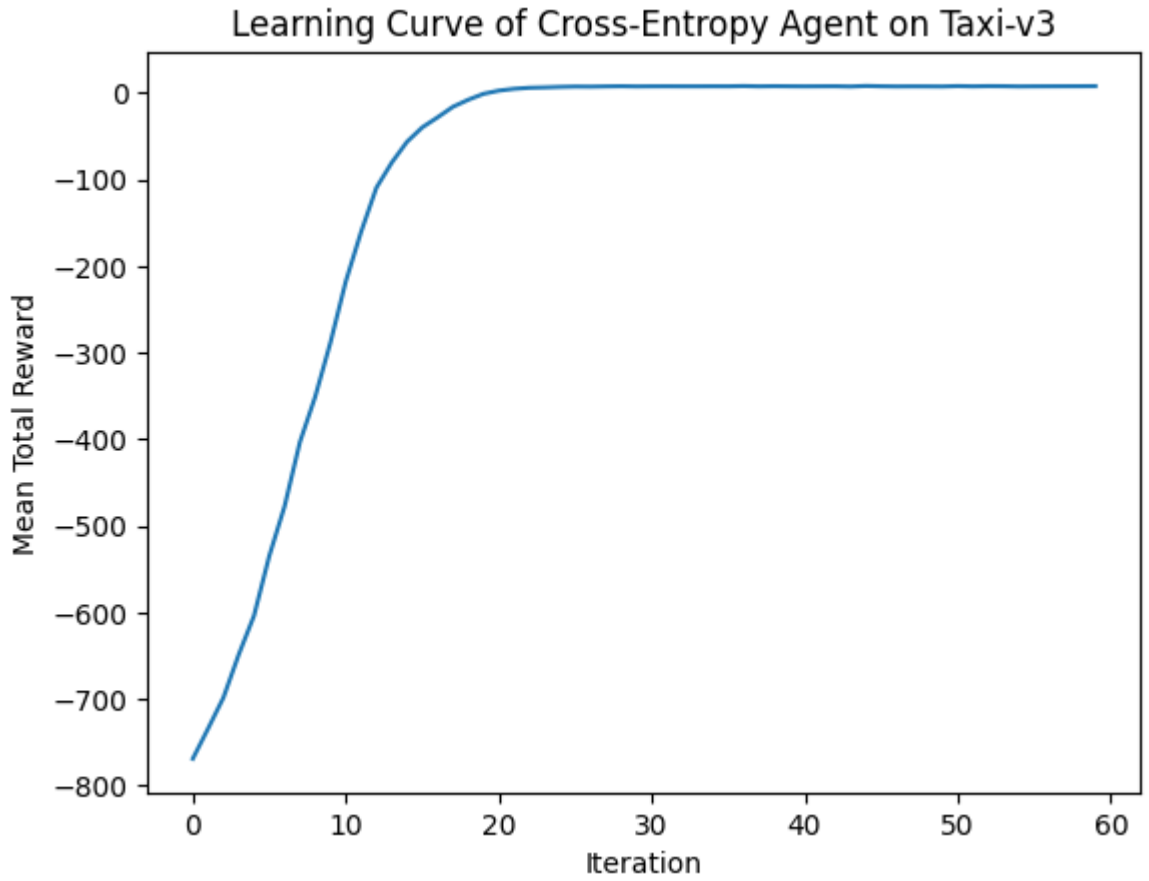
Введение

Первое задание

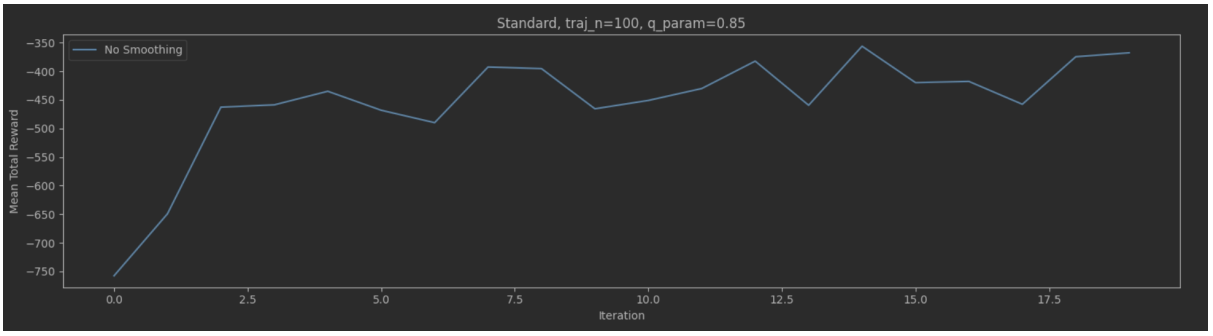
Обучение

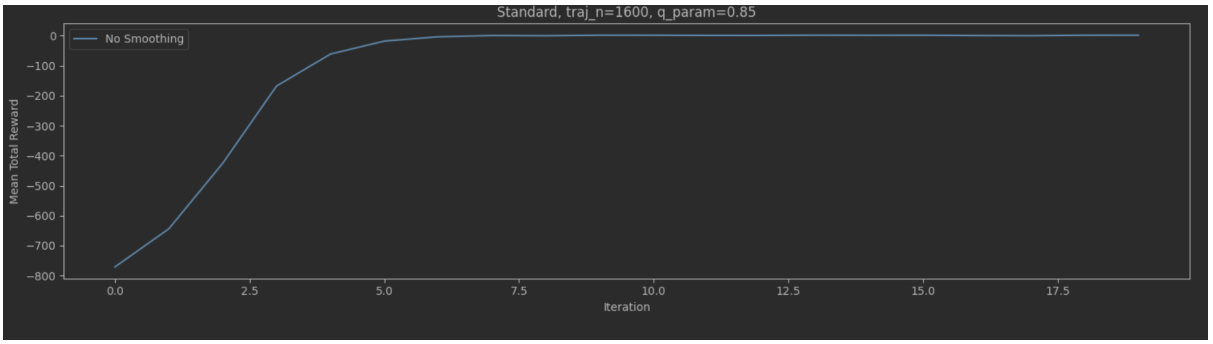
Эксперимент 1

Для начала я просто запустил модель с квантилем 0.3, и получил очень неплохой результат. Скорее всего он был рандомным и я решил поиграться.



Вот некоторые из графиков:





Примечательно то, что чем больше траекторий тем более сглаженней и стабильнее график. высокие параметры q проигрывали при меньшем количестве траекторий, но с ростом оных все уравнивалось(больше графиков во втором ipynb ноутбуке)

Вывод

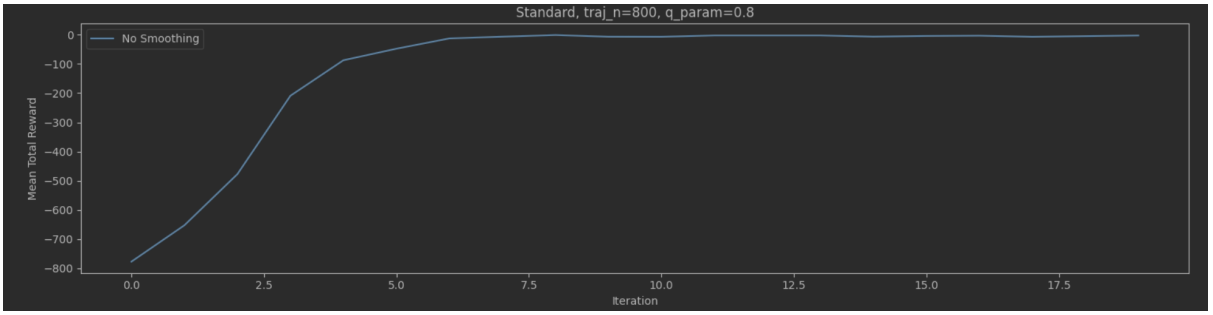
Так как среда, алгоритм, и проделанные действия достаточно простые, то все обучилось довольно быстро, и каких-то особых выводов за исключением уже названных нет.

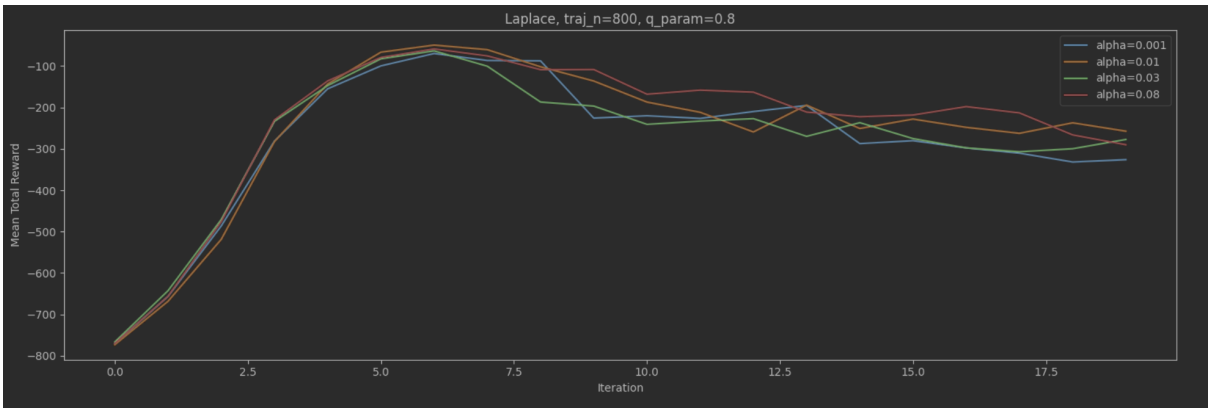
Второе задание

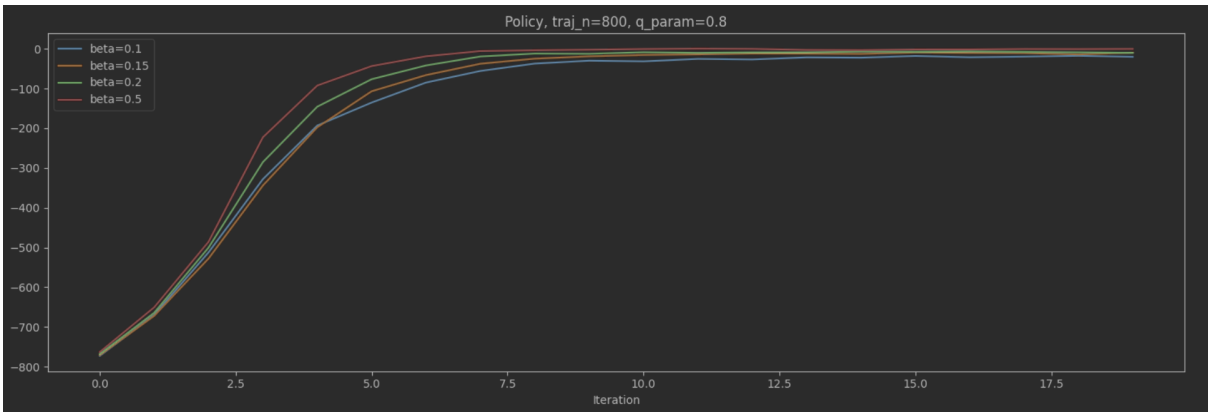
Обучение

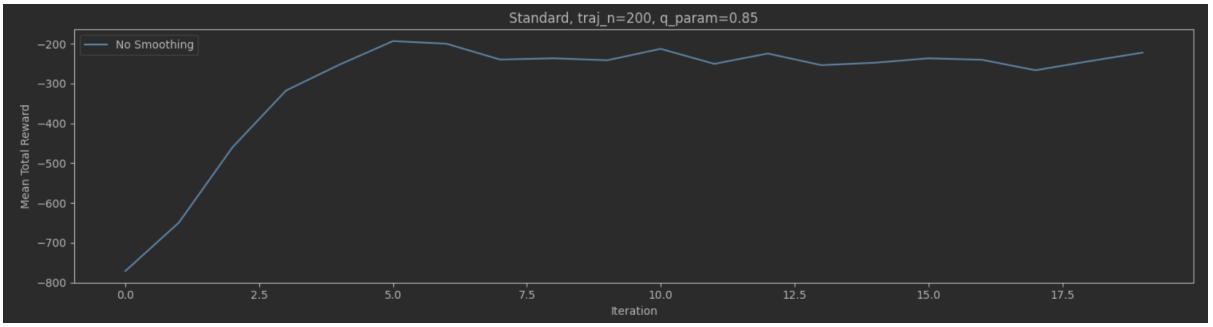
Эксперимент 1

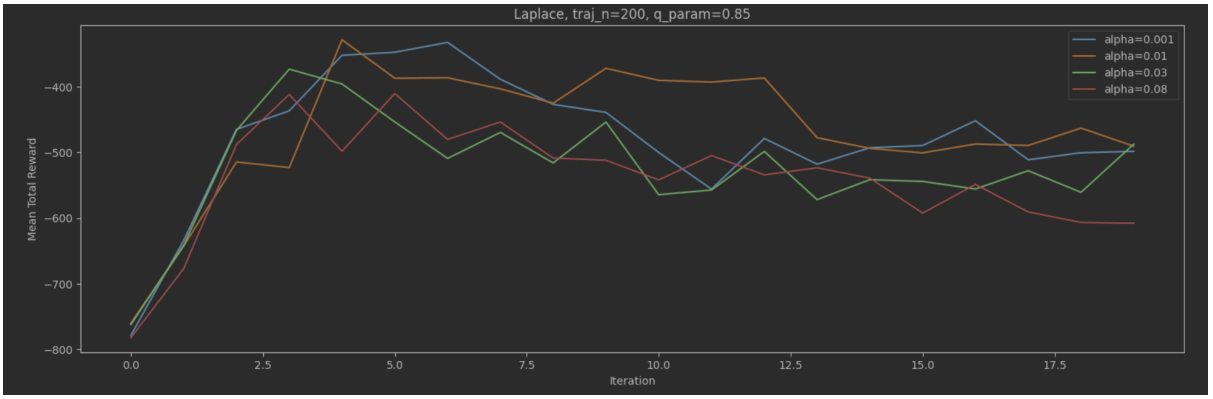
Во втором задании необходимо было применить два способа сглаживания и сравнить с обычным алгоритмом. Ниже графики(в ноутбуке больше).

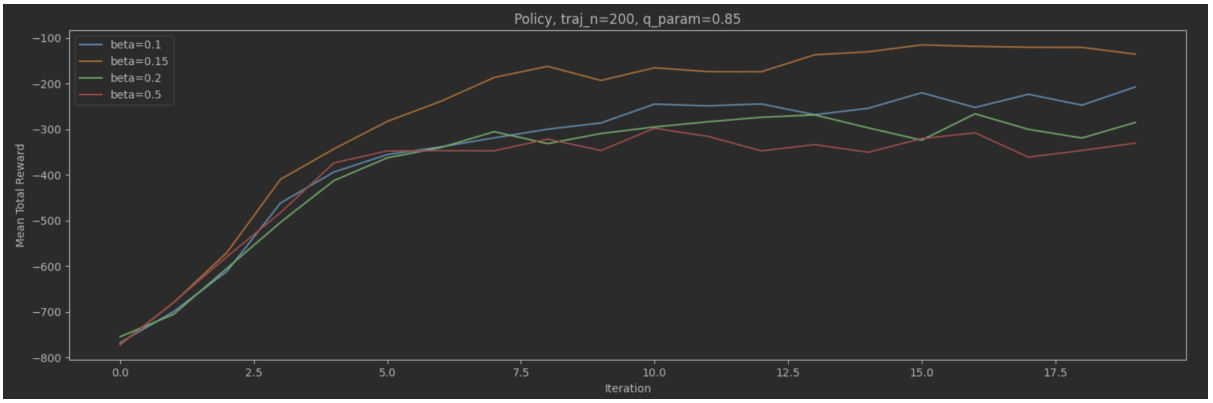

Standard, traj_n=800, q_param=0.8
No Smoothing
Mean Total Reward
Iteration


Laplace, traj_n=800, q_param=0.8
alpha=0.001
alpha=0.01
alpha=0.03
alpha=0.08
Mean Total Reward
Iteration


Policy, traj_n=800, q_param=0.8
beta=0.1
beta=0.15
beta=0.2
beta=0.5
Iteration


Standard, traj_n=200, q_param=0.85
No Smoothing
Mean Total Reward
Iteration





Опять таки в ipynb ноутбуке значительно больше графиков. Что стало понятно. При больших количествах итераций обычный и policy способы сравниваются по точности. Бета почти не влияет. А вот если говорить о маленьких количествах траекторий то тут картина иная, и модели ведут себя по разному. Очень странно поведение Laplace. Он изначально добивается неплохой точности, но потом происходит падение награды. В целом исходя из этой лабораторной не совсем понятно насколько хороший эффект оказывают policy и laplace. Возможно с другой средой они бы повели себя по другому.

Вывод

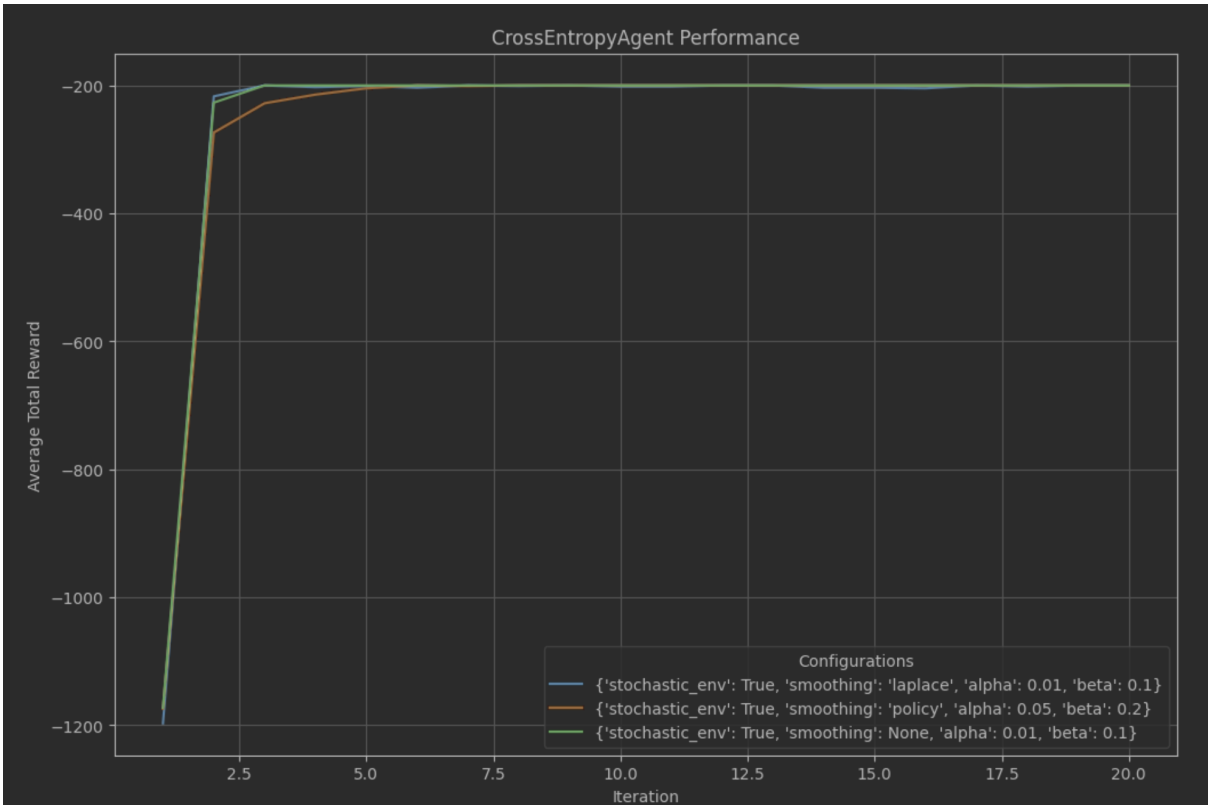
Исходя из этой лабораторной не совсем понятно как policy и laplace влияют. Единственное что видно это то, что laplace почему-то теряет точность. То ли попадает в локальный оптимум то ли еще что-то.

Третье задание

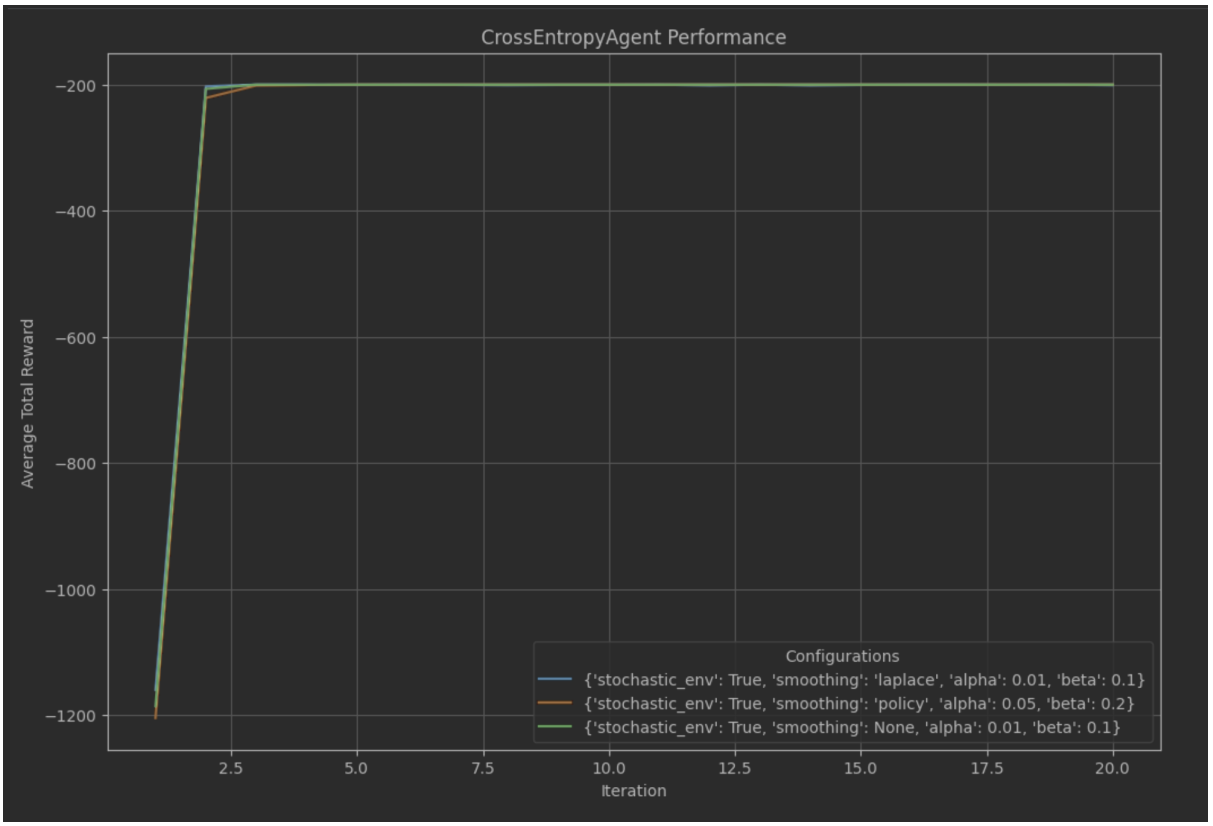
Обучение

Эксперимент 1

В третьем опциональном задании надо было написать комбинировать стохастические и детерменированные политики. Вот что получилось:



Не совсем понял, что здесь произошло, и почему так, но решил увеличить m в несколько раз и посмотреть что выйдет



Результат тот же самый модел застряла.

Вывод

Каких-то особых выводов нет, кроме того, что можно по разному прокидывать траектории и политики в модель, и экспериментировать с этим.