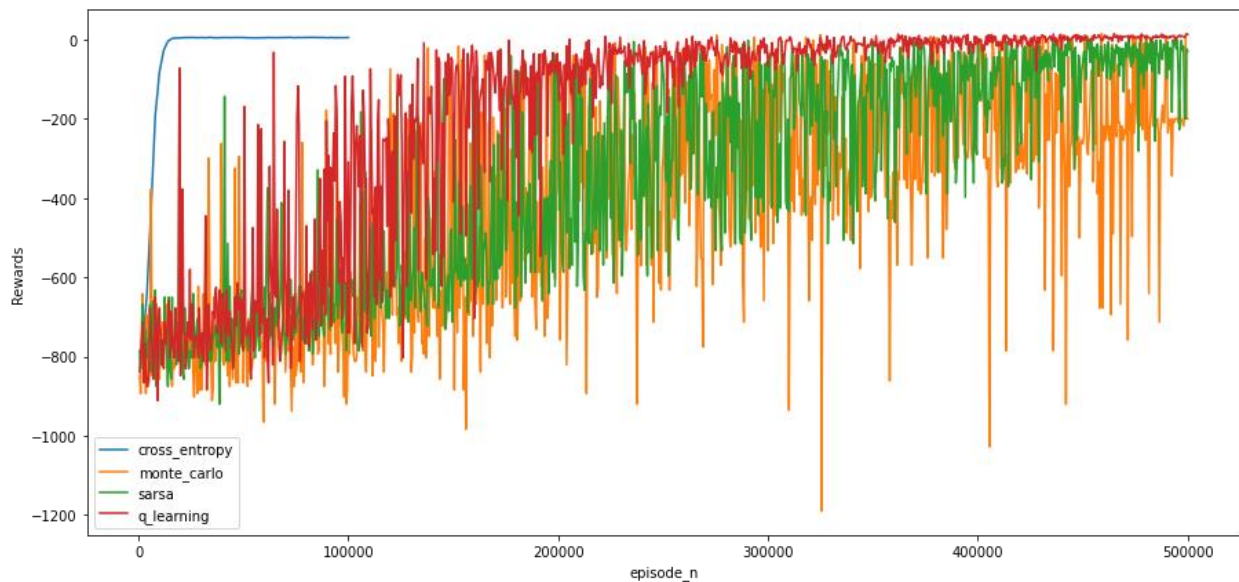


## Задание 1. Реализовать Q-Learning и сравнить его результаты с реализованными ранее алгоритмами

Эксперимент: Награда по количеству итераций для разных алгоритмов в задаче Taxi-v3.

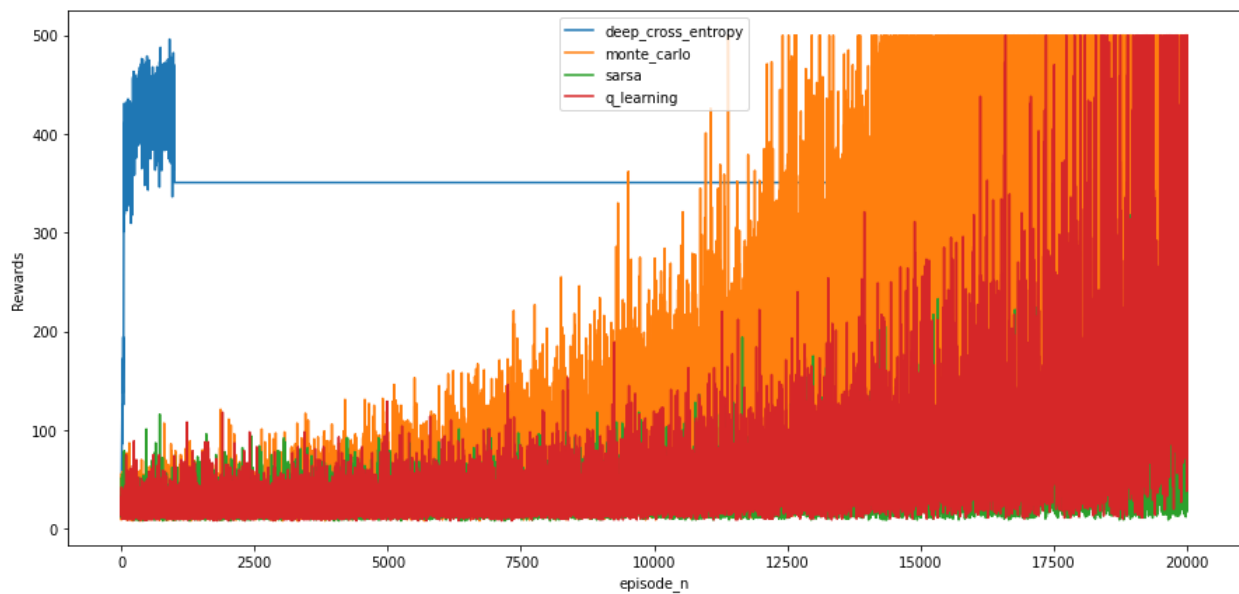


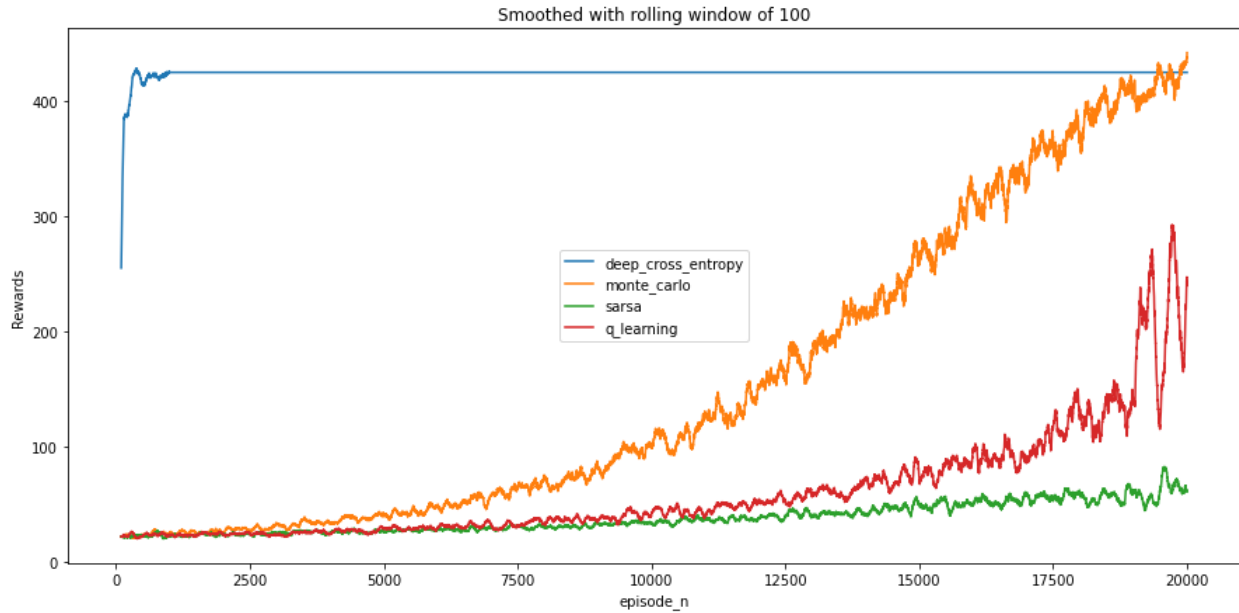
Вывод.

Из всех алгоритмов кросс энтропия сходится быстрее всего, но например q-learning в процессе обучения доходит до более высокой награды.

## Задание 2. Дискретизировать пространство состояний

Эксперимент: Награда по количеству итераций для разных алгоритмов в задаче Taxi-v3.



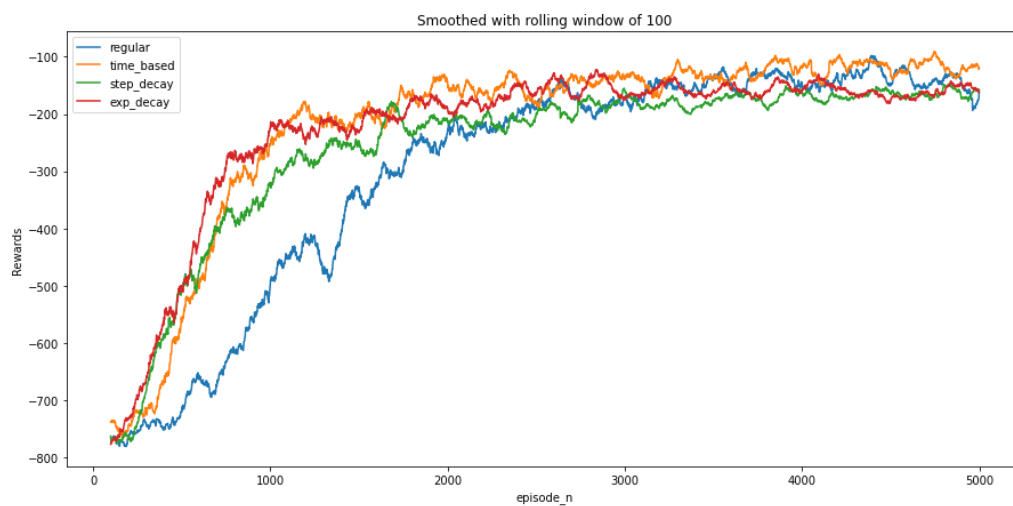


**Вывод.**

Для наглядности я сгладил графики наград, по-прежнему кросс энтропия сходится гораздо быстрее, но может уступать по качеству другим алгоритмам. SARSA обучается крайне медленно, возможно из-за метода дискретизации который я выбрал, меньшее количество бинов в теории уменьшило бы Q матрицу и могло бы быстрее учиться.

**Задание 3.** Придумать стратегию для выбора epsilon в задаче Taxi-v3

**Эксперимент:** Награда по количеству итераций для разных алгоритмов уменьшения epsilon.



### *Вывод.*

Я использовал 3 классических метода снижения learning rate из DL: временной, ступенчатый и экспоненциальный. Данные методы требуют некоторой предварительной настройки параметров, которая влияет на качество схождения. Почти все они сходятся примерно к одинаковым наградам в итоге, но делают это немного быстрее чем линейный метод. Временной метод кажется работает немного лучше остальных.