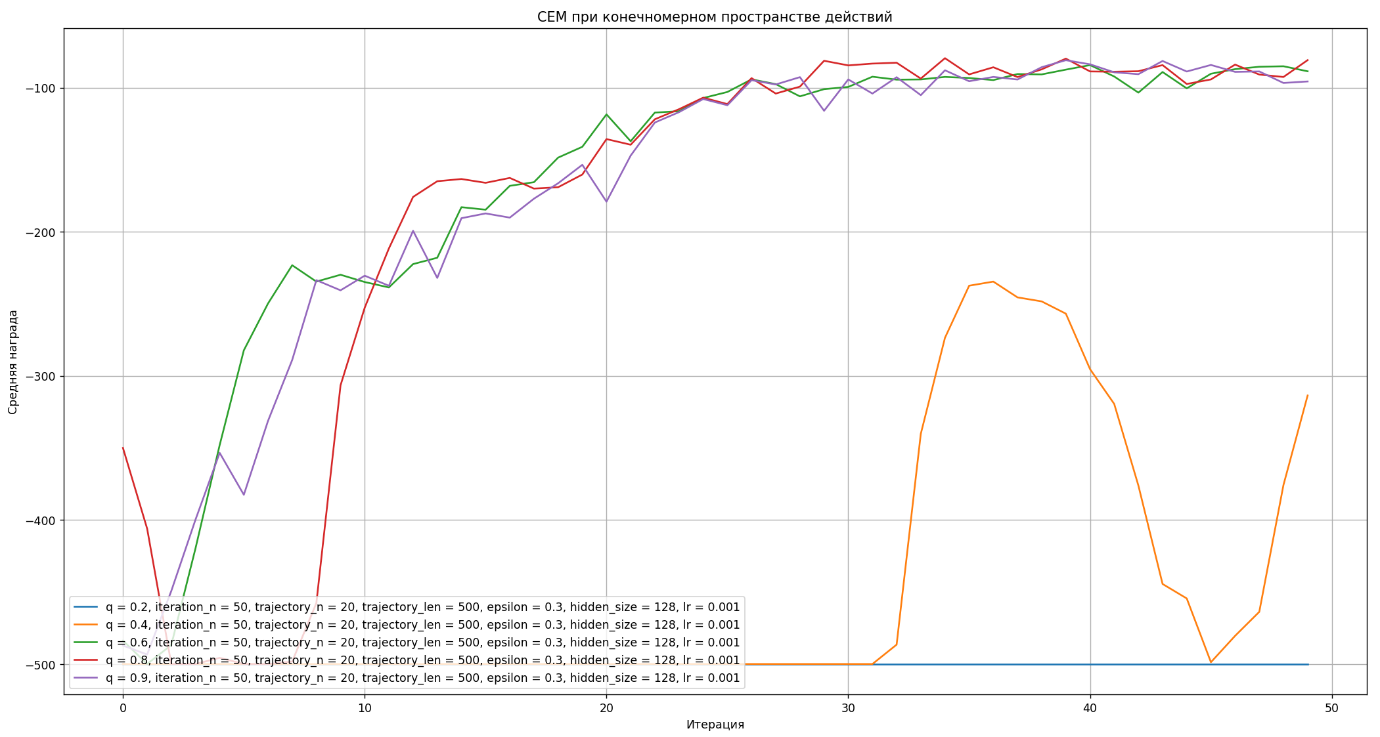
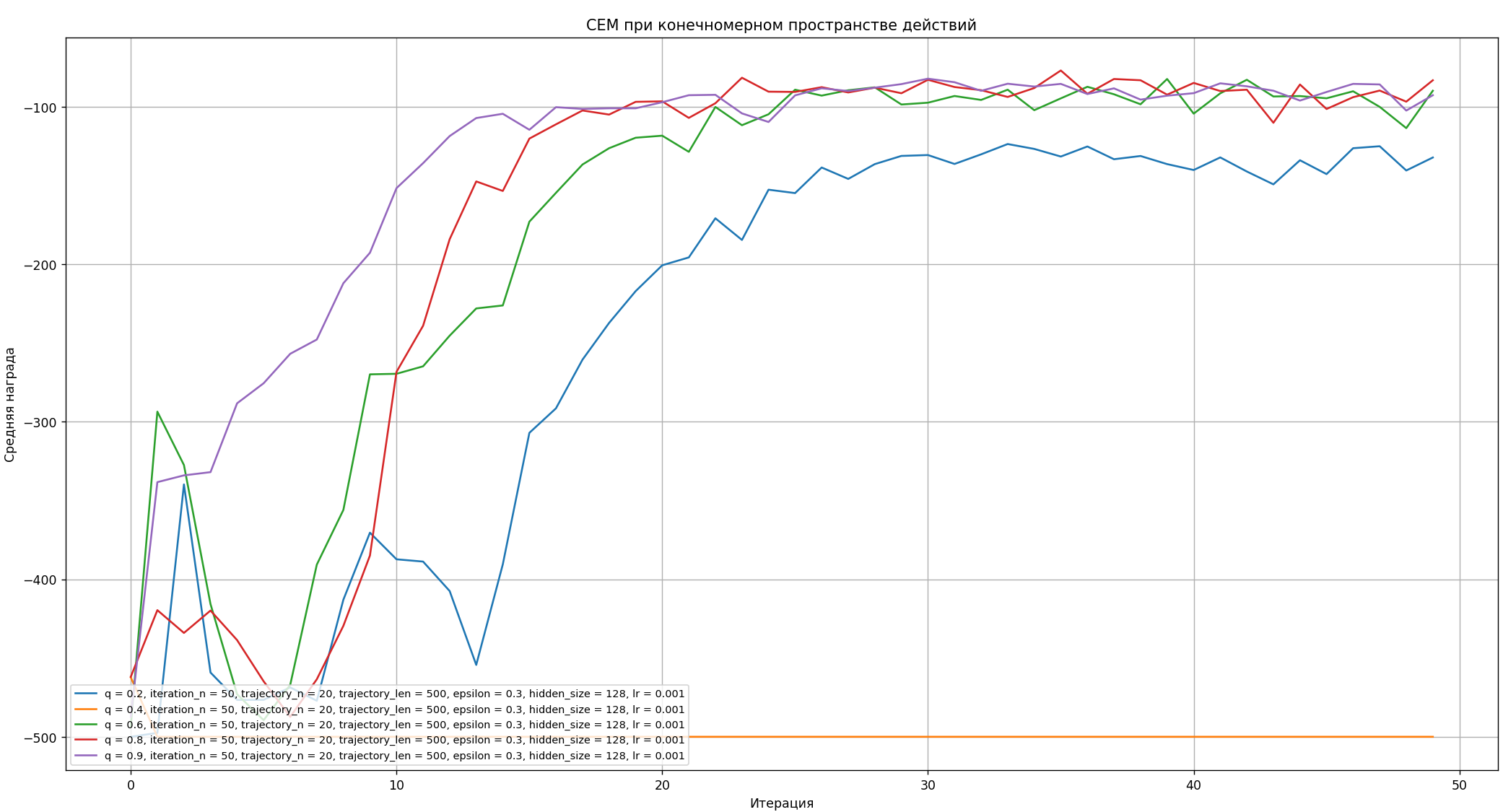
**Трапер Максим. СПБГУ. Магистратура 1 курс «Искусственный интеллект и наука о данных»**

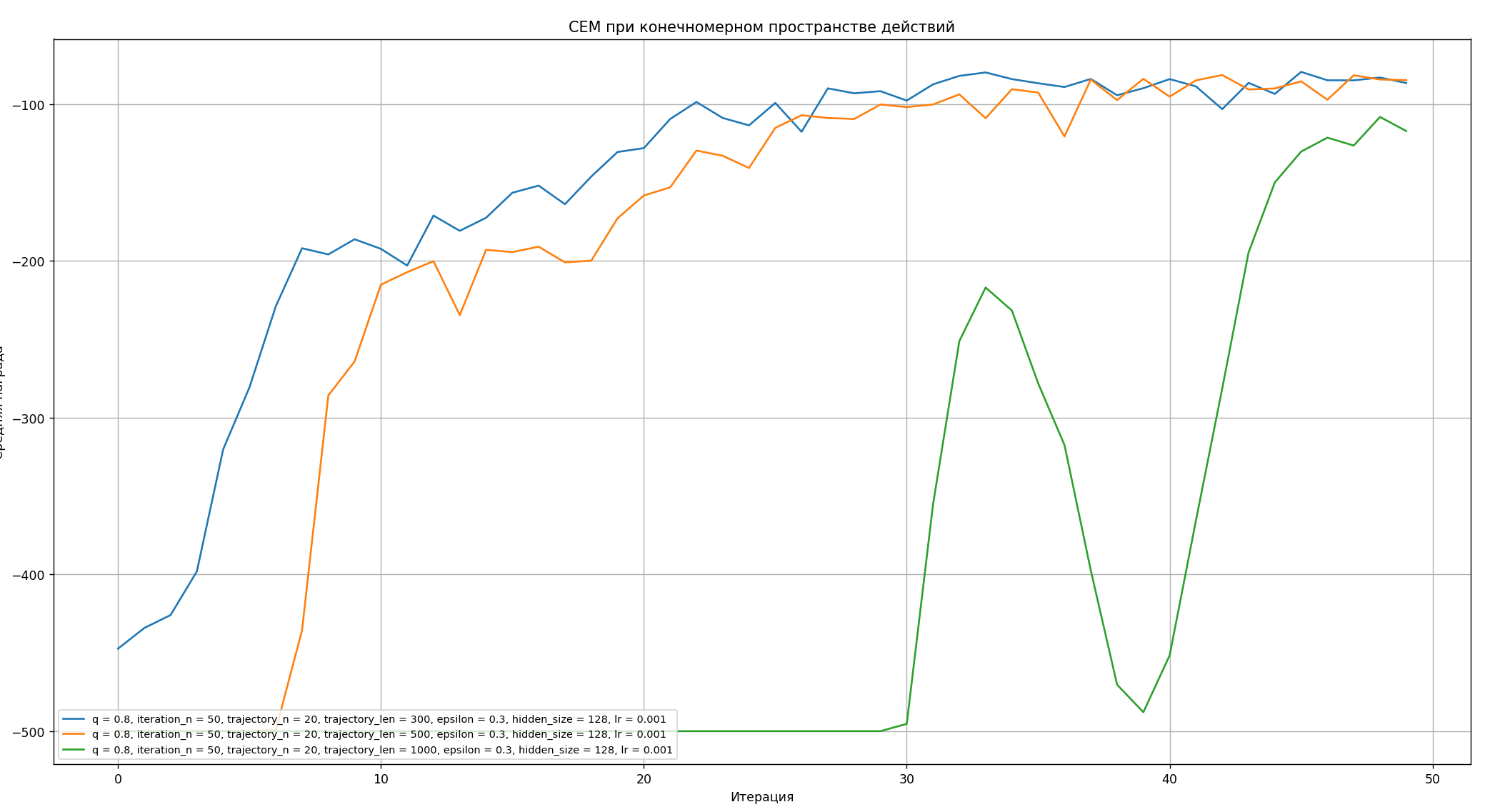
**Задание №1:** пользуясь алгоритмом Кросс-Энтропии для конечного пространства действий обучить агента решать Acrobot-v1. Исследовать гиперпараметры алгоритма и выбрать лучшие.

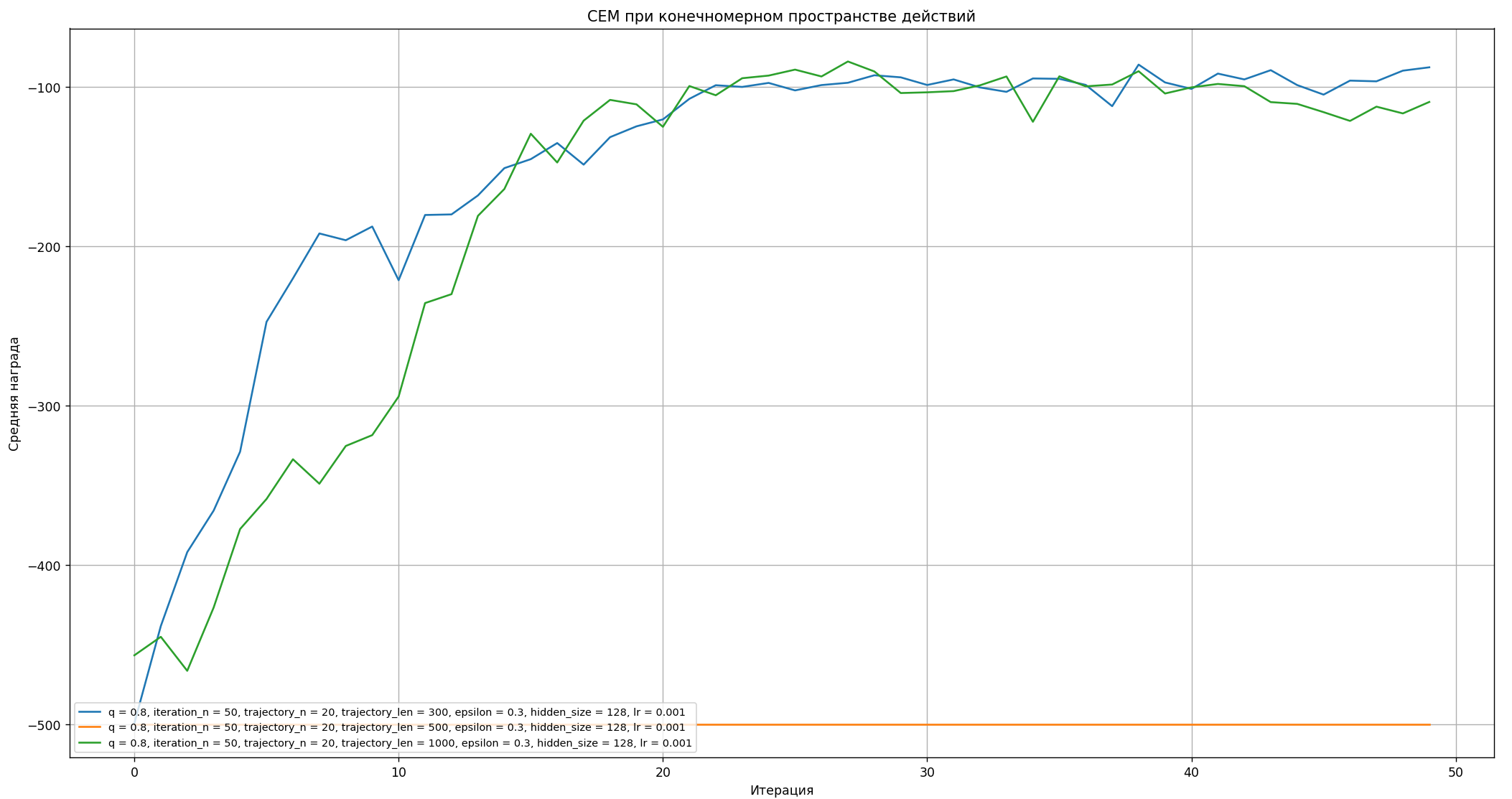
**Эксперимент №1:** попробуем перебрать параметр q. Набор для перебора = [0.2, 0.4, 0.6, 0. 8, 0.9].

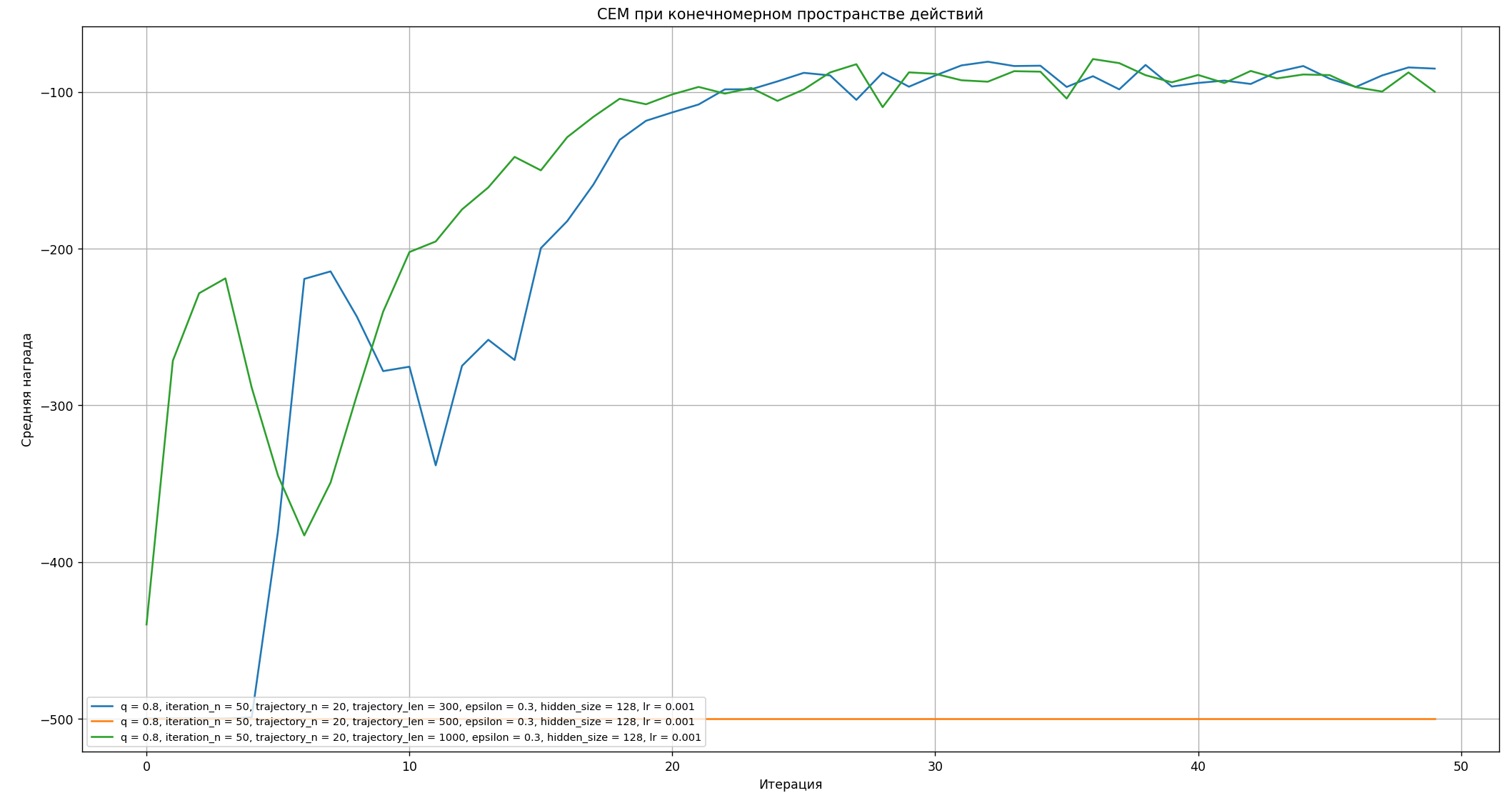
Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, длина траекторий = 500, эпсилон = 0.3, размер скрытого слоя = 128, lr = 1e-3 = 1e

**Вывод:** исходя из результатов, высокие значения q (q > 0.5) показывают в среднем лучшие результаты по качеству и скорости сходимости. За оба раза лучше примерно одинаково и лучше всего показали себя значения q: 0.6, 0.8, 0.9. Допустим, возьмём 0.8 как медиану из этих значений для дальнейших экспериментов.

**Эксперимент №2:** попробуем перебрать длину траекторий. Набор для перебора = [300, 500, 1000].

Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 128, lr = 1e-3, q = 0.8

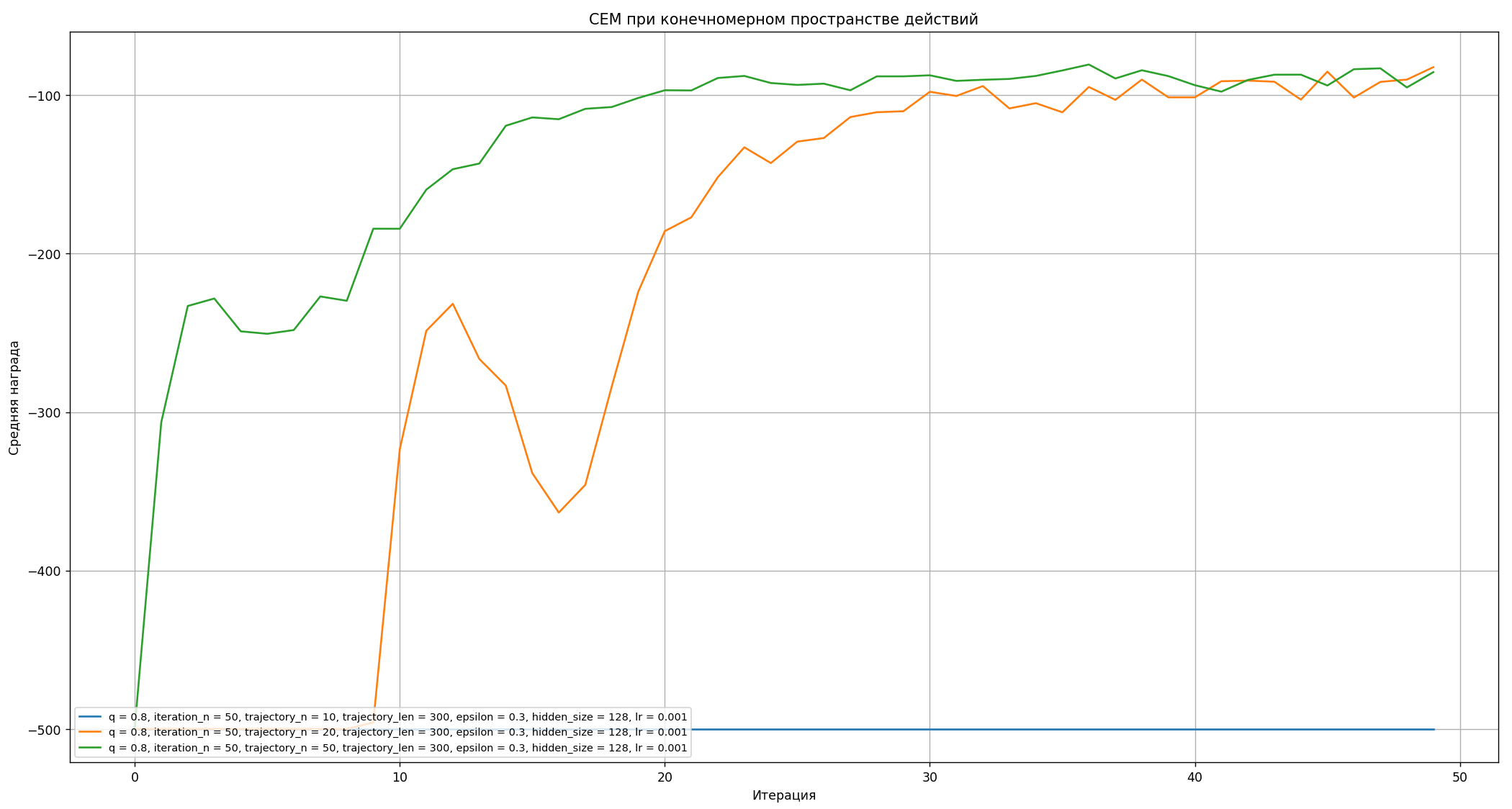


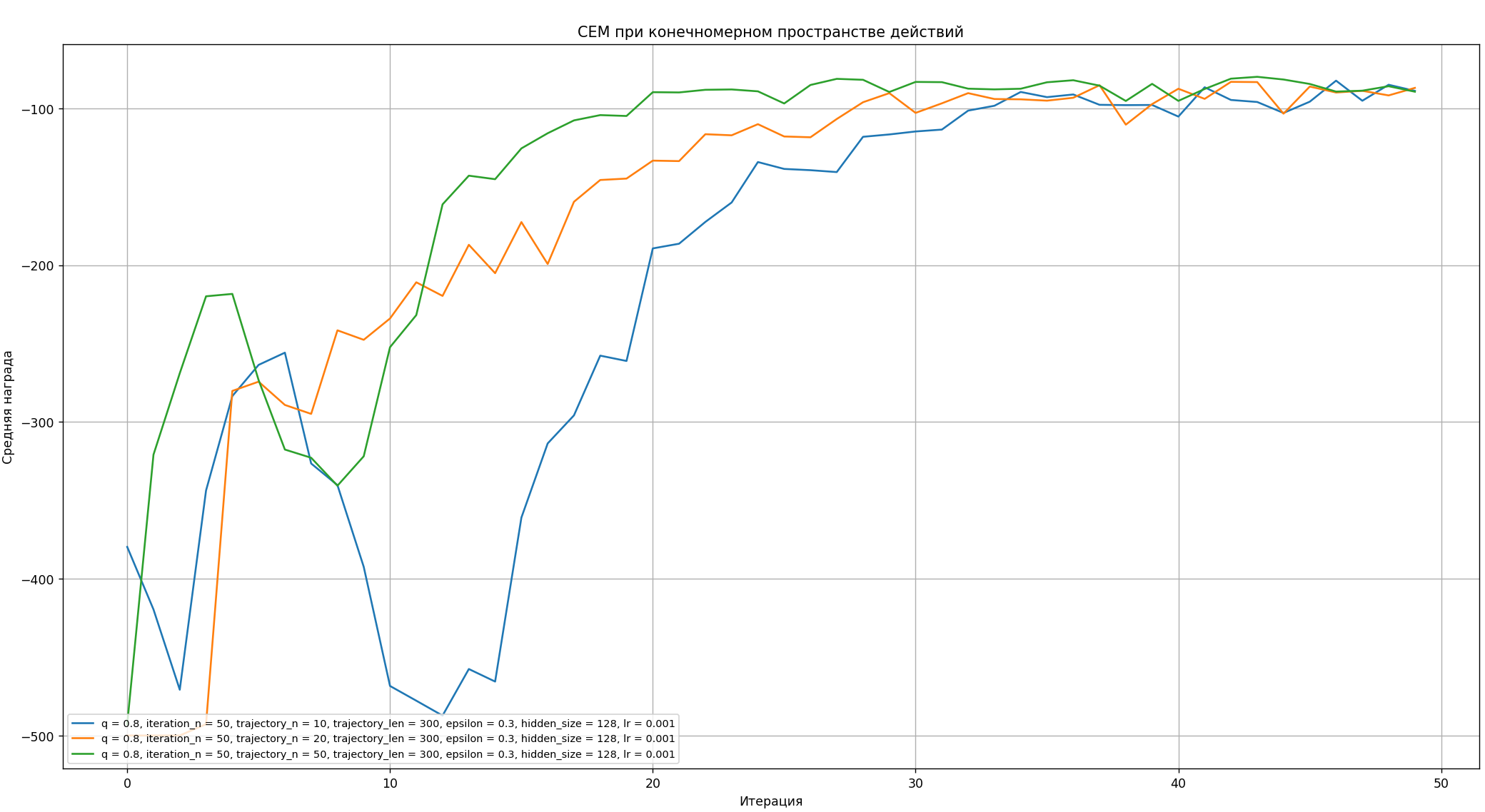


**Вывод:** если честно, объяснение результатов неочевидны. Два раза из трёх обучение при trajectory\_len = 500 не показывало себя никак, при = 1000, тоже не очень. Наверное, 300 – лучший вариант ввиду малой вариабельности действий в задаче. Возможно, стоило бы убавить ещё, но пока примем trajectory\_len = 300.

**Эксперимент №3:** попробуем перебрать количество траекторий. Набор для перебора = [10, 20, 50].

Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 128, lr = 1e-3, q = 0.8, длина траектории = 300.

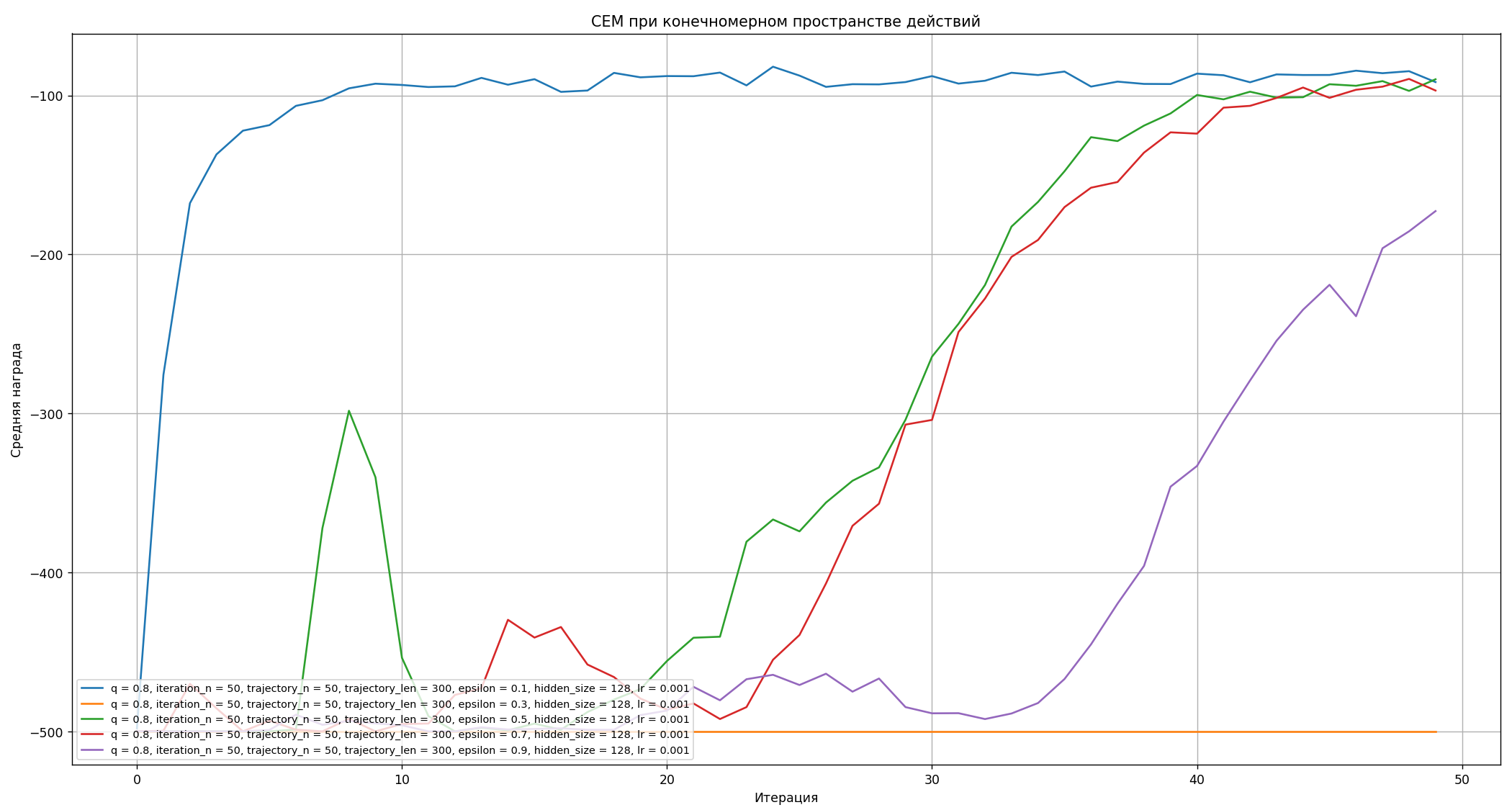


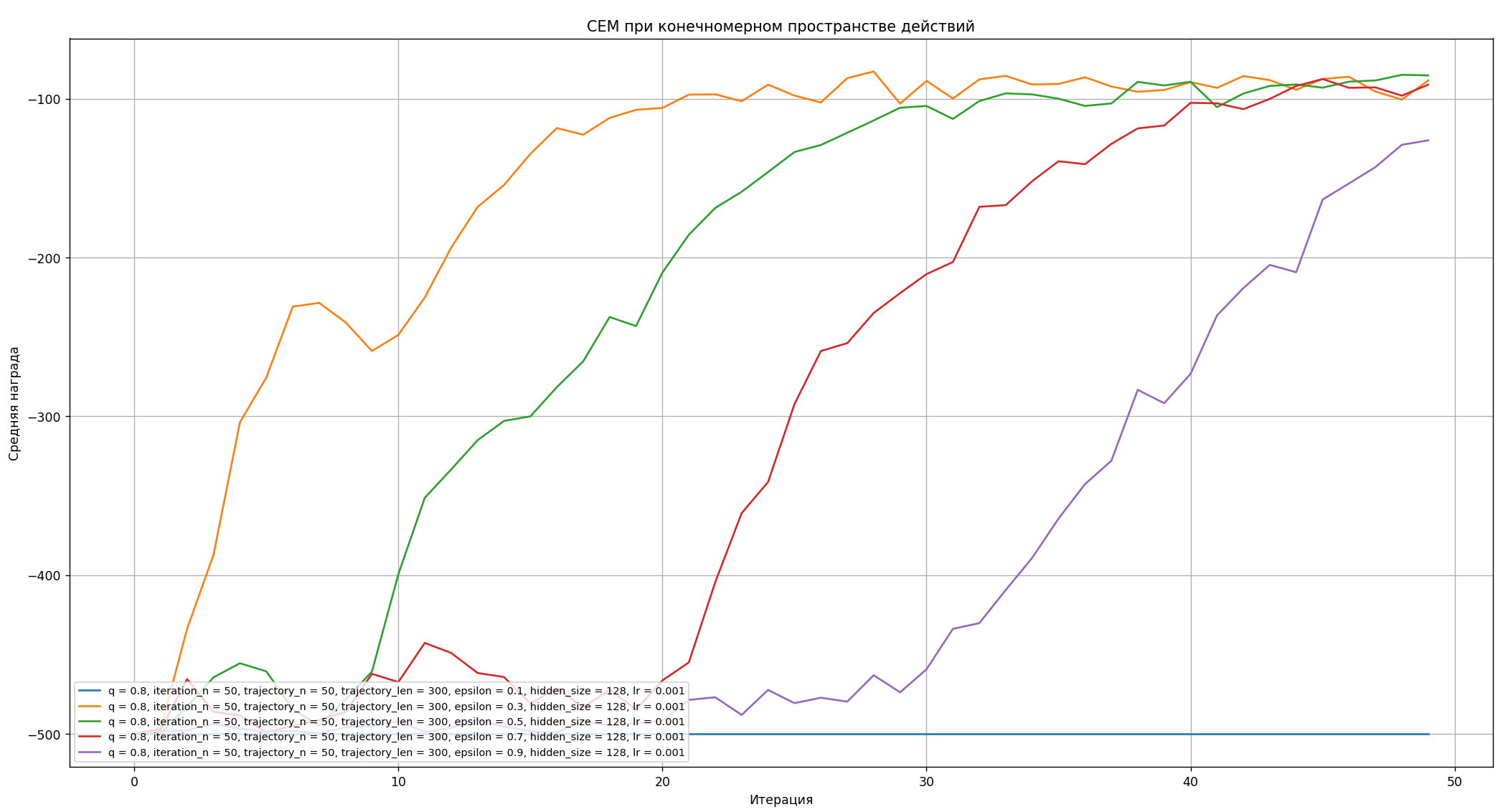


**Вывод:** ну вроде как, чем больше сэмплируем траекторий, тем лучше. Берём trajectory\_n = 50. Но на данный момент, по прошествии 3 экспериментов, главный вывод – выбранные параметры для перебора видимо не так сильно влияют на качество обучения (ну точнее должны быть какие-то более «сильные» признаки). В каждом случае агент упирается «в потолок» награды около - 70-80 баллов

**Эксперимент №4:** попробуем перебрать значения эпсилон. Набор для перебора = [0.1, 0.3, 0.5, 0.7, 0.9].

Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, размер скрытого слоя = 128, lr = 1e-3, q = 0.8, длина траектории = 300, количество траекторий = 50.



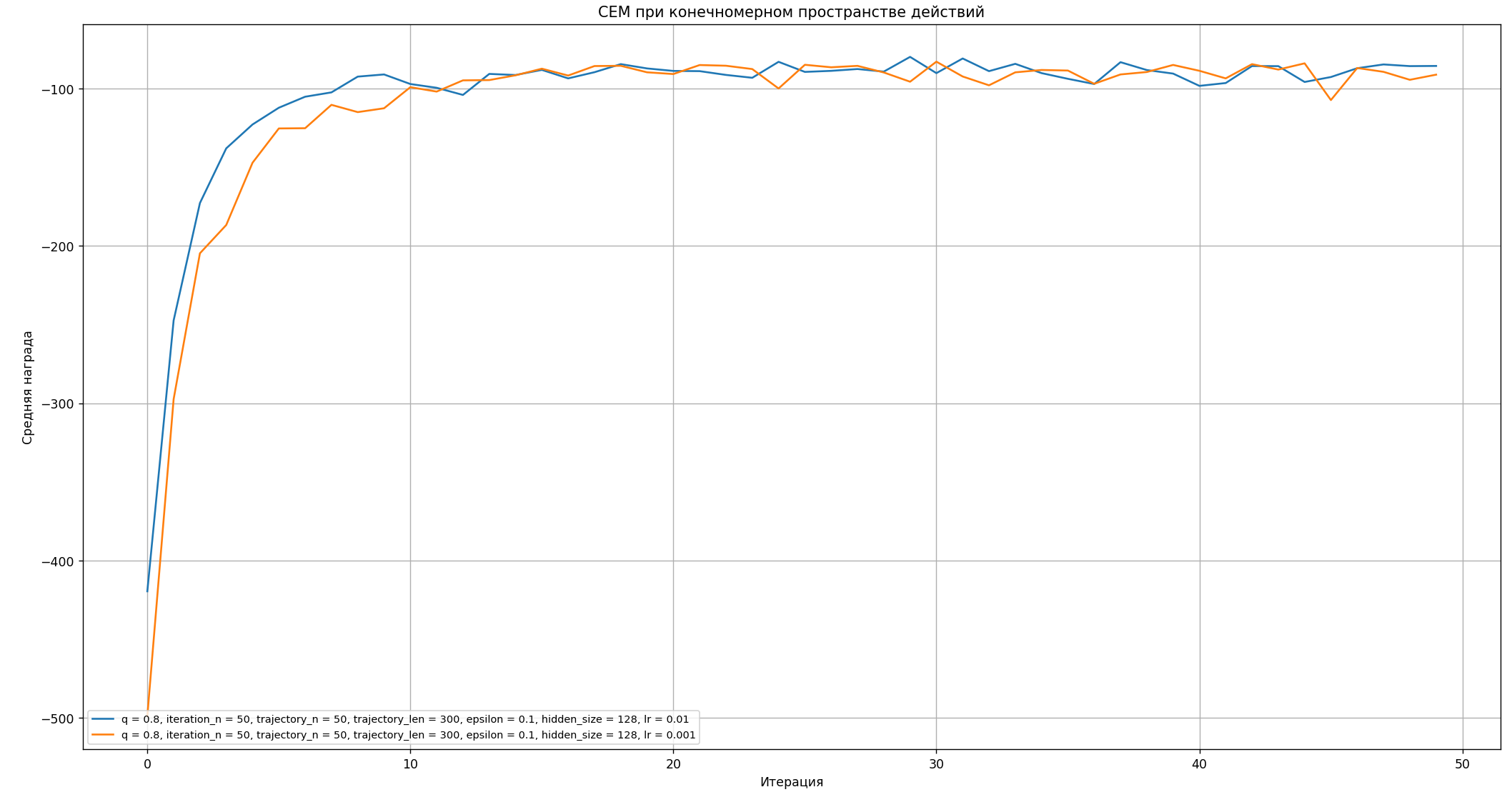


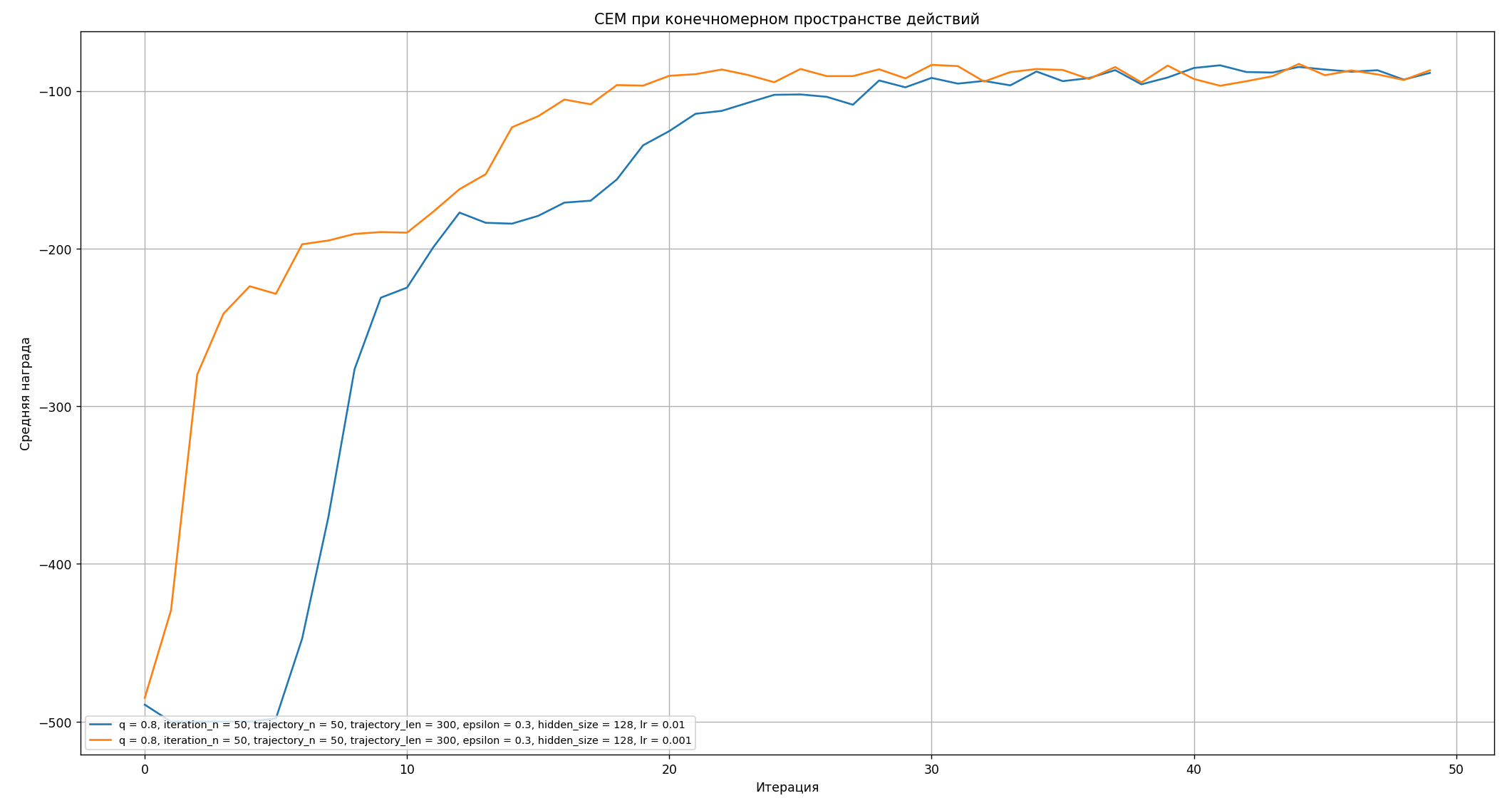


**Вывод:** реакция на изменение эпсилона тоже не однозначная. Видимо, как повезёт. Нужно повышать количество траекторий и больше итераций и тогда картина будет яснее? Пока берём эпсилон = 0.3.

**Эксперимент №5:** попробуем перебрать скорость обучения. Набор для перебора = [1e-2, 1e-3, 5e-3].

Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, эпсилон = 0.3, q = 0.8, длина траектории = 300, количество траекторий = 50, размер скрытого слоя = 128.

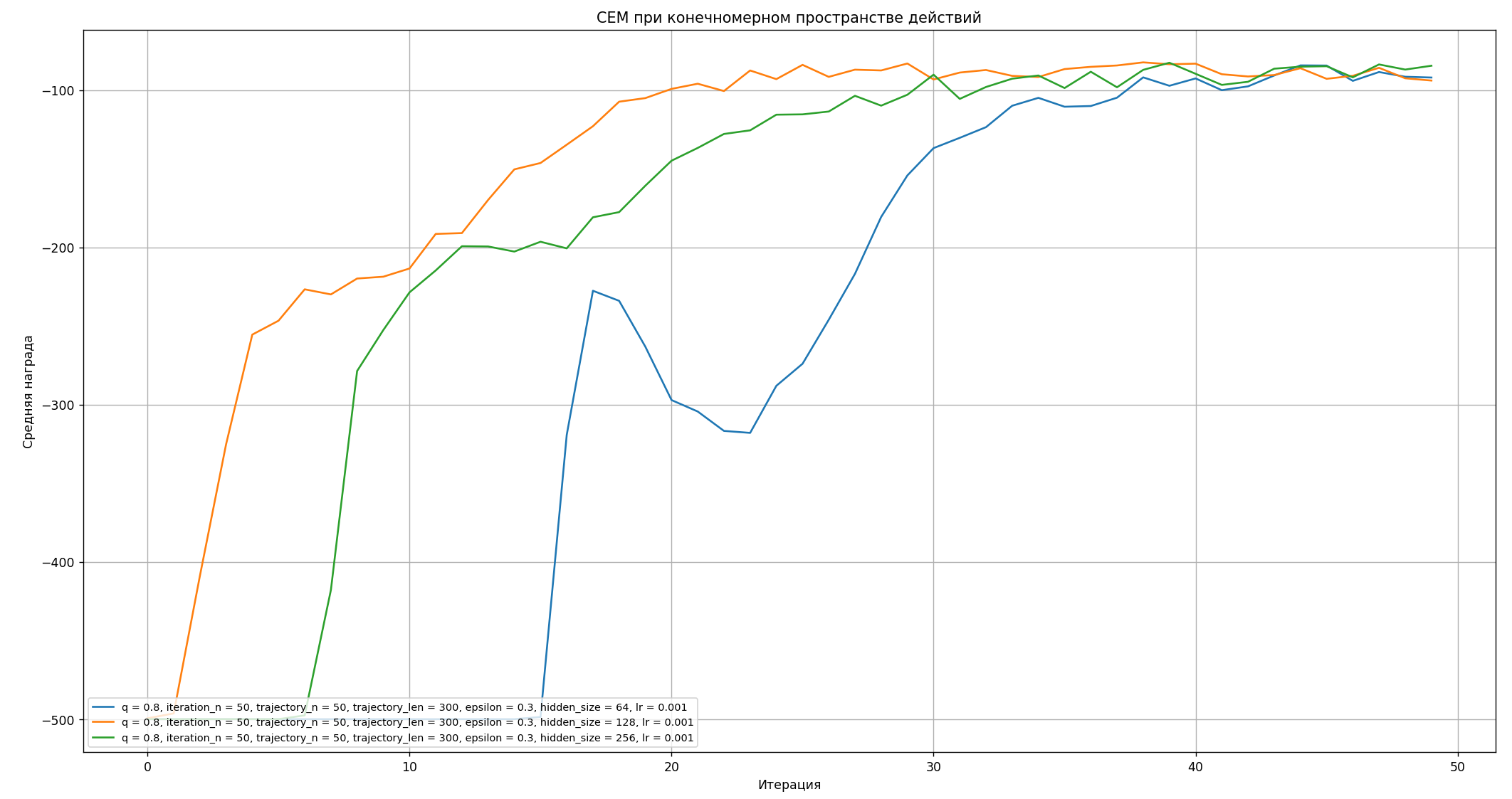




**Вывод:** лучше lr = 1e-3.

**Эксперимент №6:** попробуем перебрать размер скрытого слоя. Набор для перебора = [64, 128, 256].

Остальные параметры: кол-во итераций = 50, кол-во траекторий = 20, эпсилон = 0.3, lr = 1e-3, q = 0.8, длина траектории = 300, количество траекторий = 50.



**Вывод:** лучше показал размер скрытого слоя = 128.

**Общий вывод по задаче №1:** задачи объективно труднее, чем первое дз. Честно говоря, мне всё ещё не до конца очевидно, как гиперпараметры влияют на ход обучения агента, ибо даже проводя одни и те же эксперименты по несколько раз, результаты иногда кардинально отличались. И не был так и найден нужный параметр, с помощью которого можно было перепрыгнуть потолок в -70.

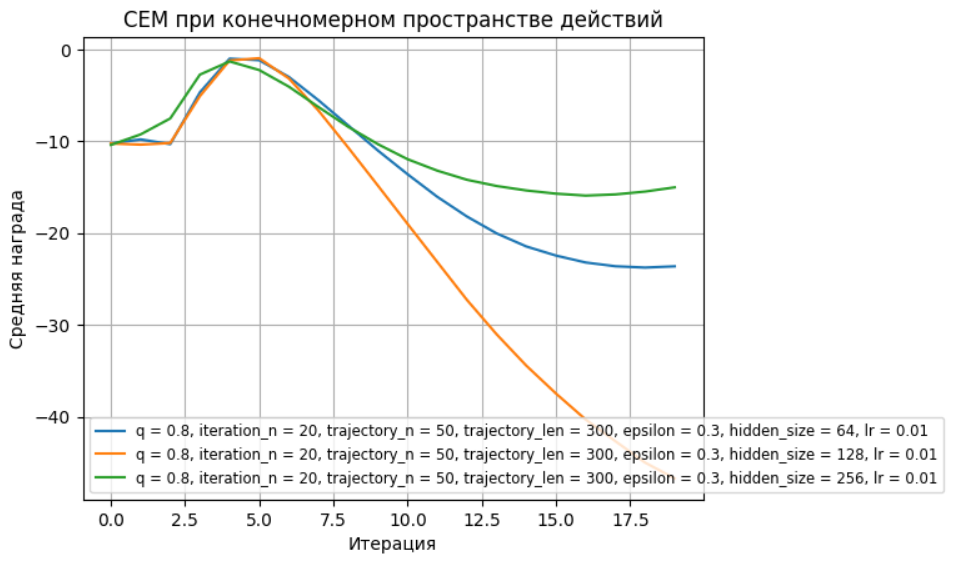
Либо для этой задачи нужна архитектура посложнее или поглубже поработать с данной на практике. Или ошибки в коде 😊

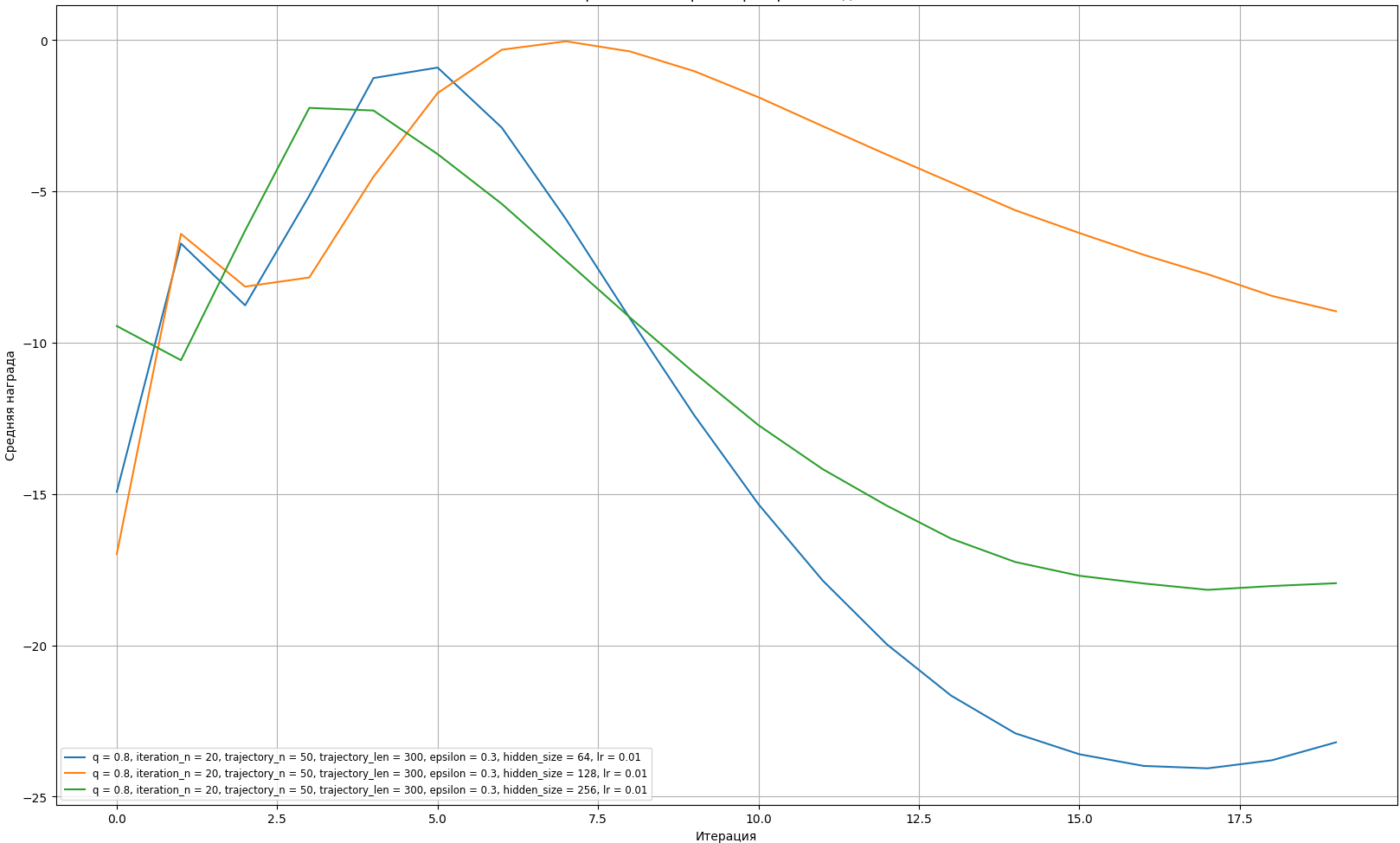
**Лучшие параметры**: кол-во итераций = 50, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 128, lr = 1e-3, q = 0.8, длина траектории = 300, количество траекторий = 50.

**Задание №2:** пользуясь алгоритмом Кросс-Энтропии для непрерывного пространства действий обучить агента решать MountainCarContinuous-v0. Исследовать гиперпараметры алгоритма и выбрать лучшие.

**Эксперимент №1:** попробуем перебрать размер скрытого слоя. Набор для перебора = [64, 128, 256].

Остальные параметры: кол-во итераций = 20, кол-во траекторий = 20, эпсилон = 0.3, lr = 1e-2, q = 0.8, длина траектории = 300, количество траекторий = 50.



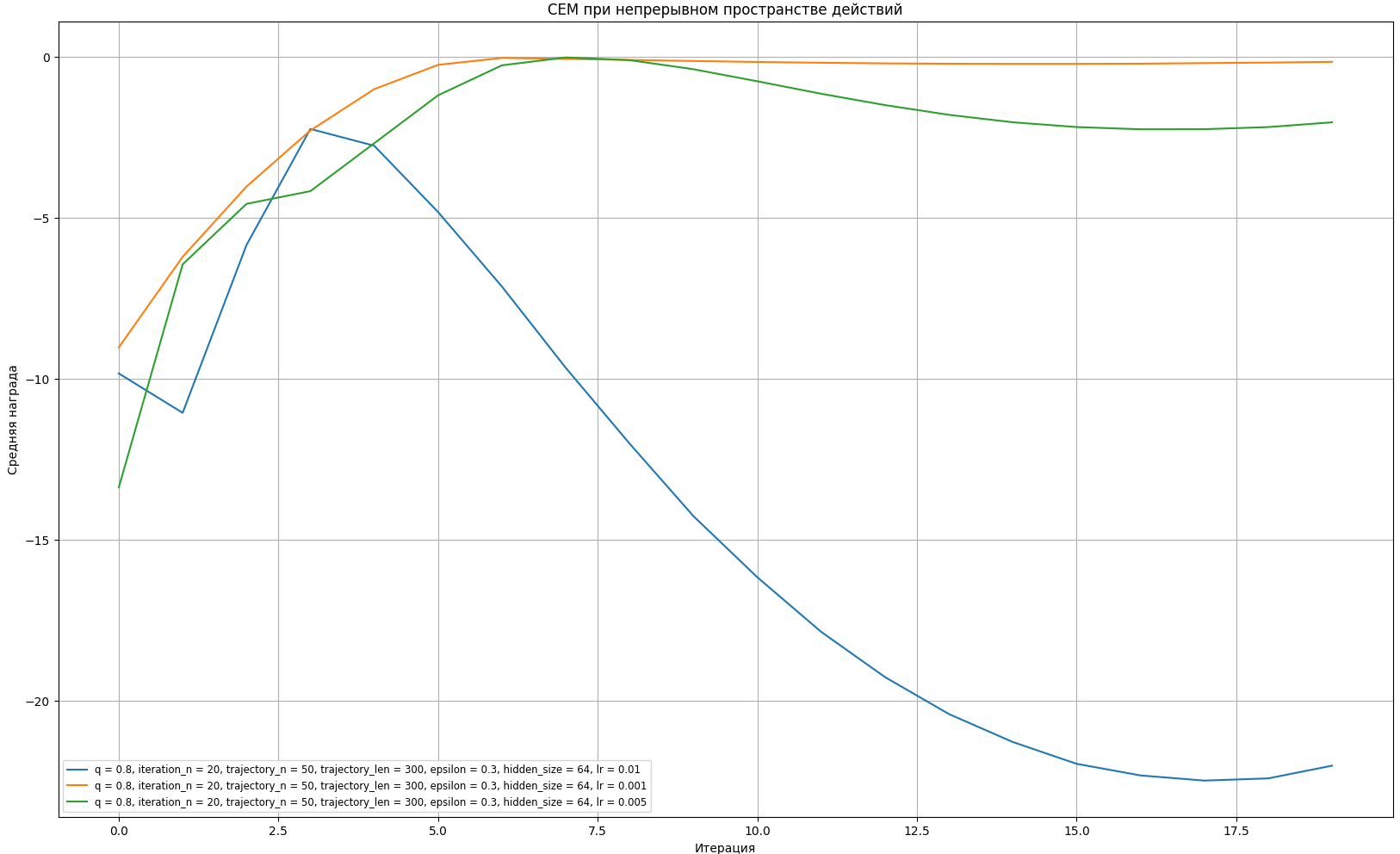


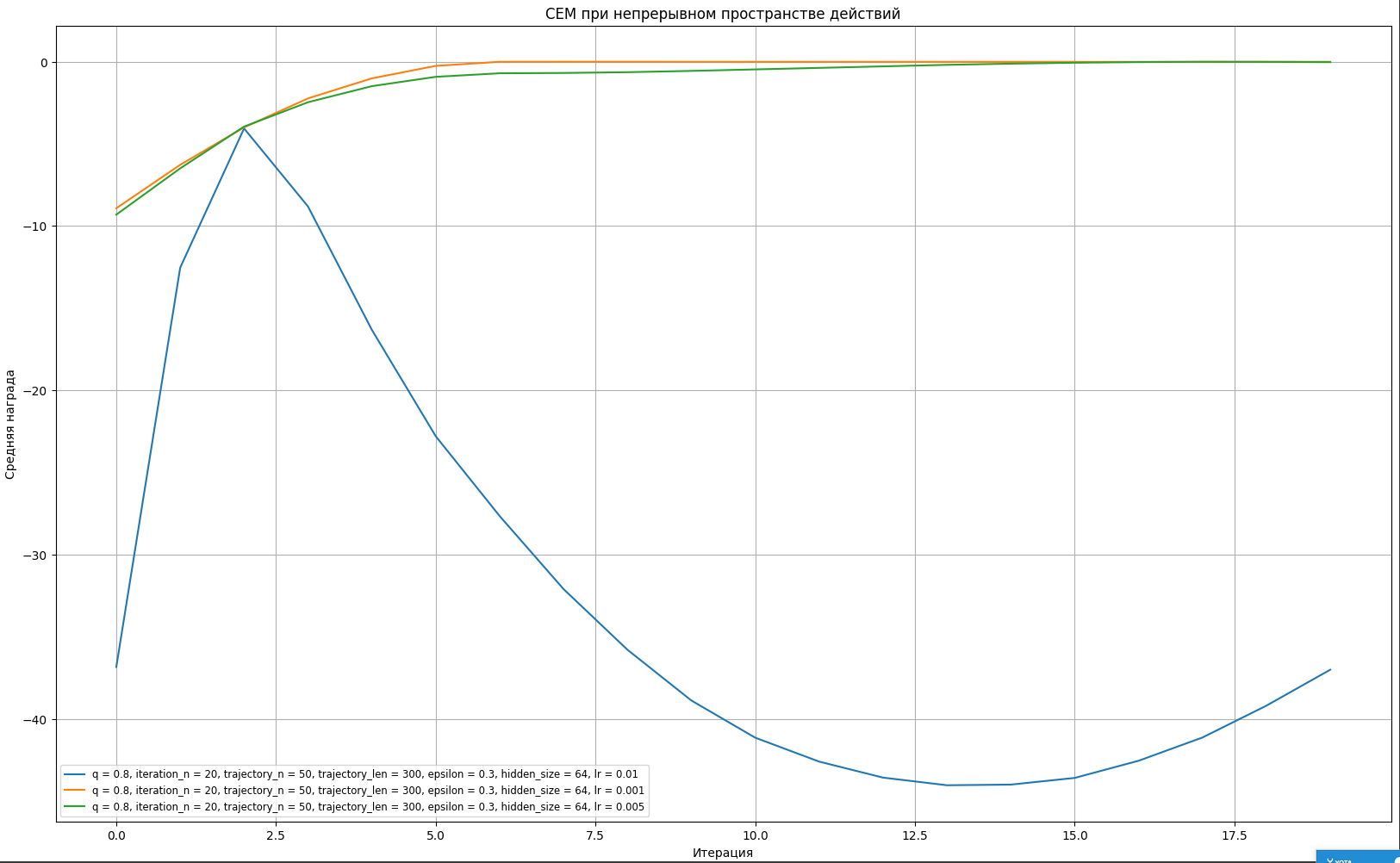
**Вывод:** вцелом, при любом размере скрытого слоя обучение агента достигает потолка, а затем начинает падать. Пока не могут сказать, почему (может переобучение?...). Вцелом можно останавливать обучение (или брать состояние агента на лучшем шаге), когда награда начинает падать.

Пока что берём размер скрытого слоя = 64, для облегчения счёта. Хотя наверняка это число можно ещё уменьшить.

**Эксперимент №2:** попробуем перебрать скорость обучения. Набор для перебора = [1e-3, 5e-3, 1e-2].

Остальные параметры: кол-во итераций = 20, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 64, q = 0.8, длина траектории = 300, количество траекторий = 50.

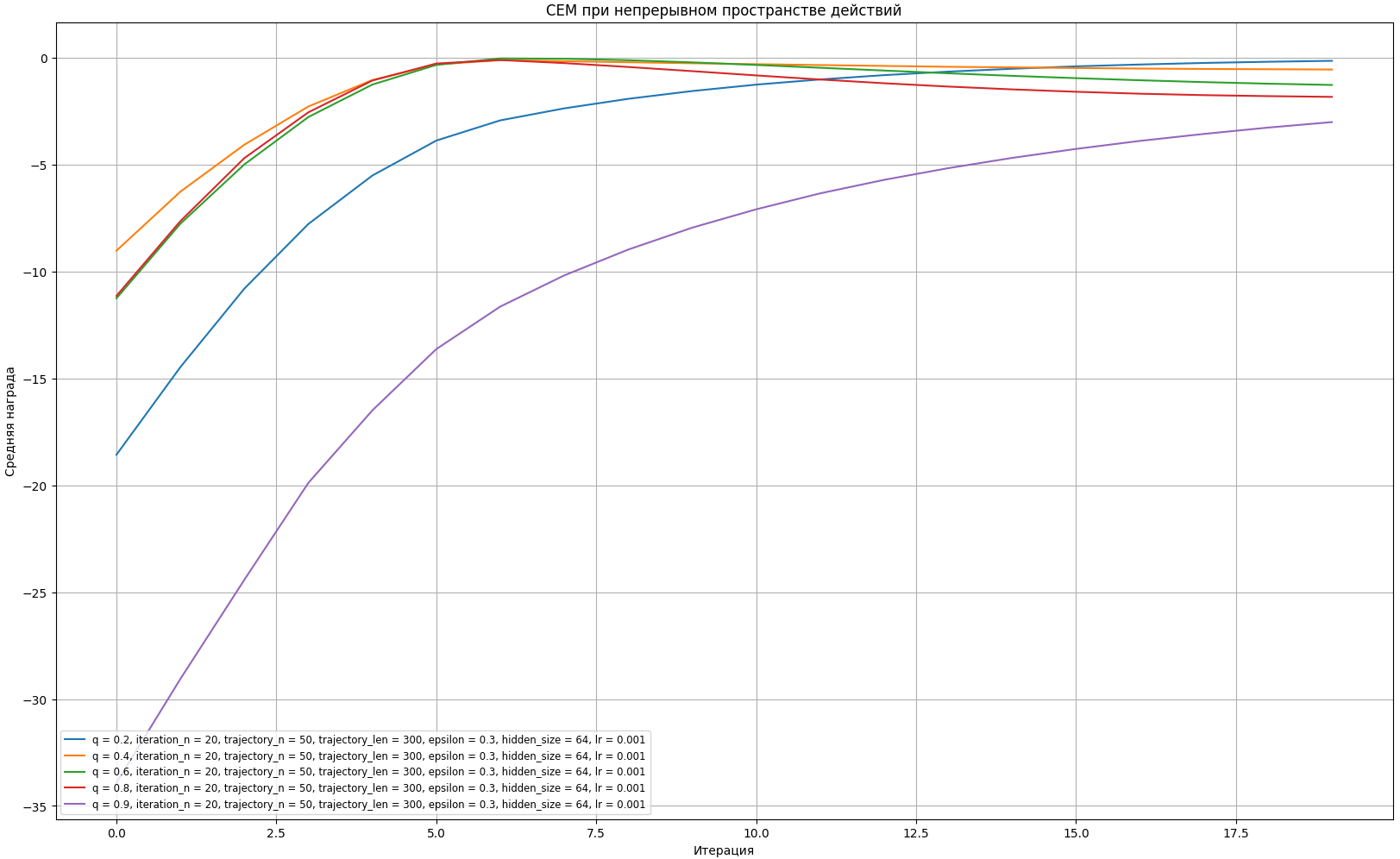




**Вывод:** лучше всего при маленьком lr = 0.001. Берём дальше его.

**Эксперимент №3:** попробуем перебрать параметр q. Набор для перебора = [0.2, 0.4, 0.6, 0. 8, 0.9].

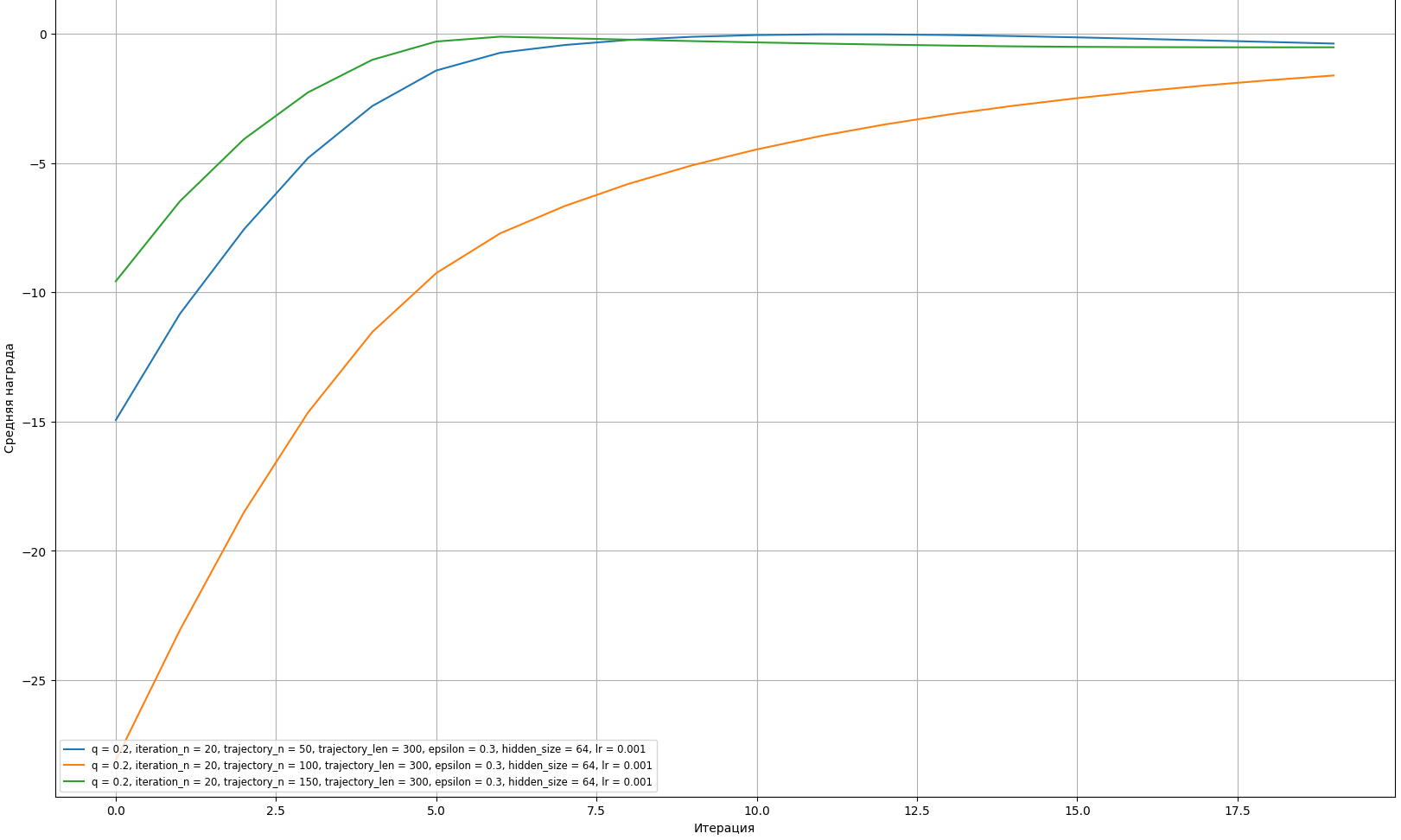
Остальные параметры: кол-во итераций = 20, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 64, длина траектории = 300, количество траекторий = 50, lr = 1e-3.



**Вывод:** закономерность такая: чем меньше q, тем лучше. Берём q = 0.2.

**Эксперимент №4:** попробуем перебрать количество траекторий. Набор для перебора = [50, 100, 150].

Остальные параметры: кол-во итераций = 20, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 64, длина траектории = 300, lr = 1e-3, q = 0.2.



**Вывод:** нужна ещё проба, потому что вывода не сделать. Ибо 50 и 150 показали себя лучше, чем 100. Ну пусть будет 150 для лучшей вариабельности.

**Общий вывод по 2 задаче:** к сожалению, из-за нехватки времени, не смог перебрать все гиперпараметры. Но из того, что получилось, хотелось бы особо отметить шаг по смене скорость обучения, что при схождении в точку экстремума при маленьком lr, график останавливался «на плато», а не начинал уходить вниз. А так при любых перебираемых параметрах, самый лучший график достигал потолка (около 0) на 5-6 итерации. Поэтому вцелом, для хорошего обучения этой задаче почти любые гиперпараметры подходят.

**Лучшие параметры:** кол-во итераций = 20, кол-во траекторий = 20, эпсилон = 0.3, размер скрытого слоя = 64, длина траектории = 300, количество траекторий = 150, lr = 1e-3, q = 0.2.