МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ОТЧЕТ
по лабораторной работе №1
по дисциплине «Обучение с подкреплением»
Тема: Реализация DQN для среды CartPole-v1

Студент гр. 0310 _____ Якушкин Д.А.

Преподаватель _____ Глазунов С.А.

Санкт-Петербург
2025 г.

Цель работы.

Реализация DQN для среды CartPole-v1. Исследование влияния различных параметров: архитектура сети, значения gamma и epsilon_decay, влияние epsilon на скорость обучения

Задание.

1. Реализовать DQN для среды CartPole-v1
2. Изменить архитектуру нейросети.
3. Попробовать разные значения gamma и epsilon_decay.
4. Провести исследование как изначальное значение epsilon влияет на скорость обучения

Выполнение работы.

1. **Реализация DQN**

DQNAgent это класс, который реализует сам алгоритм DQN. Его задача принимать решения, которые приведут к наибольшей выгоде. Стандартные параметры для него будут следующими:

```
base_config = {
        "layers”: [64, 64],
        "gamma": 0.99,
        "epsilon_decay": 0.99,
        "epsilon_start": 1.0
    }
```

2. **Влияние архитектуры на обучение**

Для экспериментов были выбраны следующие структуры слоев:

```
layer_params = {
  "default": [64, 64],
  "deep":    [128, 128, 64],
  "wide":    [256, 128],
```

```
    "small":   [32, 32],
}
```

Были получены сводные графики loss и reward для каждого эксперимента. График loss можно увидеть на рисунке 1. График reward можно увидеть на рисунке 2.

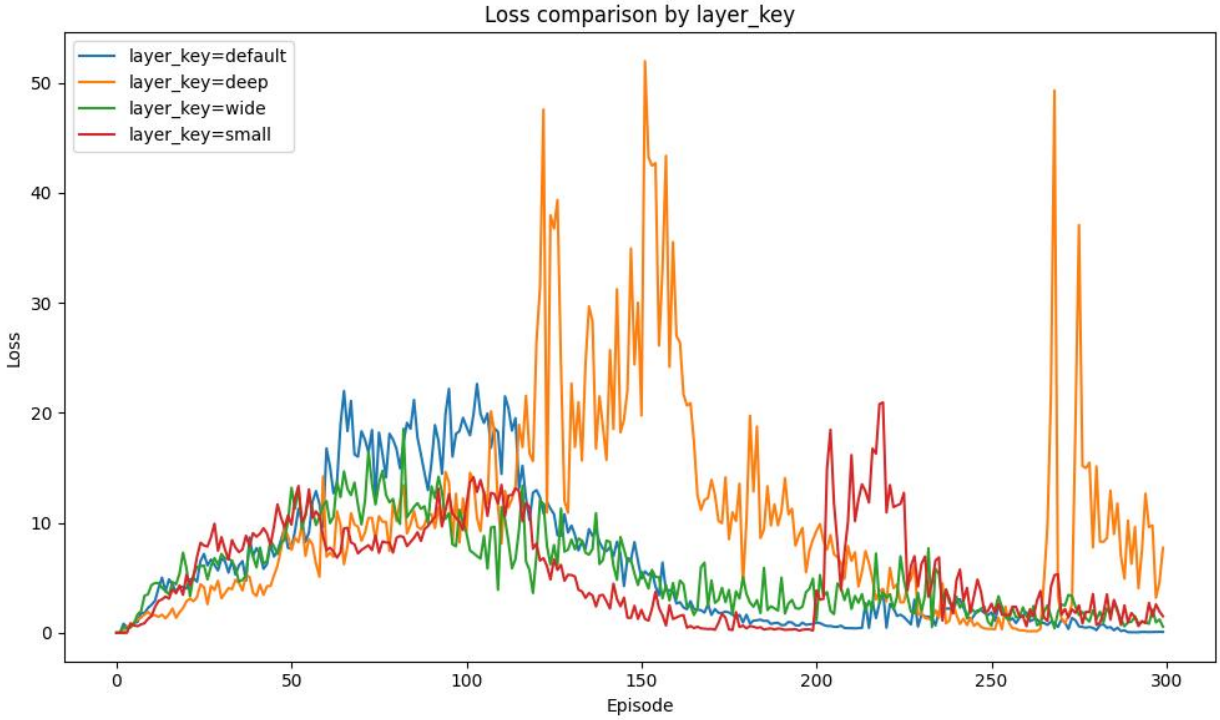

Рисунок 1 - Сравнение loss для различных конфигураций слоёв

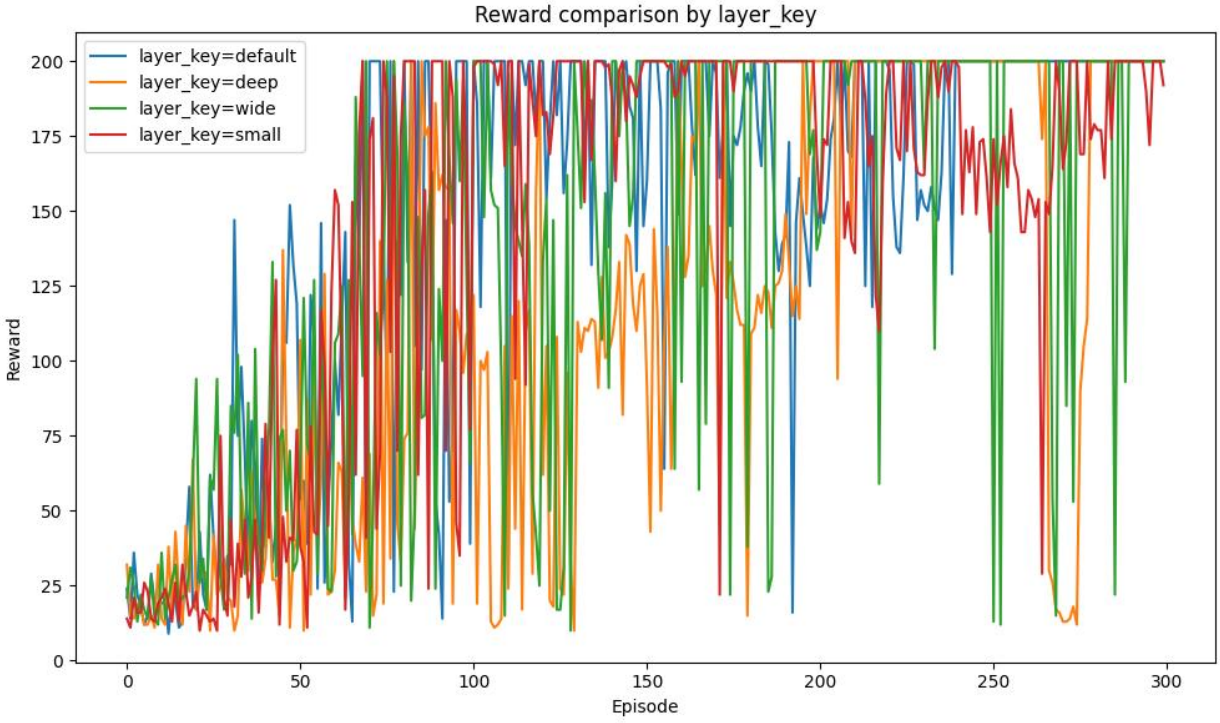

Рисунок 2 - Сравнение loss для различных конфигураций слоёв

Исходя из графиков можно судить что наилучший показатель нейросеть показывает при small и default конфигурациях слоев. Это может происходить из за того, что задача может быть слишком легкой для больших нейросетей и возможно переобучение, для больших нейросетей также может понадобиться большее количество эпизодов ввиду более медленной сходимости.

3. Влияния gamma на обучение

Gamma - это параметр дисконтирования. Он условно указывает то, насколько нейросеть учитывает будущие награды. Чем выше gamma, тем дольше мы можем ожидать большую выгоду.

Для эксперимента были выбраны значения 0.95 и 0.99. Результаты можно увидеть на рисунках 3 и 4.

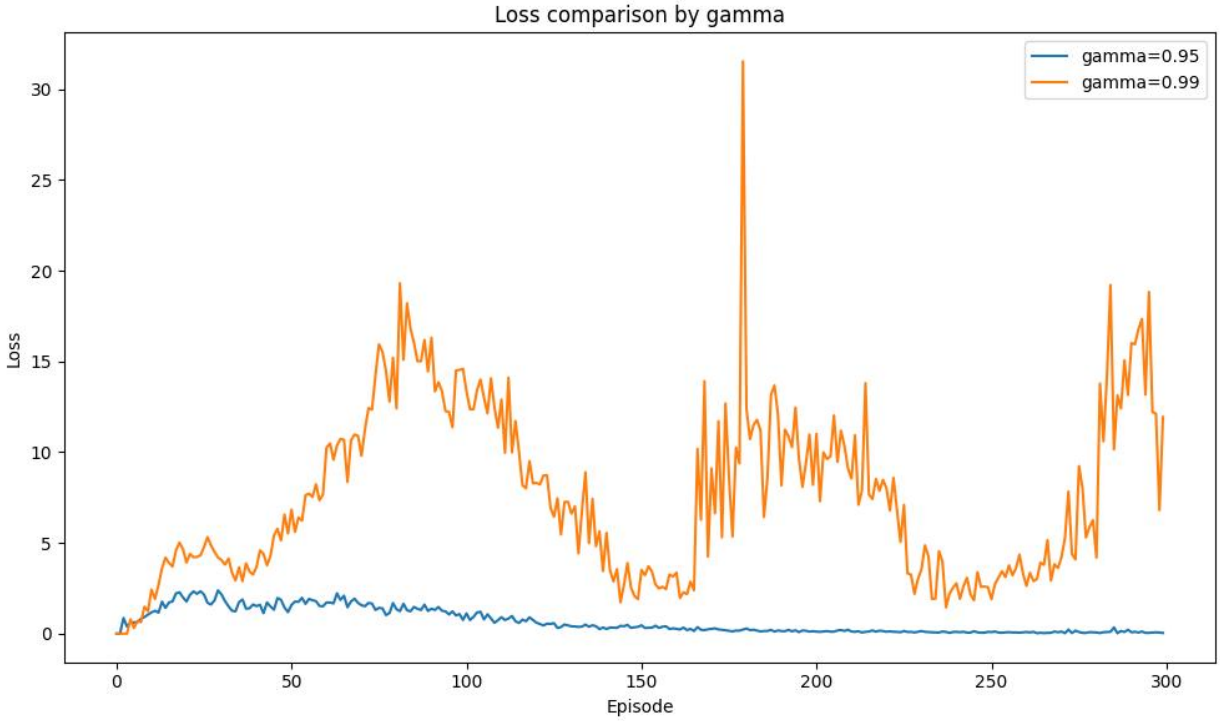

Рисунок 3 - Сравнение loss для различных параметров gamma

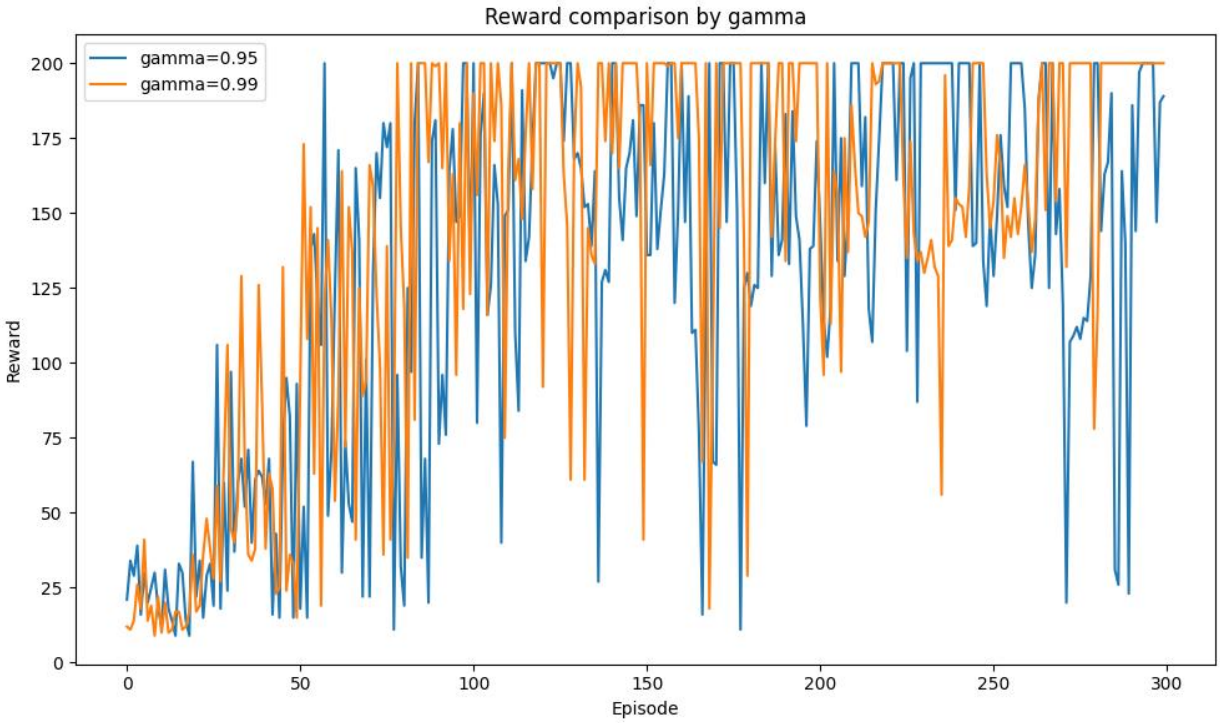

Рисунок 4 - Сравнение loss для различных параметров gamma

Исходя из графиков можно сделать вывод что gamma=0.95 даёт более быстрое достижение результата с меньшим количеством ошибок.

4. Влияние epsilon на обучение

Epsilon - это возможность случайного действия. Этот параметр задает то, как часто нейросеть будет предпринимать новые действия, вместо выполнения уже изученных. Для ее задания применяются два параметра: epsilon_start и epsilon_decay. Первый параметр это начальное значение, а второй это множитель который применяется к начальному параметру каждый эпизод.

Для эксперимента были выбраны значения множителя 0.95 и 0.99. Результаты можно увидеть на рисунках 5 и 6.

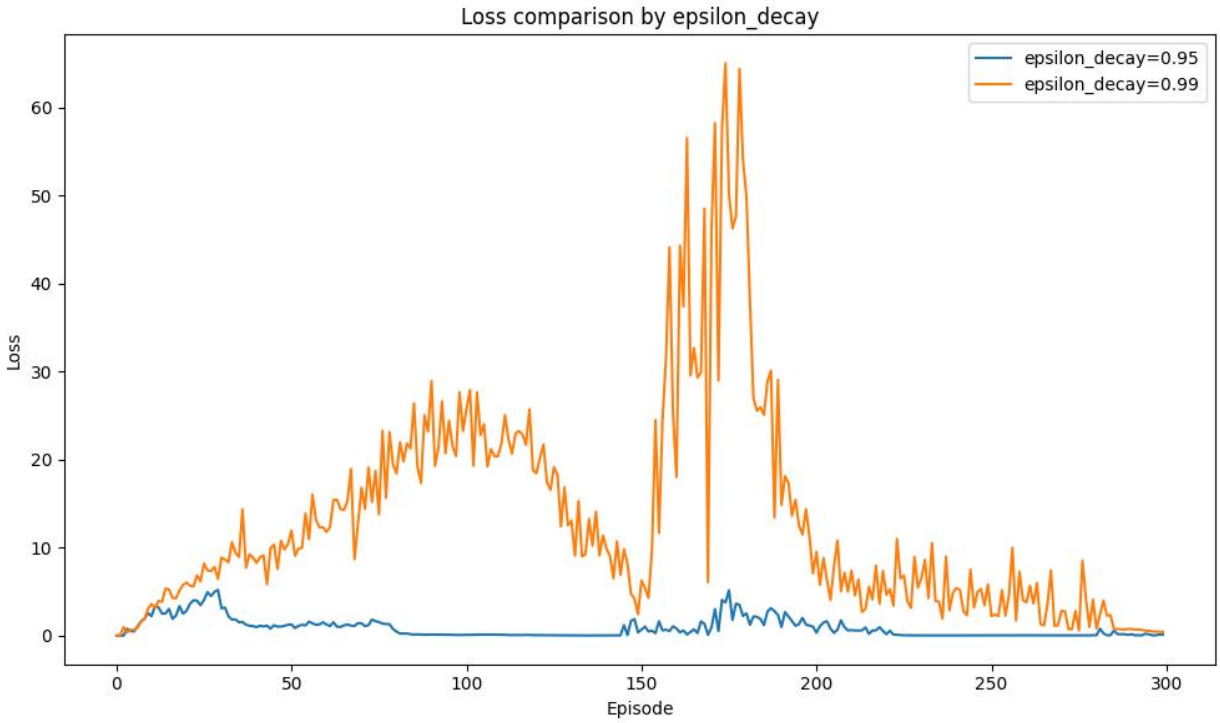

Рисунок 5 - Сравнение loss для различных параметров epsilon_decay

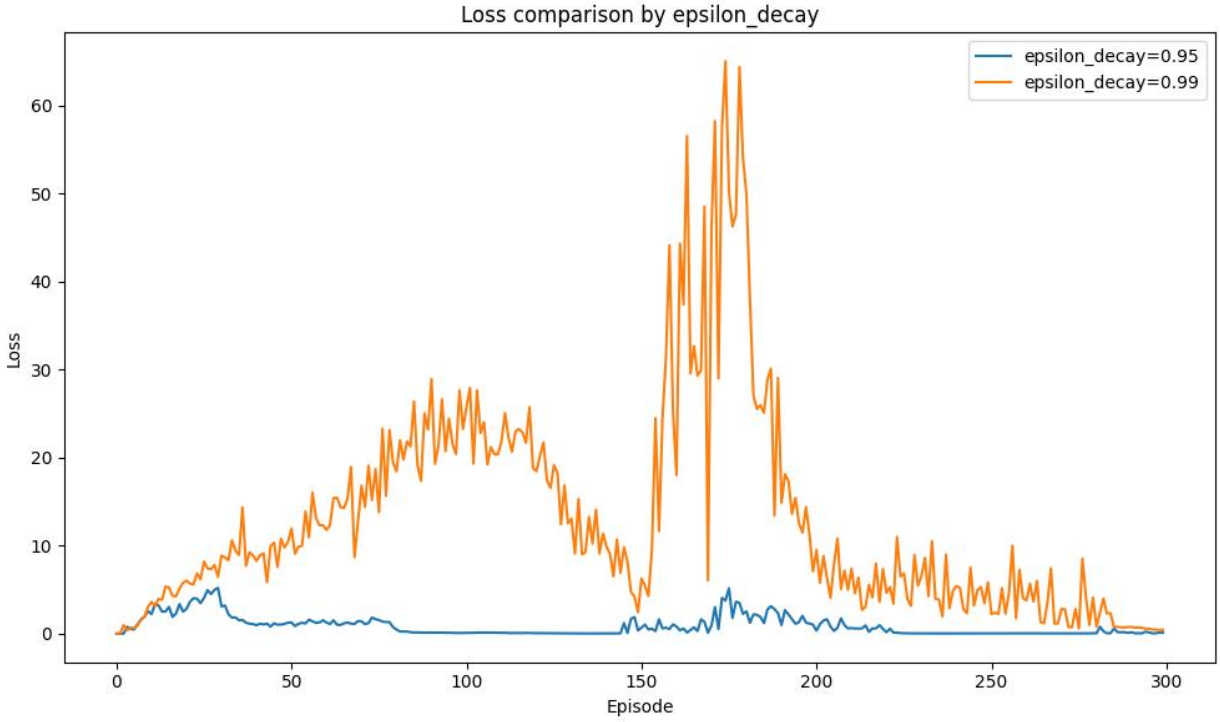

Рисунок 6 - Сравнение reward для различных параметров epsilon_decay

Можно сделать вывод что при более быстром уменьшении вероятности случаности действия мы получаем более быстрое достижение результата и меньшее количество loss.

На графиках 7 и 8 можно увидеть влияние параметра epsilon_start на

обучение.

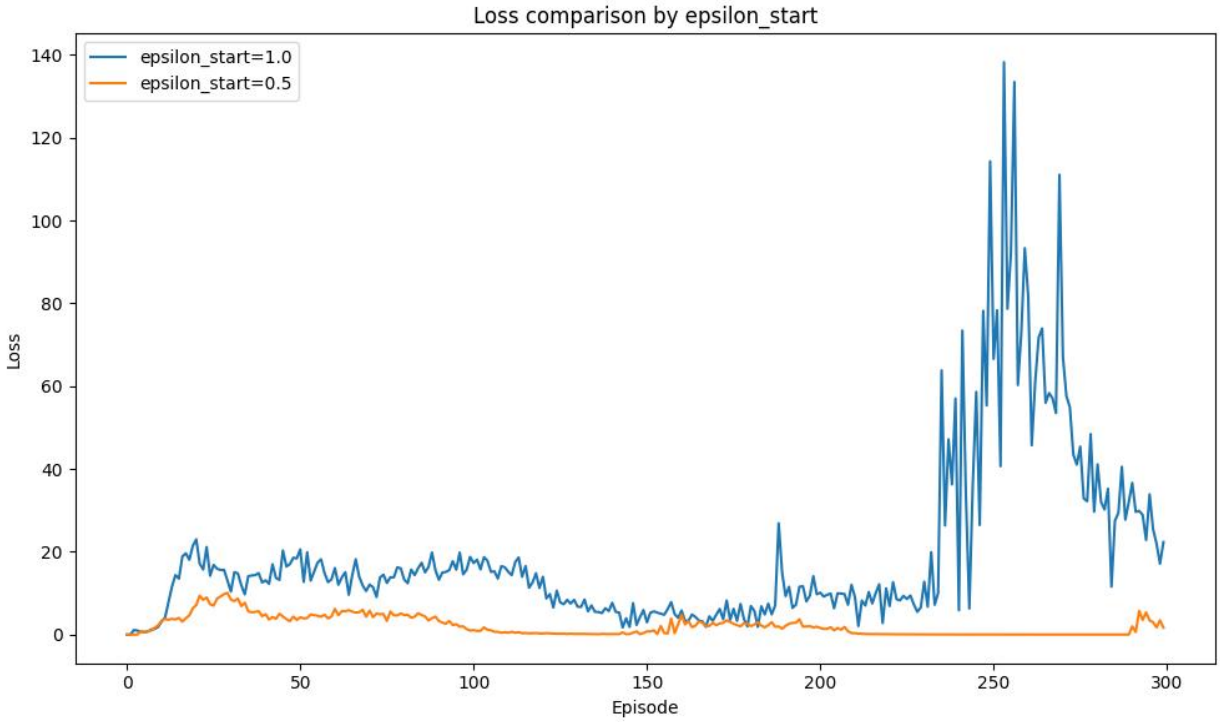

Рисунок 5 - Сравнение loss для различных параметров epsilon_start

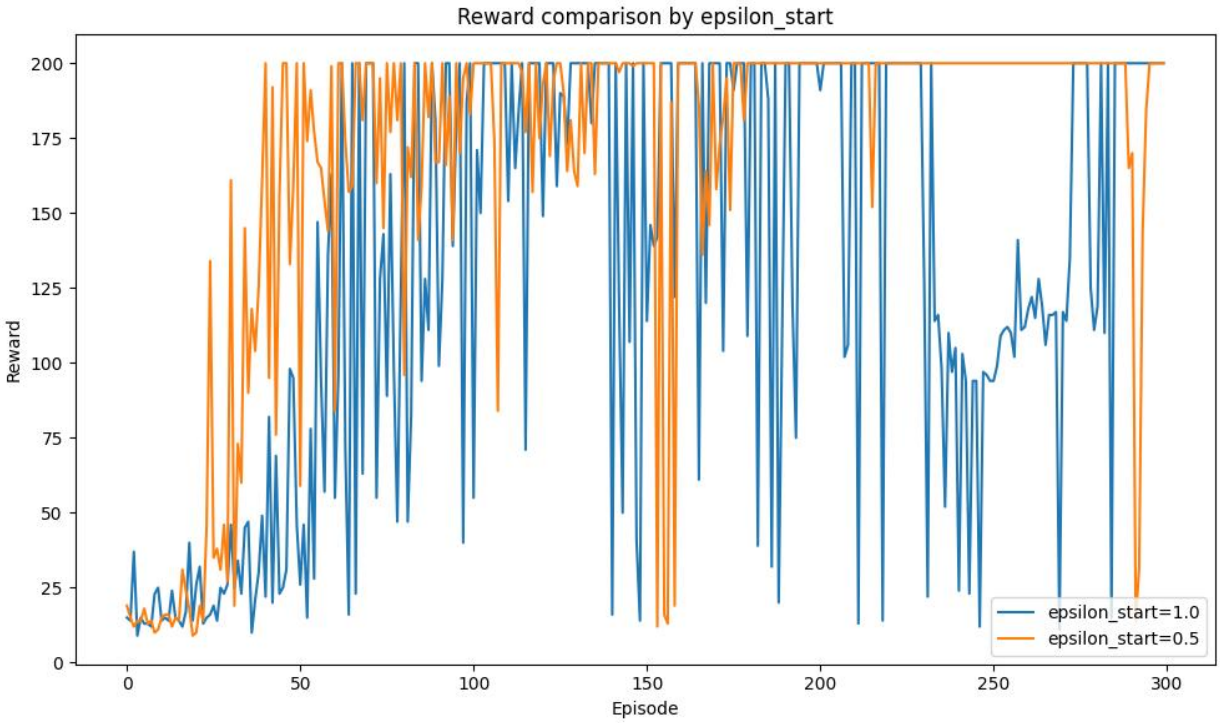

Рисунок 5 - Сравнение reward для различных параметров epsilon_start

Исходя из графиков можно заметить что более низкое значение параметра дает более быстрое обучение.

Выводы.

Был реализован DQN для среды CartPole-v1. Было проведено исследование влияния некоторых параметров сети на результат ее обучения.