# AE598 - HW1 - Dynamic Programming

Maulik Bhatt (mcbhatt2)

March 7, 2023

## 1 Introduction

our goal is to implement five reinforcement learning algorithms in a "tabular" setting (i.e., assuming small finite state and action spaces).

Two algorithms are model-based:

- Policy iteration (Chapter 4.3, Sutton and Barto)
- Value iteration (Chapter 4.4, Sutton and Barto)

Two algorithms are model-free:

- SARSA with an epsilon-greedy policy, i.e., on-policy TD(0) to estimate Q (Chapter 6.4, Sutton and Barto)
- Q-learning with an epsilon-greedy policy, i.e., off-policy TD(0) to estimate Q (Chapter 6.5, Sutton and Barto)

One final algorithm, which is also model-free, computes the value function associated with a given policy:

- TD(0) (Chapter 6.1, Sutton and Barto)

The two algorithms are to be tested in two different environments.

## 2 Grid world-results

### 2.1 Policy Iteration

Fig:1-3 are for Policy Iteration

### 2.2 Value Iteration:

Fig:4-6 are for Value Iteration

### 2.3 SARSA

Fig:7-21 are for this section

### 2.4 Q-learning(off-policy)

Fig:22-36 are for this section

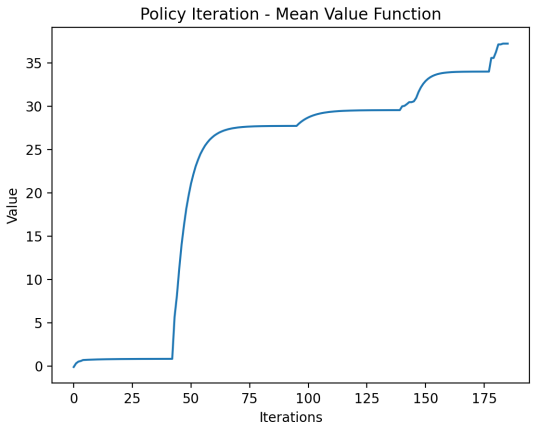


Figure 1:

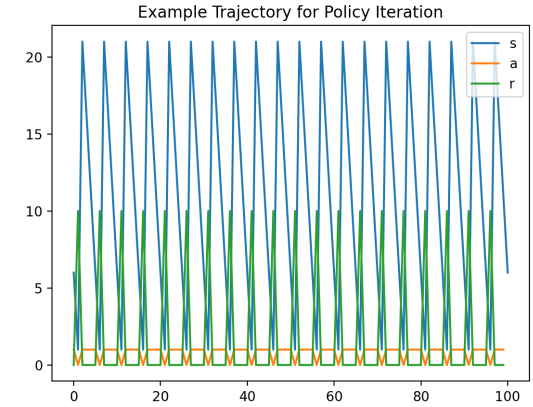


Figure 2:

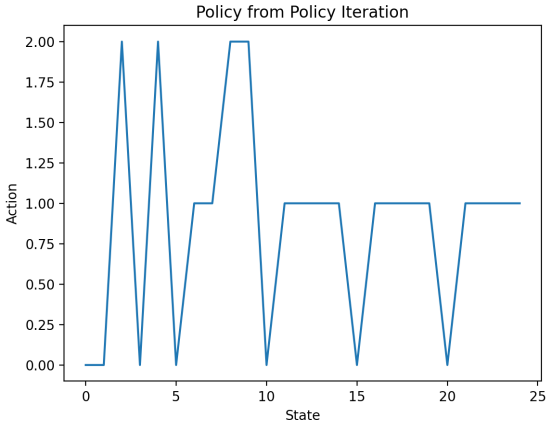


Figure 3:

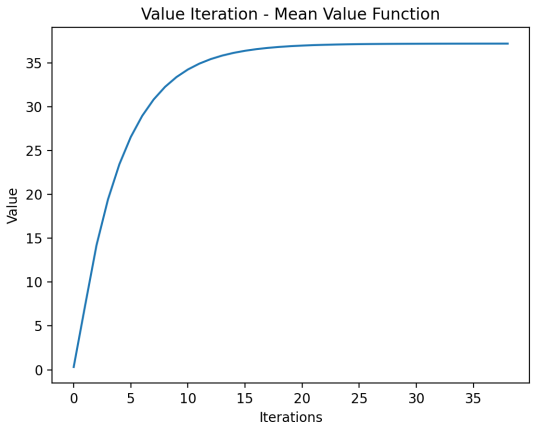


Figure 4:

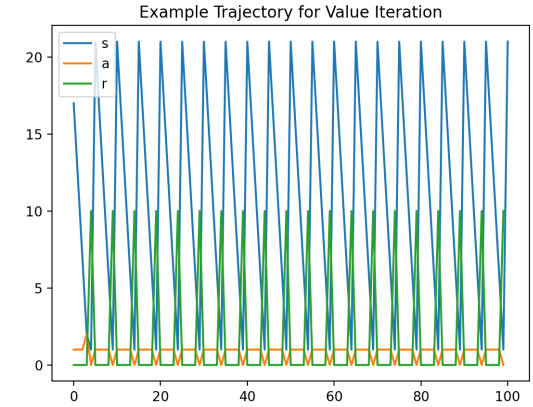


Figure 5:

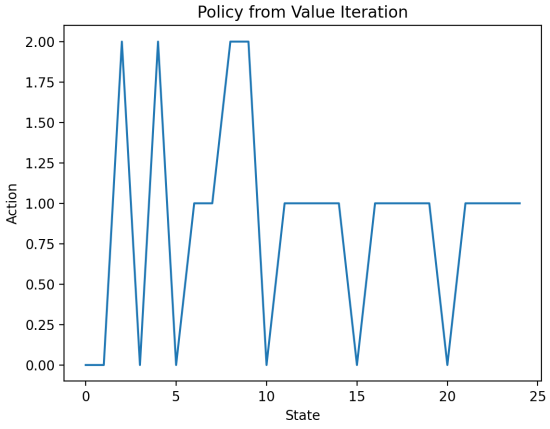


Figure 6:

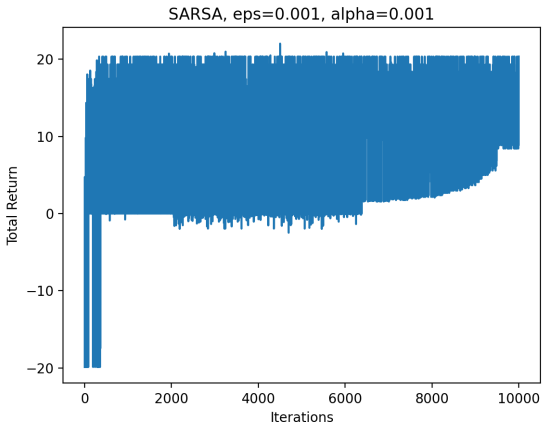


Figure 7:

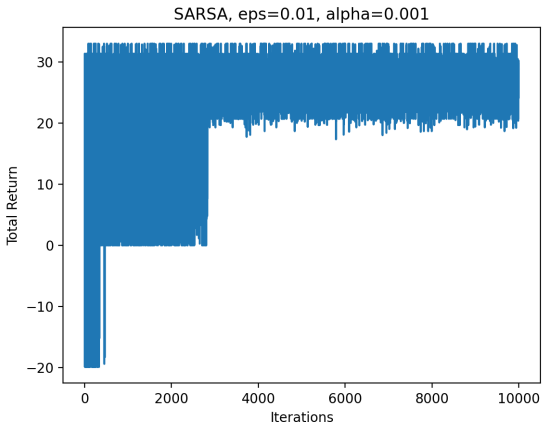


Figure 8:

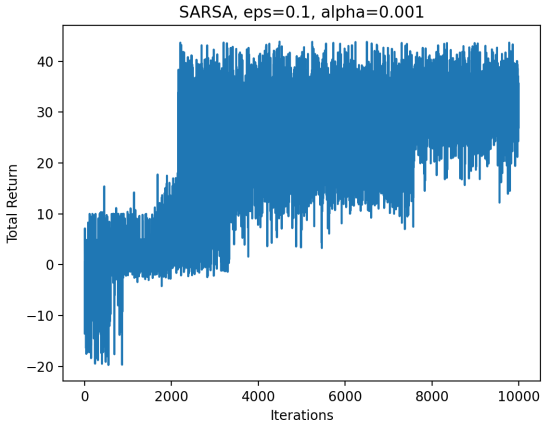


Figure 9:

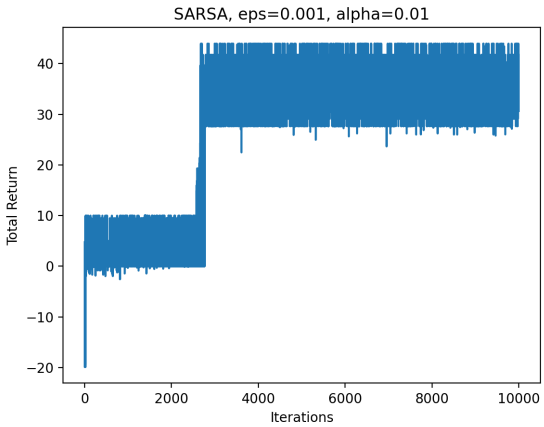


Figure 10:

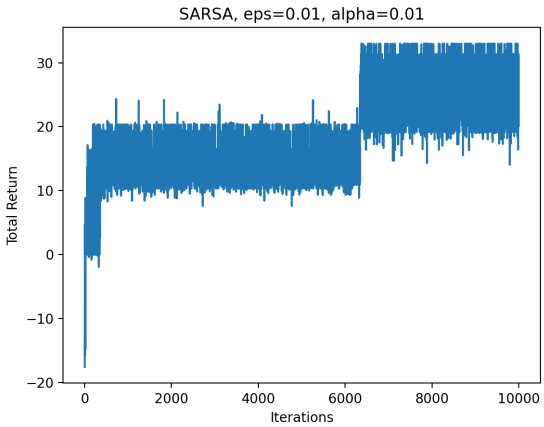


Figure 11:

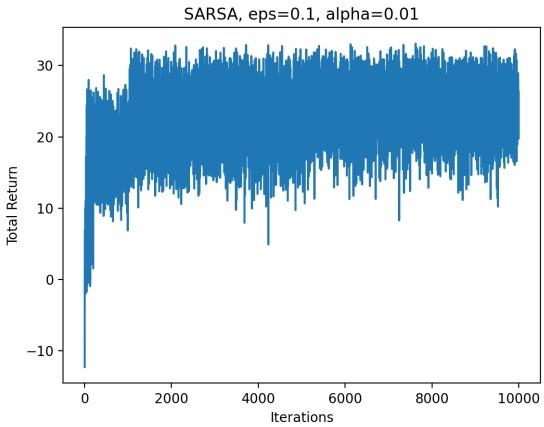


Figure 12:

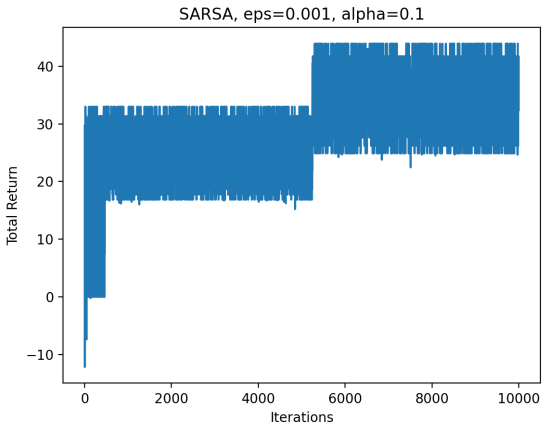


Figure 13:

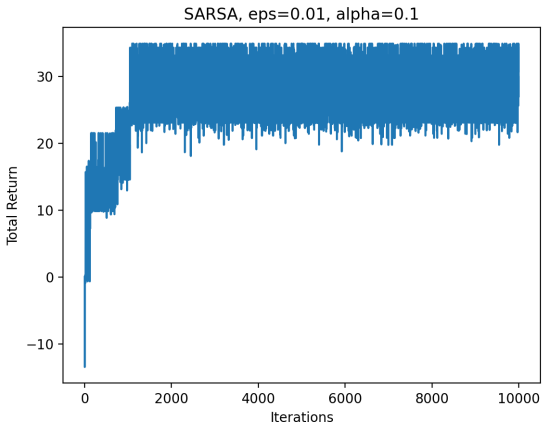


Figure 14:

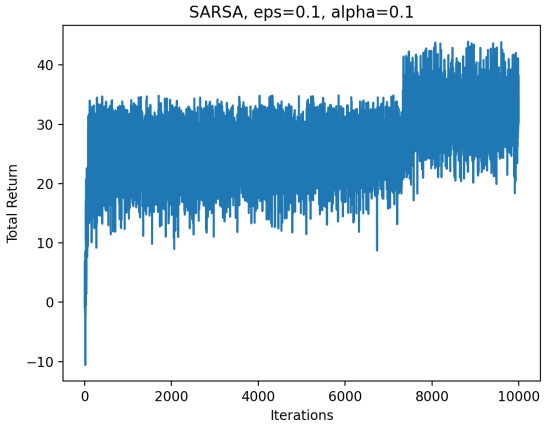


Figure 15:

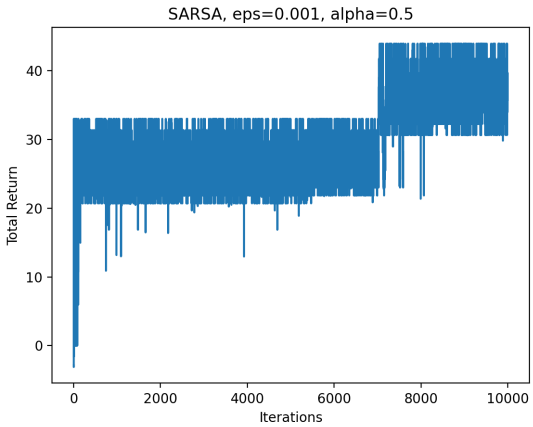


Figure 16:

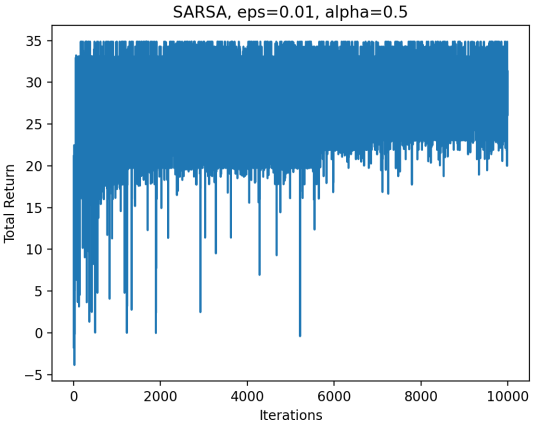


Figure 17:

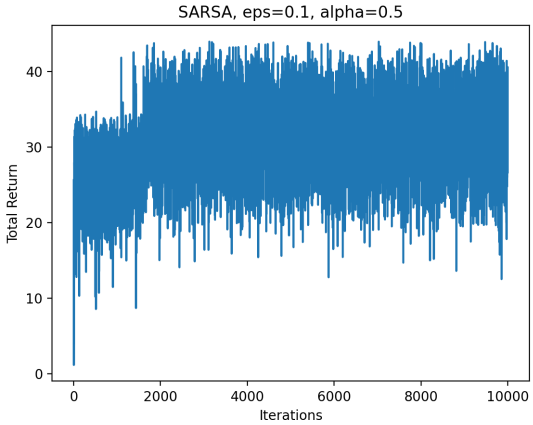


Figure 18:

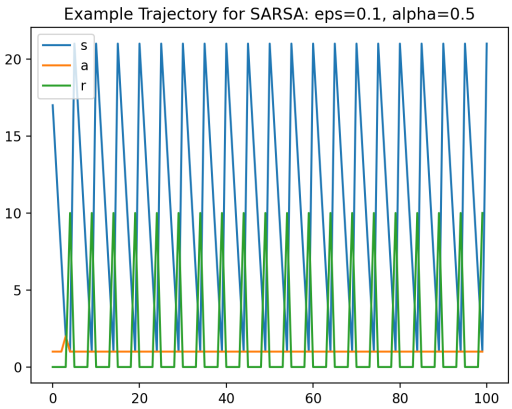


Figure 19:

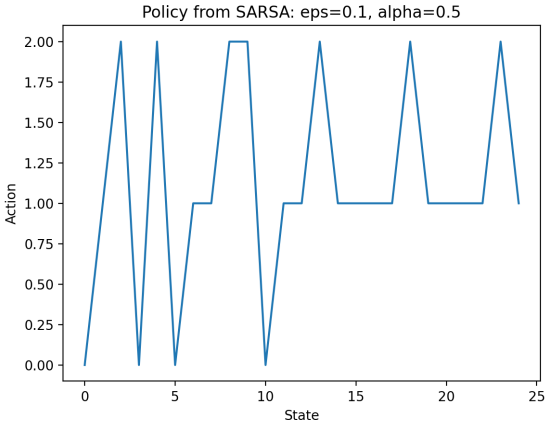


Figure 20:

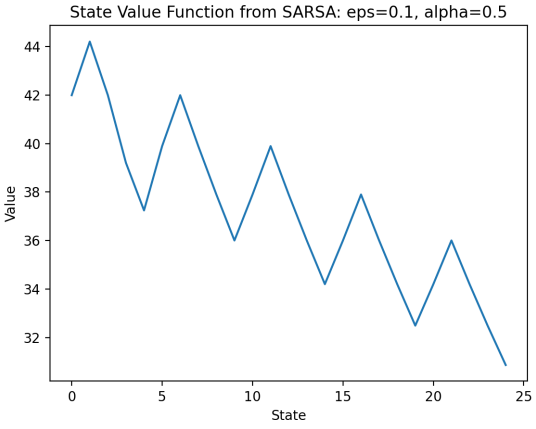


Figure 21:

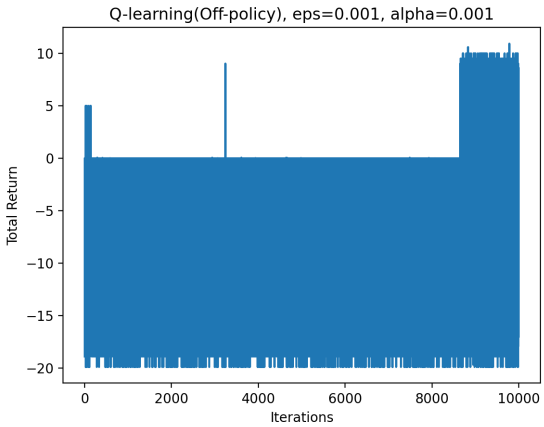


Figure 22:

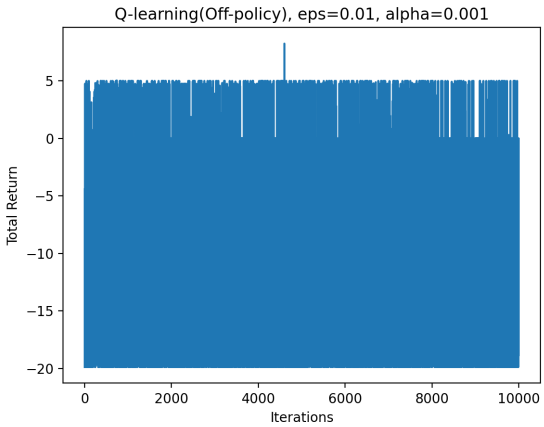


Figure 23:

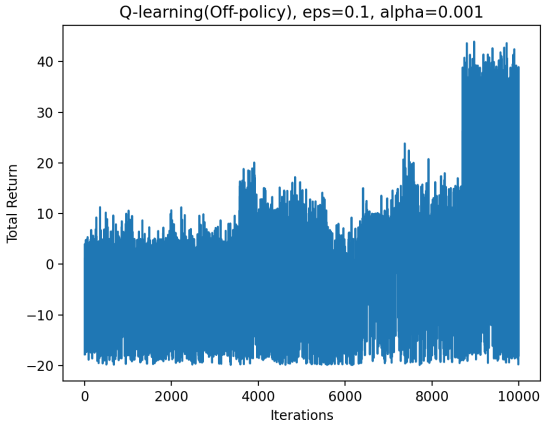


Figure 24:

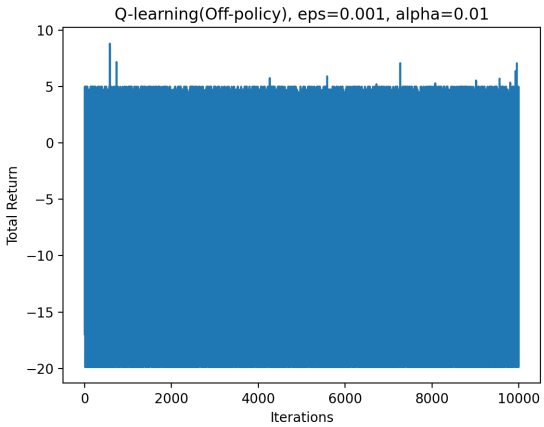


Figure 25:

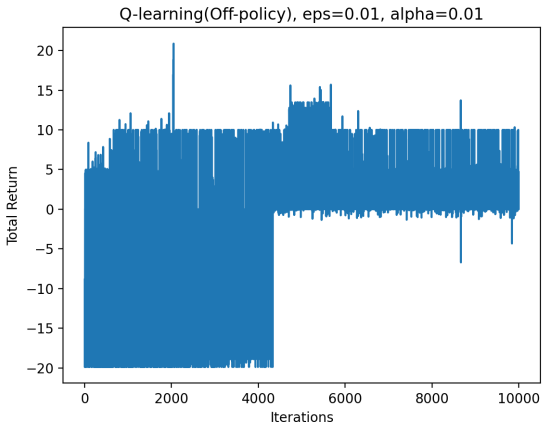


Figure 26:

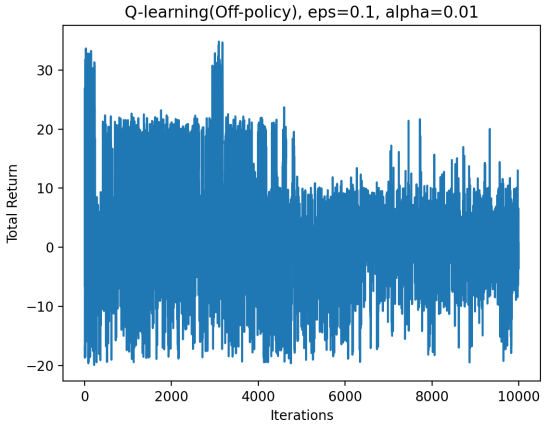


Figure 27:

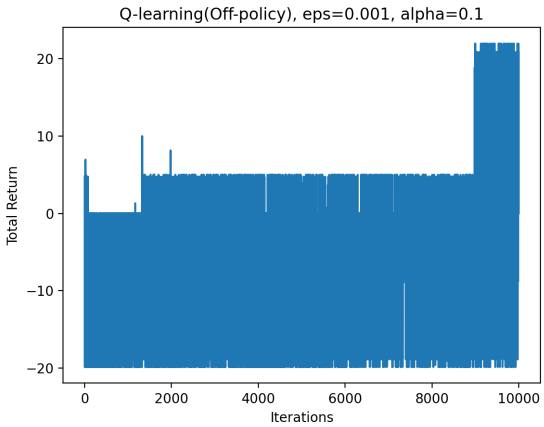


Figure 28:

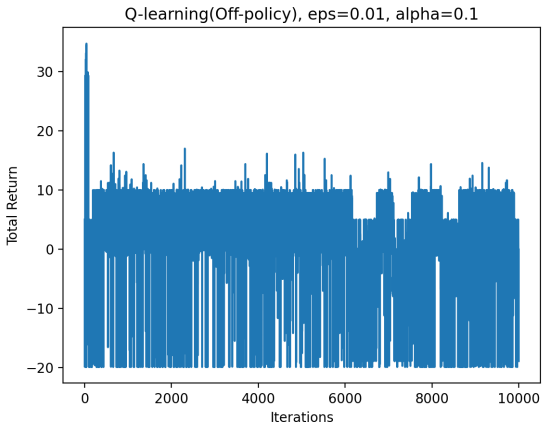


Figure 29:

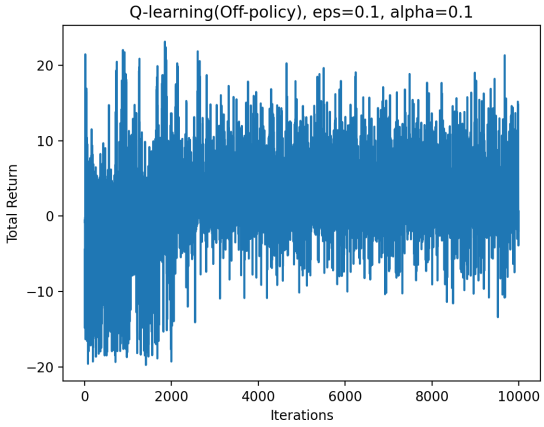


Figure 30:

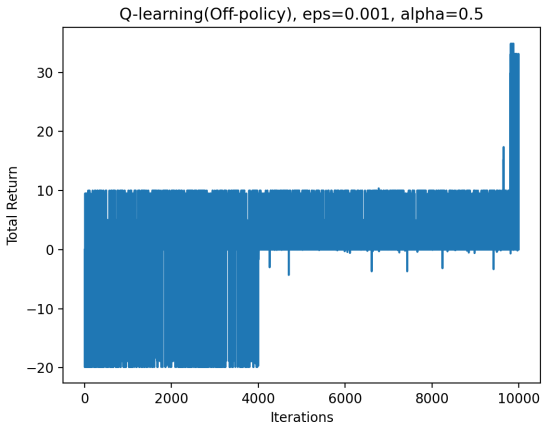


Figure 31:

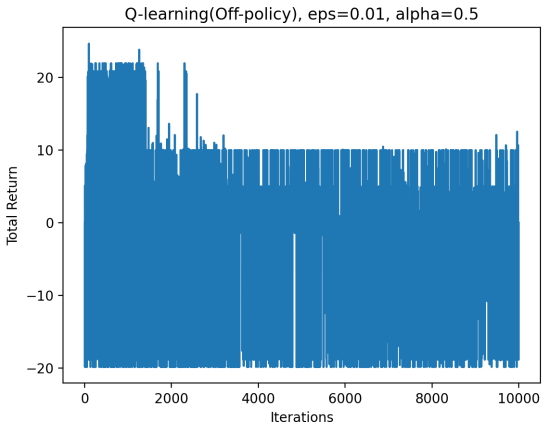


Figure 32:

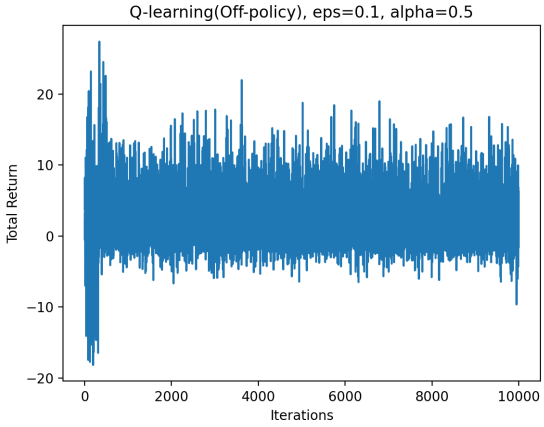


Figure 33:

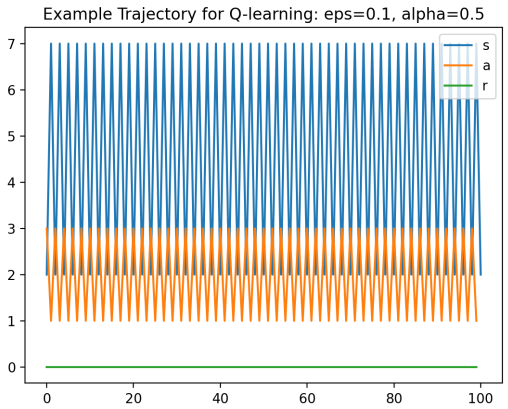


Figure 34:

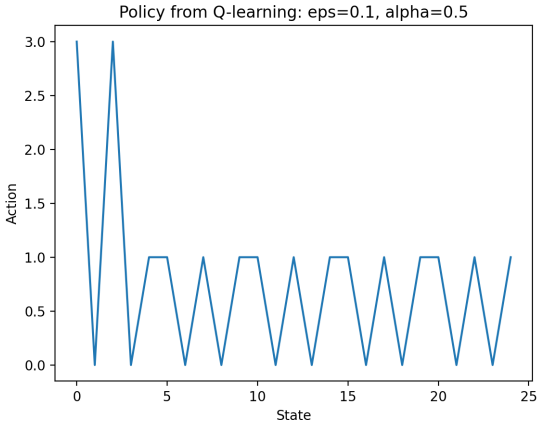


Figure 35:

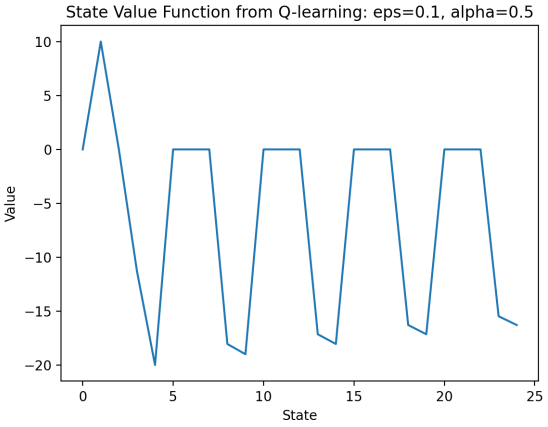


Figure 36:

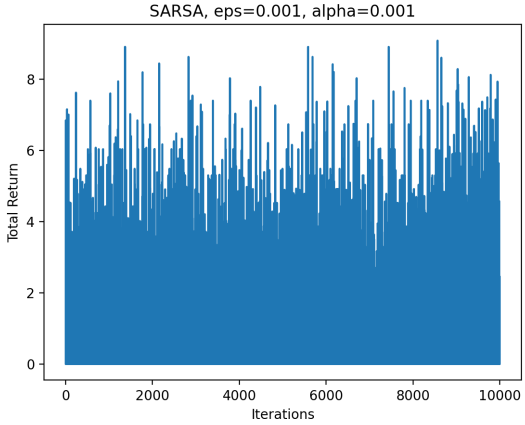


Figure 37:

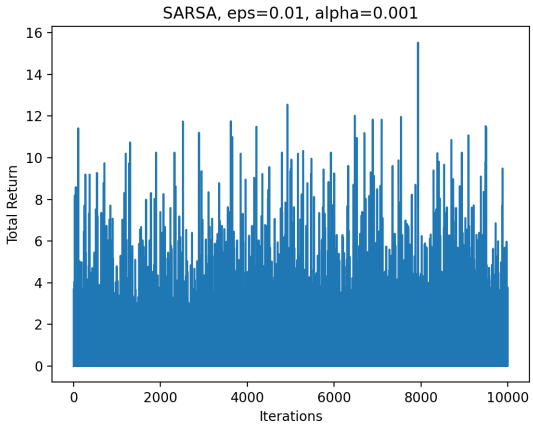


Figure 38:

# 3 Discrete-Pendulum

## 3.1 SARSA

Fig:39-51

## 3.2 Q-learning(off-policy)

Fig:52-66

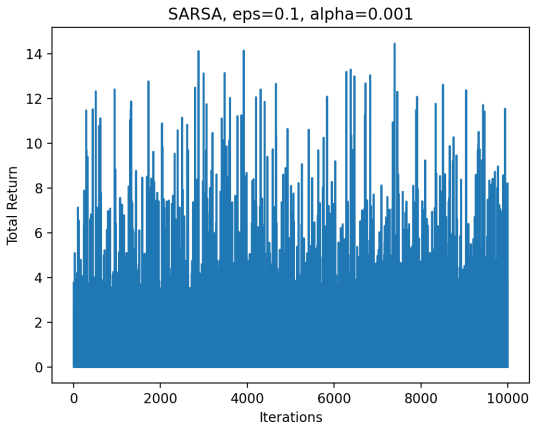


Figure 39:

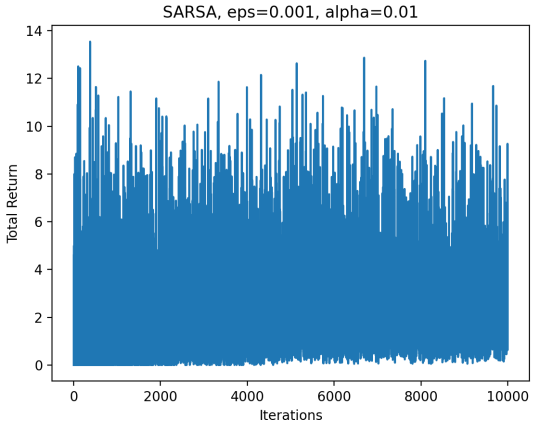


Figure 40:

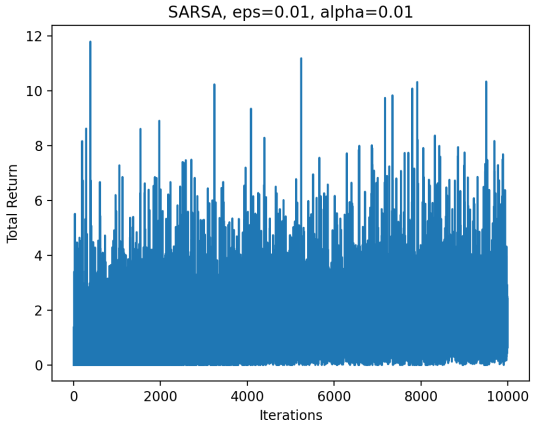


Figure 41:

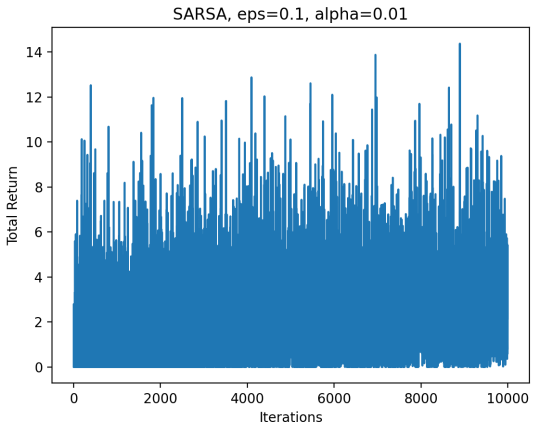


Figure 42:

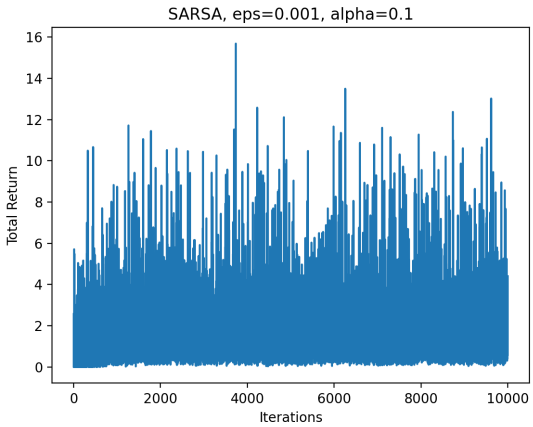


Figure 43:

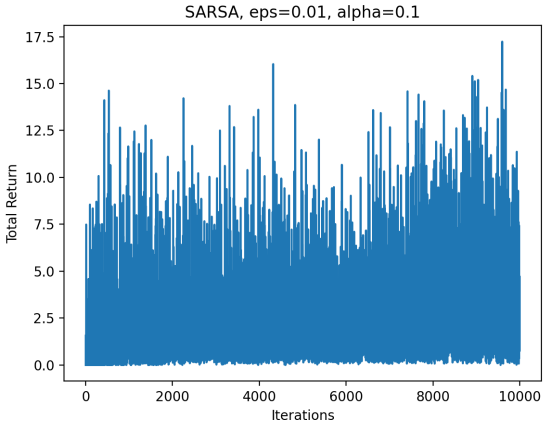


Figure 44:

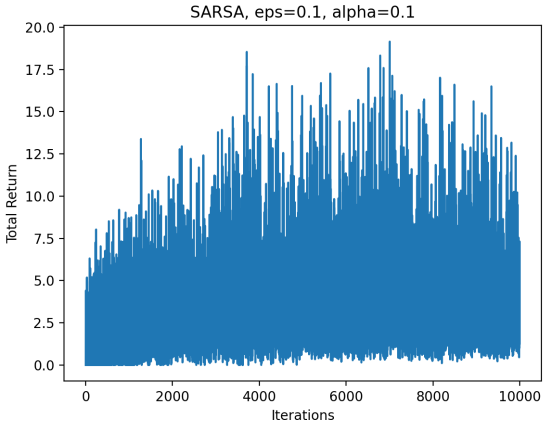


Figure 45:

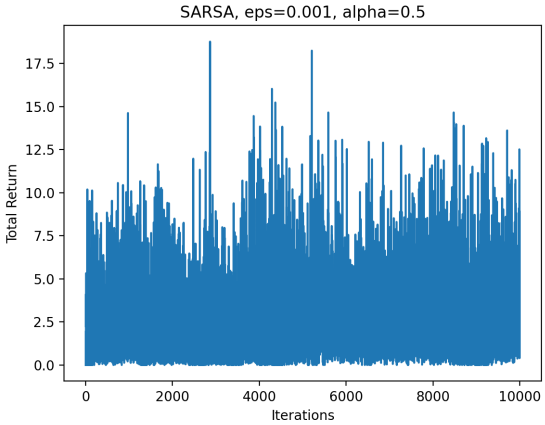


Figure 46:

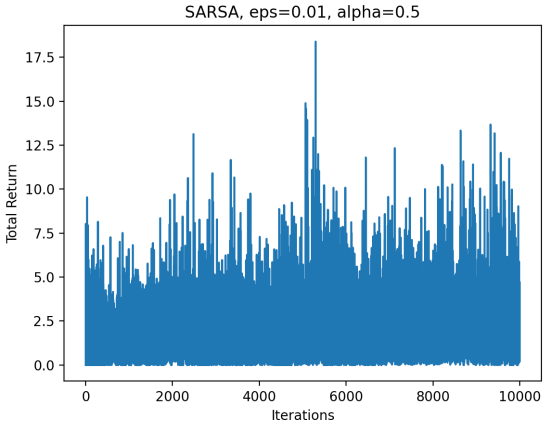


Figure 47:

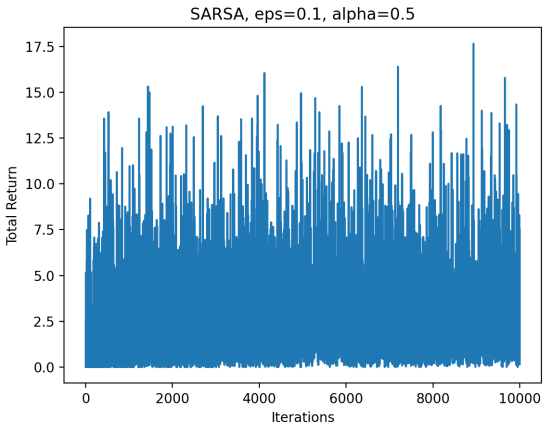


Figure 48:

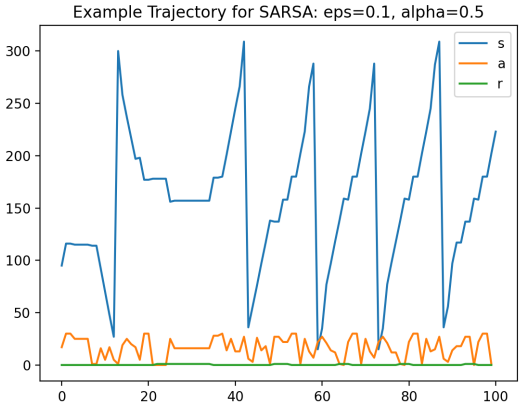


Figure 49:

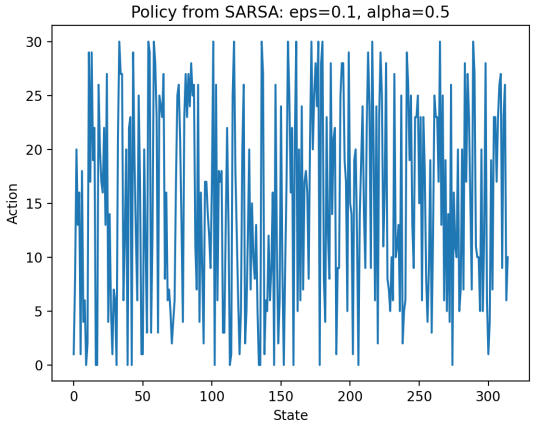


Figure 50:

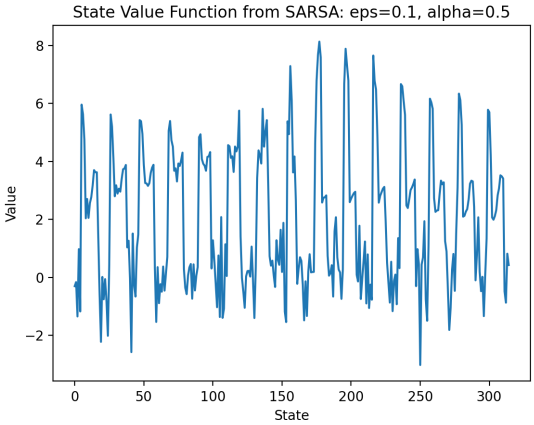


Figure 51:

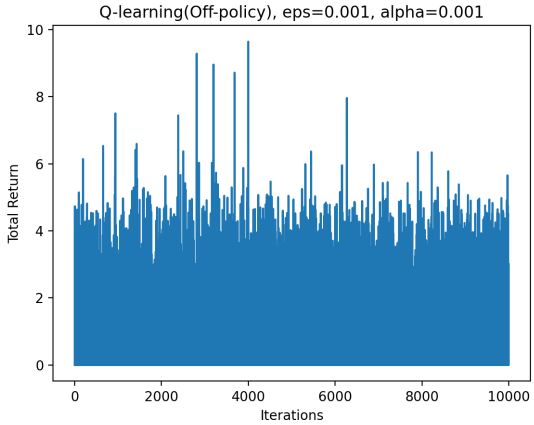


Figure 52:

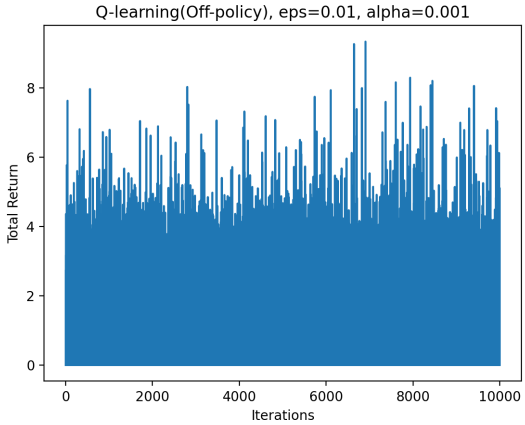


Figure 53:

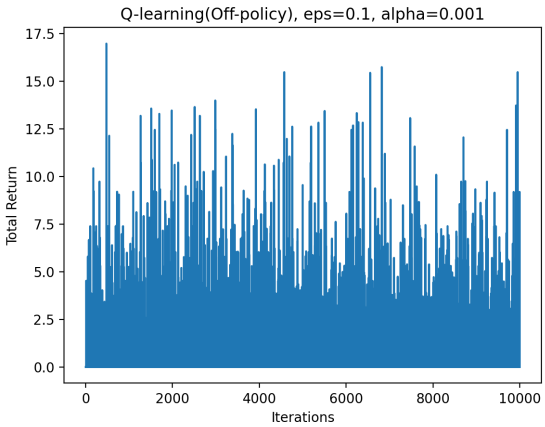


Figure 54:

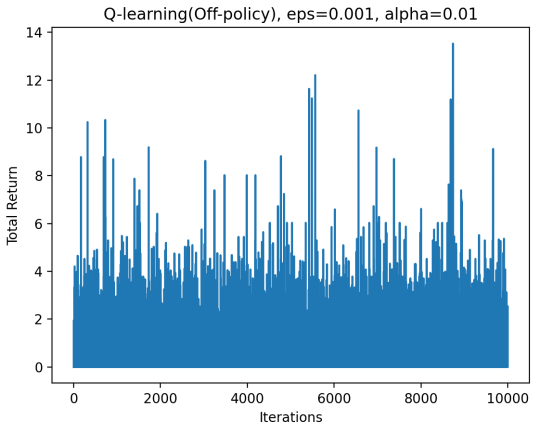


Figure 55:

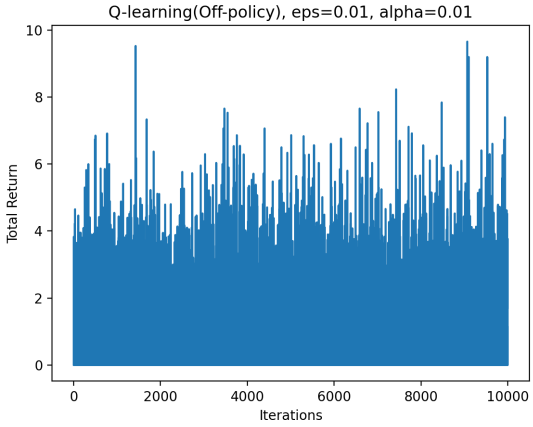


Figure 56:

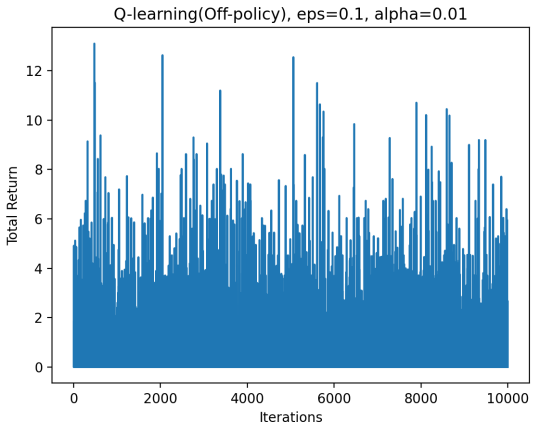


Figure 57:

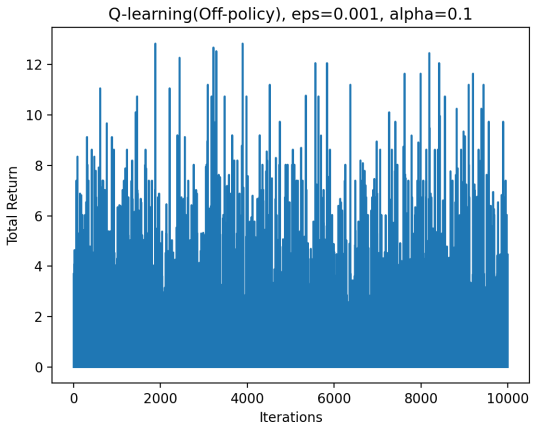


Figure 58:

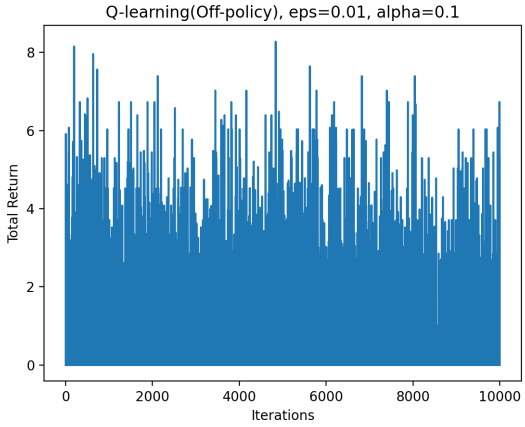


Figure 59:

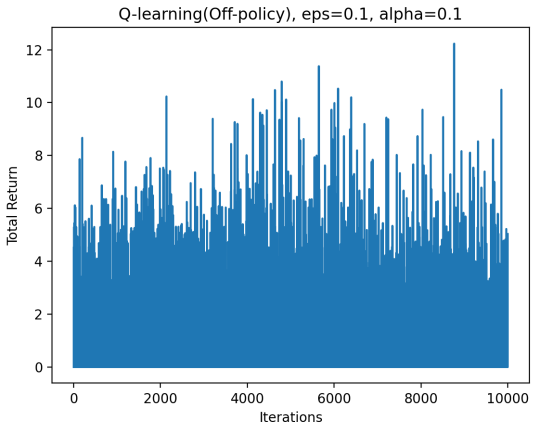


Figure 60:

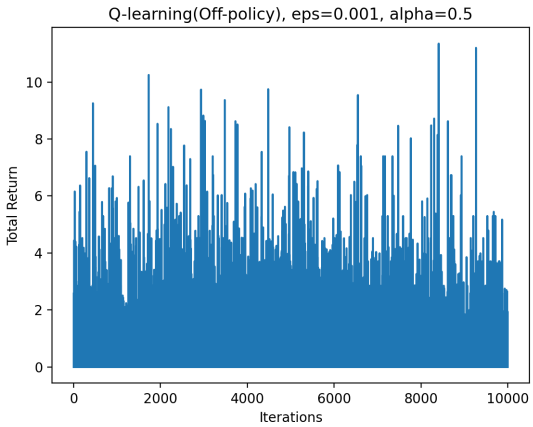


Figure 61:

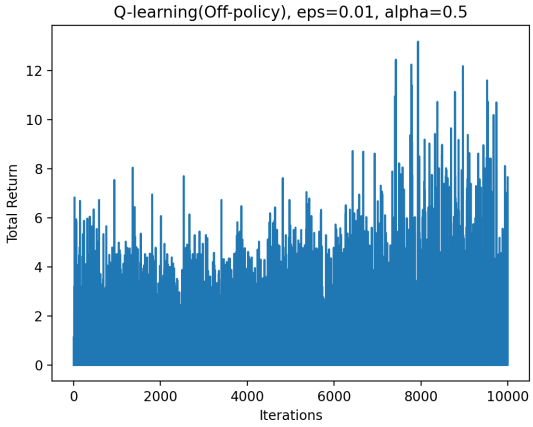


Figure 62:

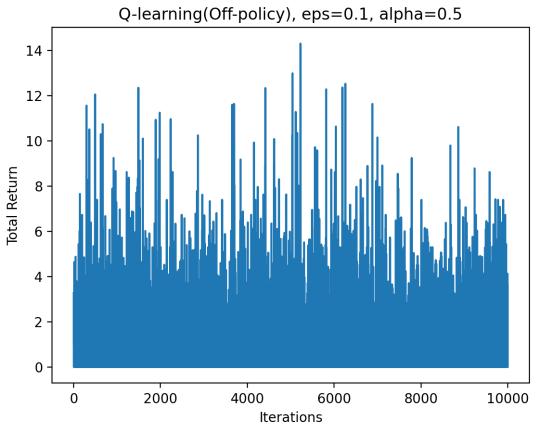


Figure 63:


Figure 64:

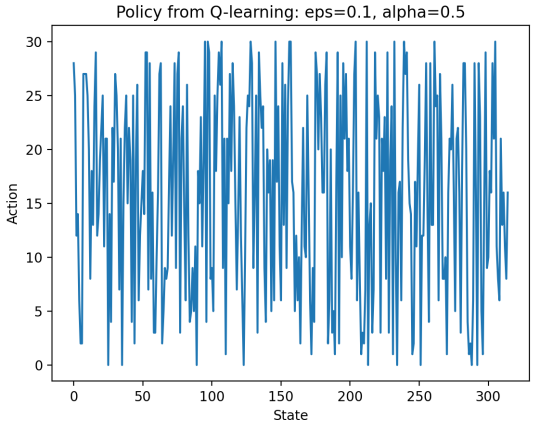


Figure 65:

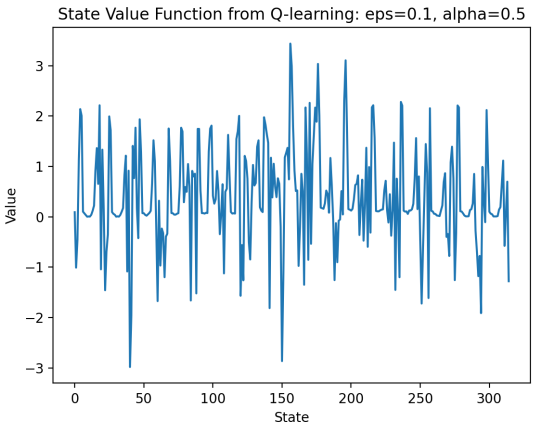
State Value Function from Q-learning: eps=0.1, alpha=0.5
Value
State
3
2
1
0
−1
−2
−3
0
50
100
150
200
250
300

Figure 66: