AE 598-RL HW1 Report

Gokul Puthumanaillam (gokulp2
Department of Aerospace Engineering, University of Illinois Urbana-Champaign

I. Introduction

This report documents the generated graphs for the following algorithms in the Grid World and the Discrete Pendulum (in a tabular setting):

1) Value Iteration: Grid World
2) Policy Iteration: Grid World
3) SARSA: Grid World and Discrete Pendulum
4) QLearning: Grid World and Discrete Pendulum

II. Hyperparameters:

A. Grid World

1. Value Iteration and Policy Iteration

1) Max Number of episodes: 5000
2) γ: 0.95
3) θ=1e-6

2. SARSA

1) Max Number of episodes: 5000
2) γ: 0.95
3) θ: 1e-6
4) ϵ: 0.1

3. QLearning

1) Max Number of episodes: 5000
2) γ: 0.95
3) θ: 1e-6
4) ϵ: 0.1

B. Discrete Pendulum

1. SARSA

1) Max Number of episodes: 700
2) γ: 0.95
3) θ: 1e-6
4) ϵ: 0.1

2. QLearning

1) Max Number of episodes: 700
2) γ: 0.95
3) θ: 1e-6
4) ϵ: 0.1

III. Plots

A. Grid World

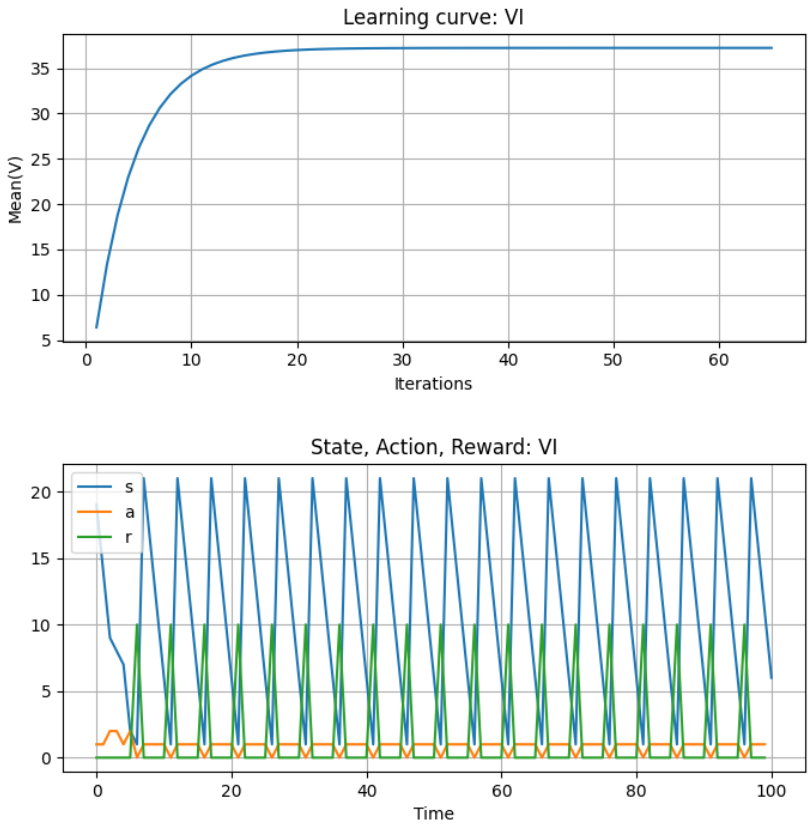

Fig. 1 Plots showing the learning curve for Value Iteration and the State, Action, Rewards

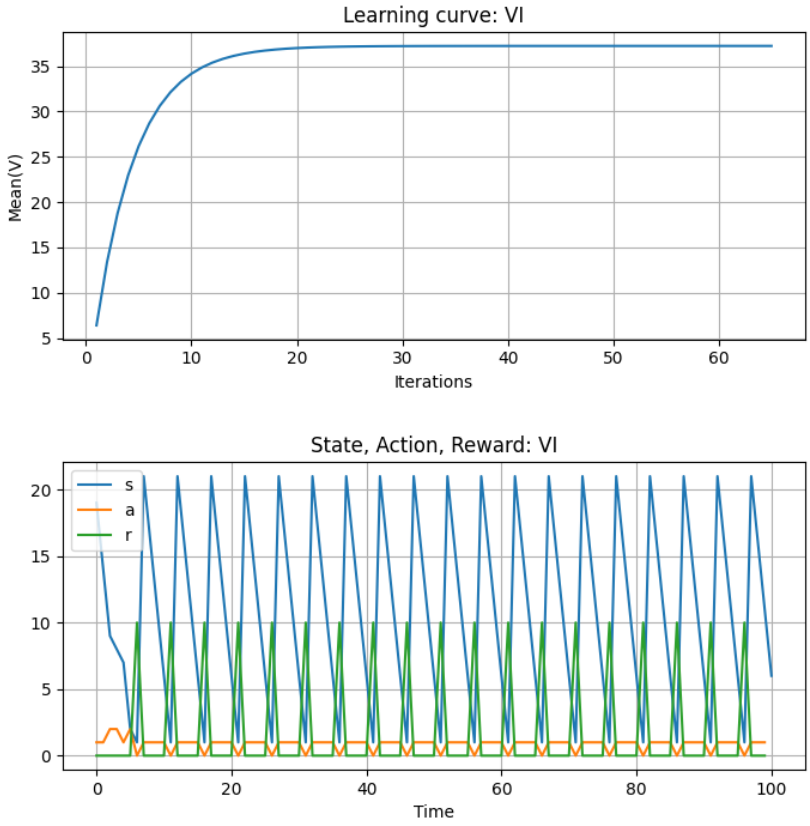

Fig. 2 Plots showing the learning curve for Policy Iteration and the State, Action, Rewards

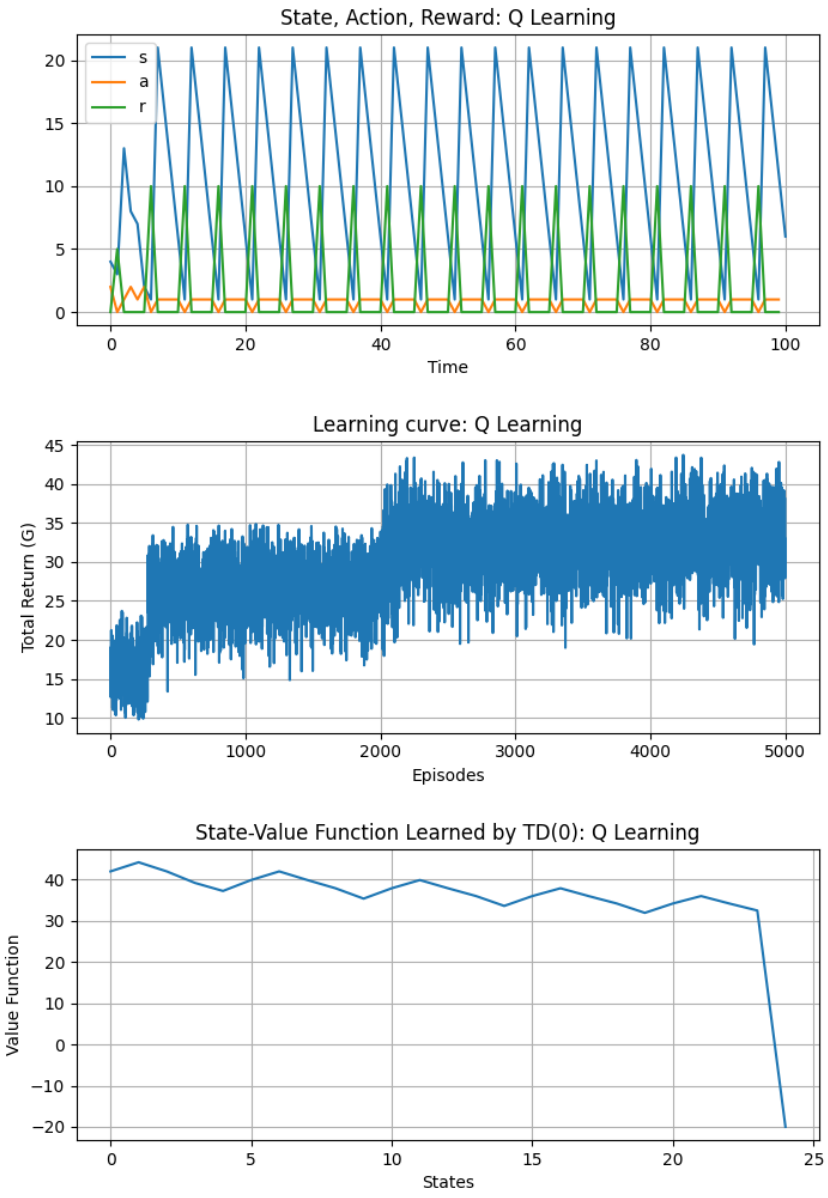

Fig. 3 Plots for QLearning: (a)State, Action, Reward (b) Learning Curve (c) State Value Function Estimation using TD(0)

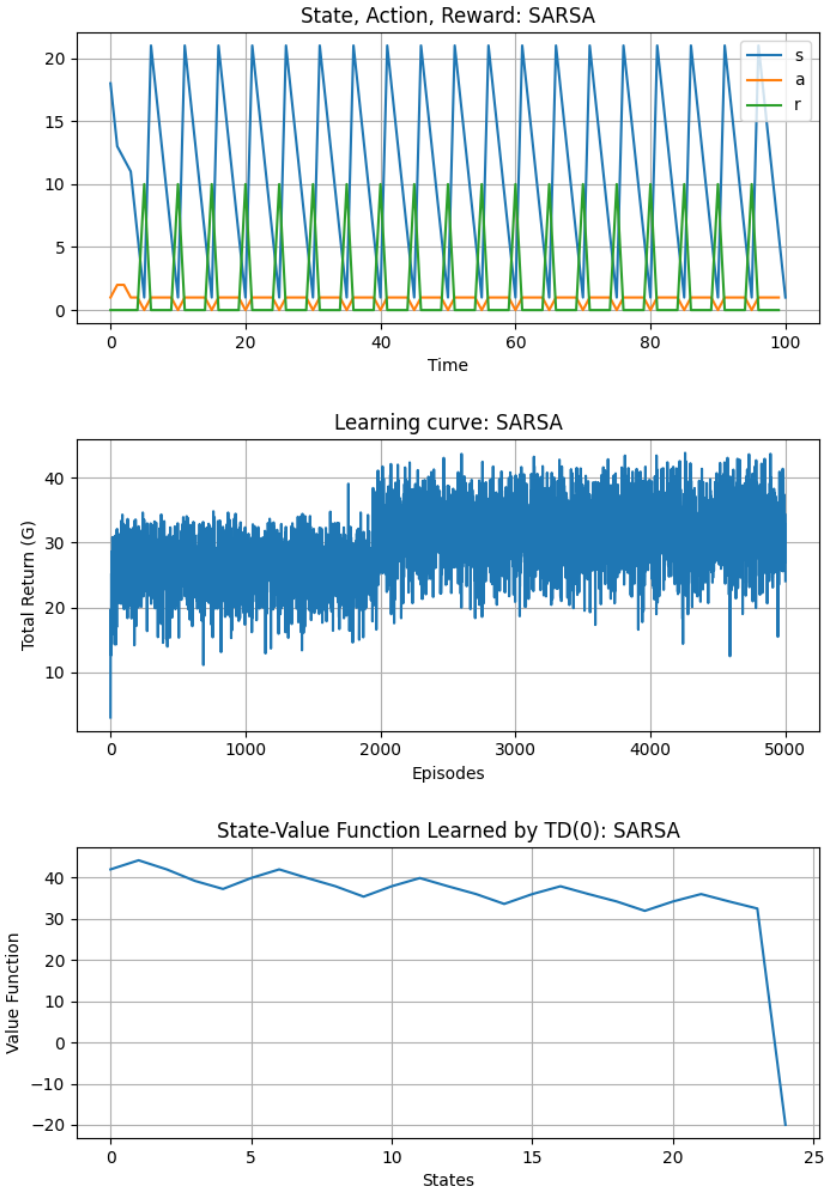

Fig. 4 Plots for SARSA: (a)State, Action, Reward (b) Learning Curve (c) State Value Function Estimation using TD(0)

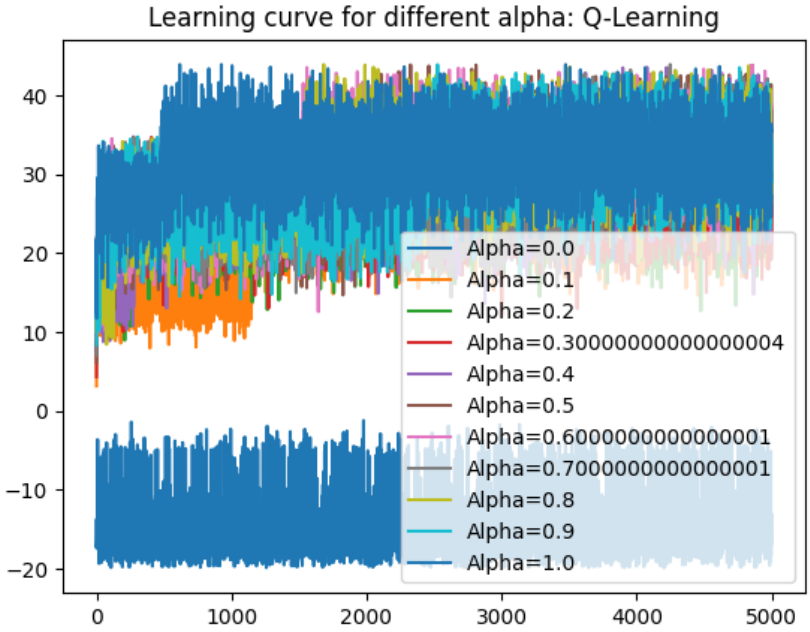

Fig. 5 Learning Curve for different Alpha: QLearning

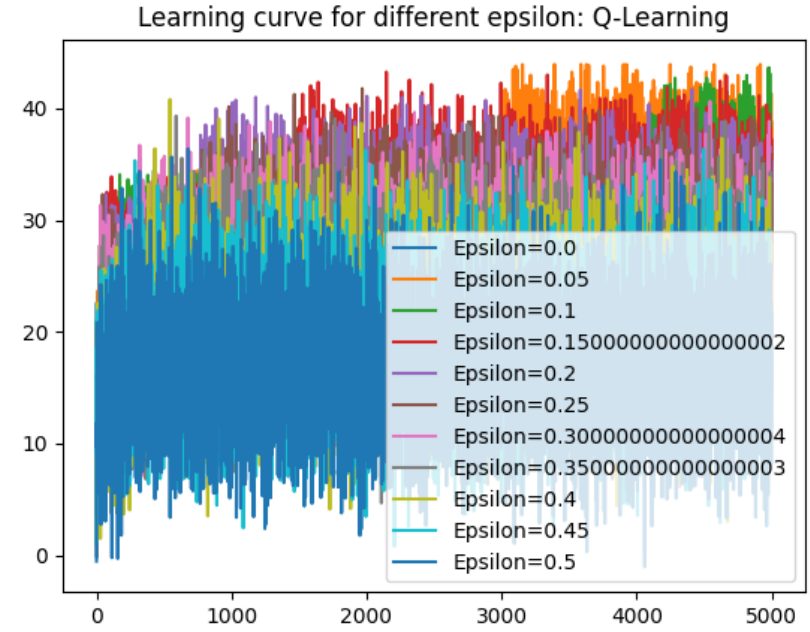

Fig. 6 Learning Curve for different Epsilon: QLearning

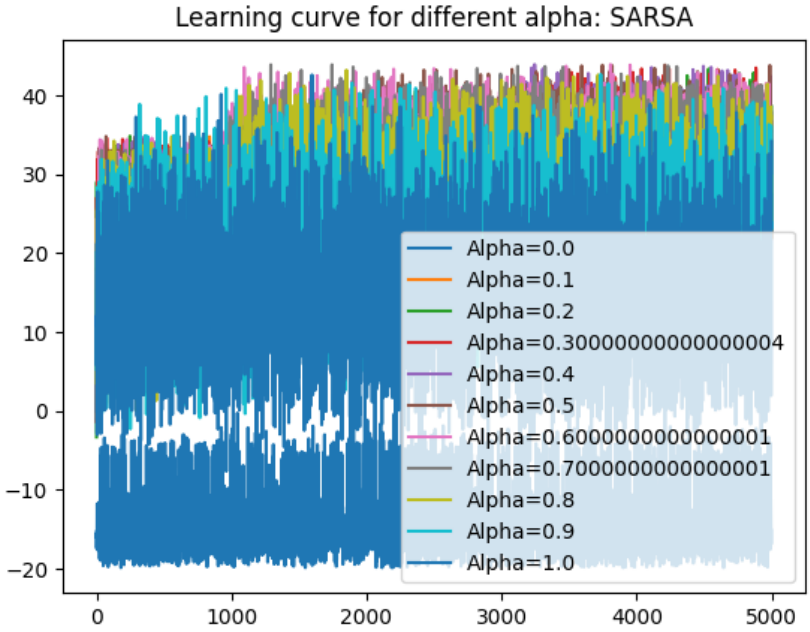

Fig. 7 Learning Curve for different Alpha: SARSA

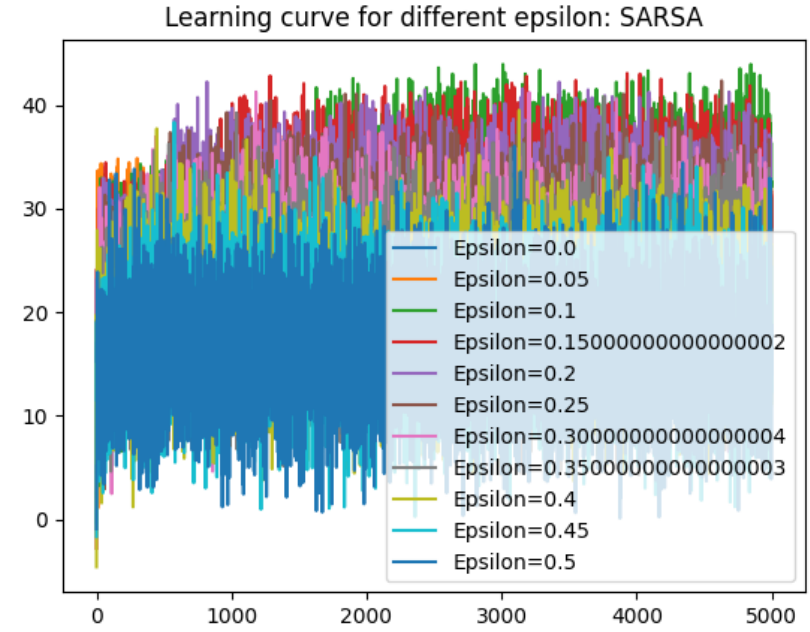

Fig. 8 Learning Curve for different Epsilon: SARSA

B. Discrete Pendulum

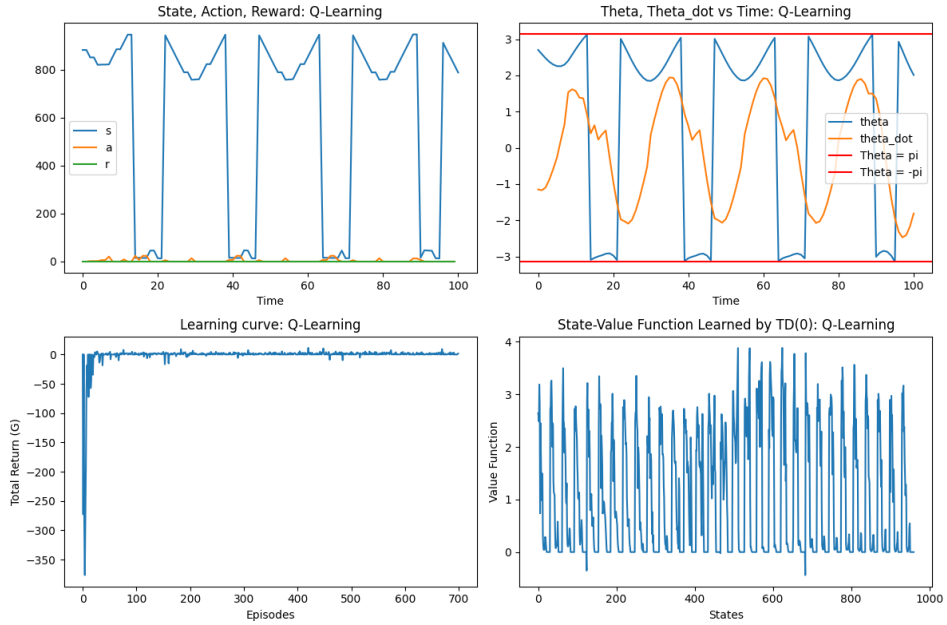

Fig. 9 Plots for QLearning: (a)State, Action, Reward (b) Theta, Theta dot vs time(c) Learning Curve (d) State Value Function Estimation using TD(0)

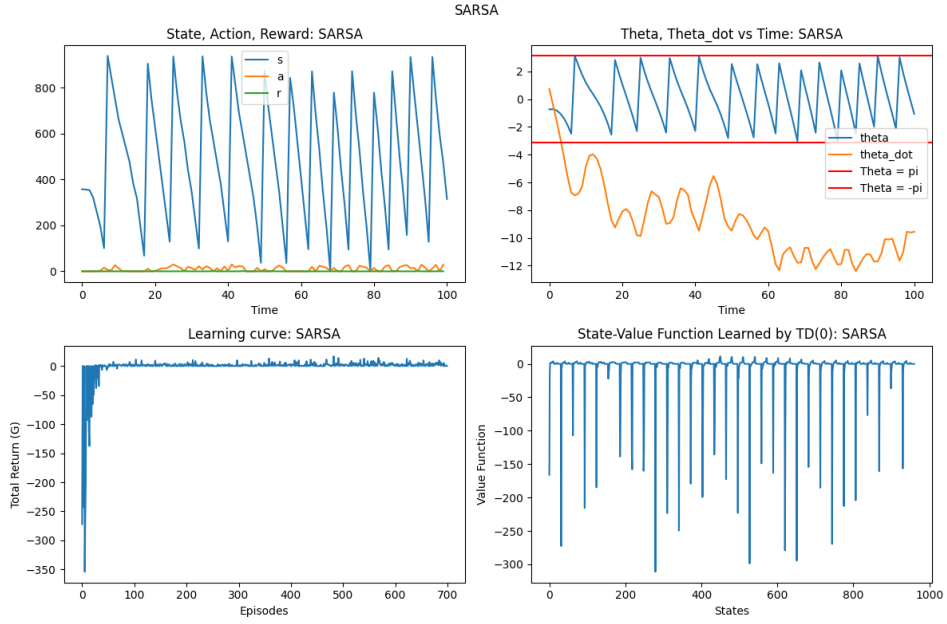

Fig. 10 Plots for SARSA: (a)State, Action, Reward (b) Theta, Theta dot vs time(c) Learning Curve (d) State Value Function Estimation using TD(0)

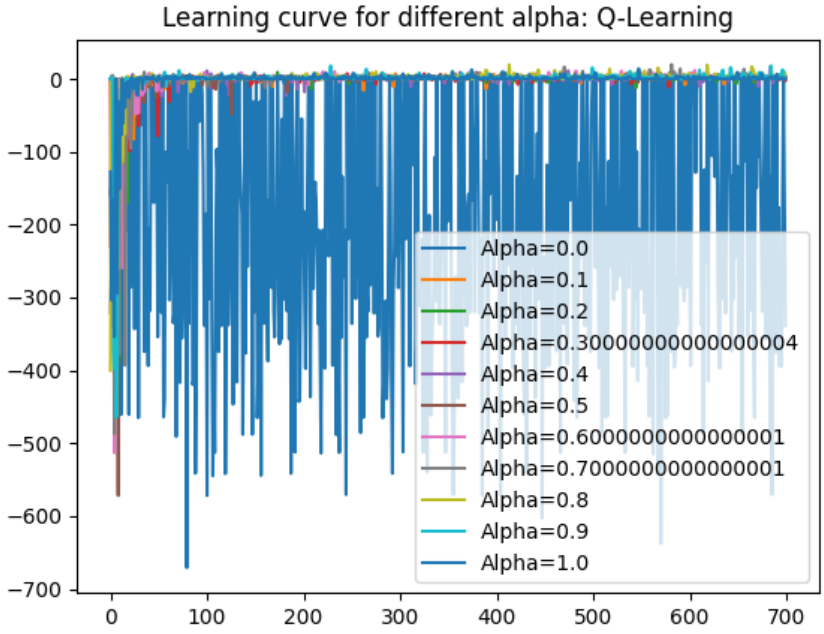

Fig. 11 Learning Curve for different Alpha: QLearning

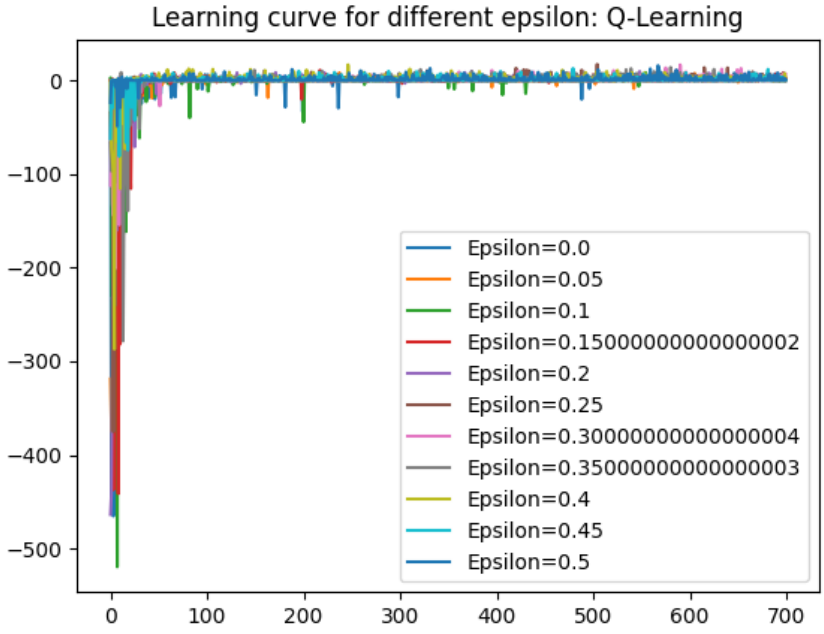

Fig. 12 Learning Curve for different Epsilon: QLearning

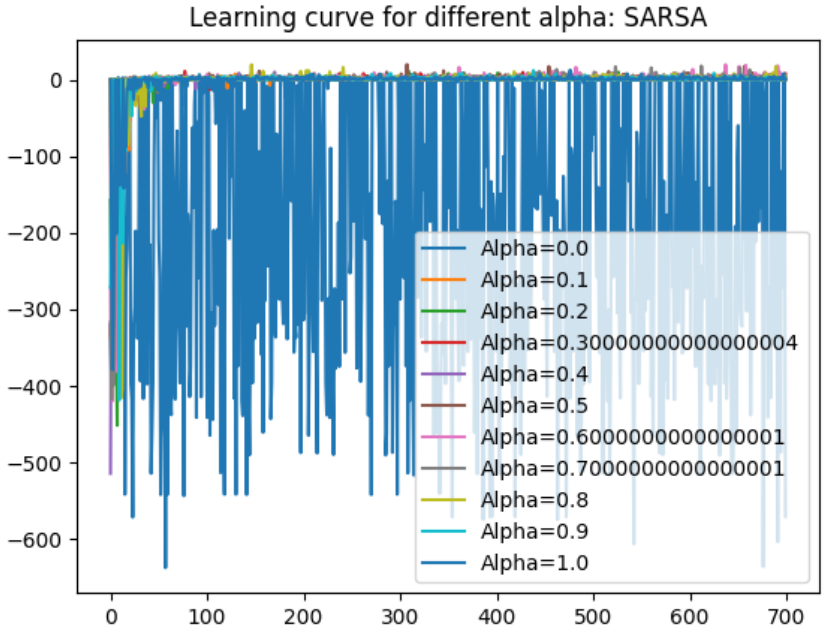

Fig. 13 Learning Curve for different Alpha: SARSA

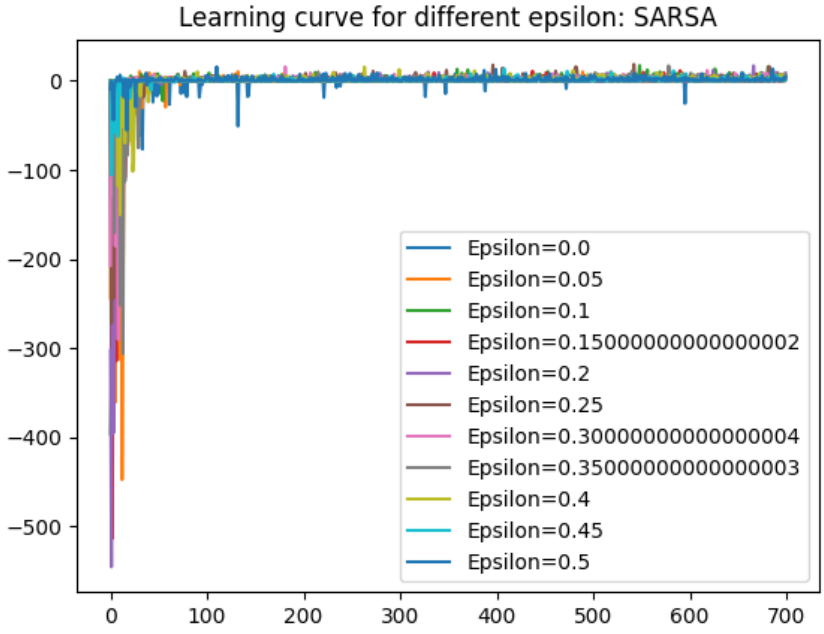

Fig. 14 Learning Curve for different Epsilon: SARSA