DA6400 : Reinforcement Learning
Programming Assignment #1
Report

Ritabrata Mandal
EE24E009

March 26, 2025

Contents

1 Introduction

1.1 Environments

In this programming task, we are utilize the following Gymnasium environments for training and evaluating your policies. The links associated with the environments contain descriptions of each environment.

- CartPole-v1 : A pole is attached by an un-actuated joint to a cart, which moves along a frictionless track. The pendulum is placed upright on the cart and the goal is to balance the pole by applying forces in the left and right direction on the cart.
- MountainCar-v0 : The Mountain Car MDP is a deterministic MDP that consists of a car placed stochastically at the bottom of a sinusoidal valley, with the only possible actions being the accelerations that can be applied to the car in either direction. The goal of the MDP is to strategically accelerate the car to reach the goal state on top of the right hill. There are two versions of the mountain car domain in gymnasium: one with *discrete actions* and one with *continuous*. This version is the one with discrete actions.
- MiniGrid-Dynamic-Obstacles-5x5-v0 : This environment is an empty room with moving obstacles. The goal of the agent is to reach the green goal square without colliding with any obstacle. A large penalty is subtracted if the agent collides with an obstacle and the episode finishes. This environment is useful to test Dynamic Obstacle Avoidance for mobile robots with Reinforcement Learning in Partial Observability.

1.2 Algorithms

Training each of the below algorithms and assessing their comparative performance.

- **SARSA**→ with ϵ-**greedy exploration**
- **Q-Learning**→ with **Softmax exploration**

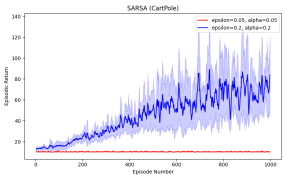
2 Implementation

2.1 CartPole-v1

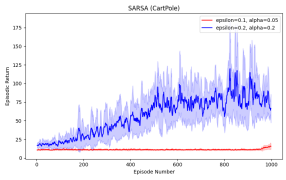
2.1.1 SARSA Code Snippets

2.1.2 SARSA Runs

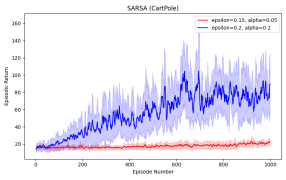
below are the graphs of runs of SARSA with different hyper-parameter



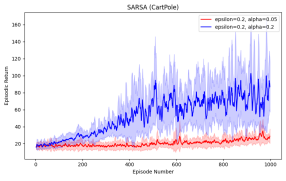
(a) $\alpha = 0.05$



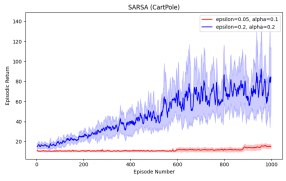
(b) $\epsilon = 0.1$



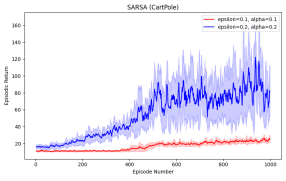
(c) $\epsilon = 0.15$



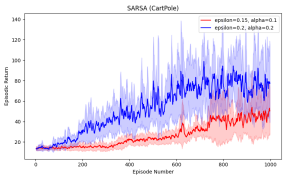
(d) $\epsilon = 0.2$



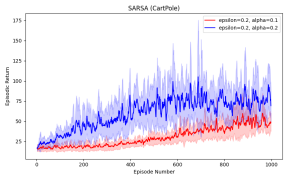
(e) $\epsilon = 0.25$



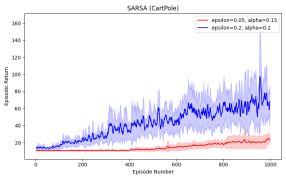
(f) $\epsilon = 0.3$



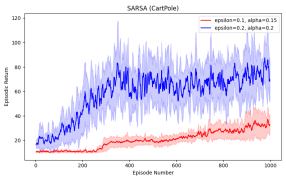
(g) $\epsilon = 0.35$



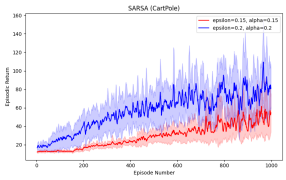
(h) $\epsilon = 0.4$



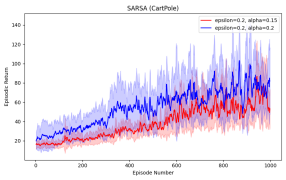
(i) $\epsilon = 0.45$



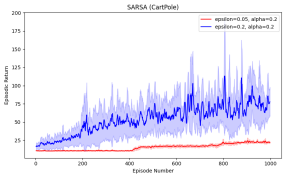
(j) $\epsilon = 0.5$



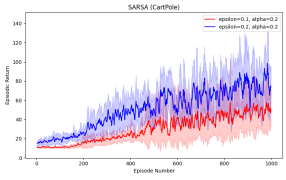
(k) $\epsilon = 0.55$



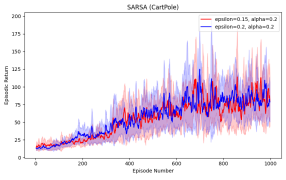
(l) $\epsilon = 0.6$



(m) $\epsilon = 0.65$



(n) $\epsilon = 0.7$



(o) $\epsilon = 0.75$

Figure 1: SARSA(Cartpole) with different ϵ and α