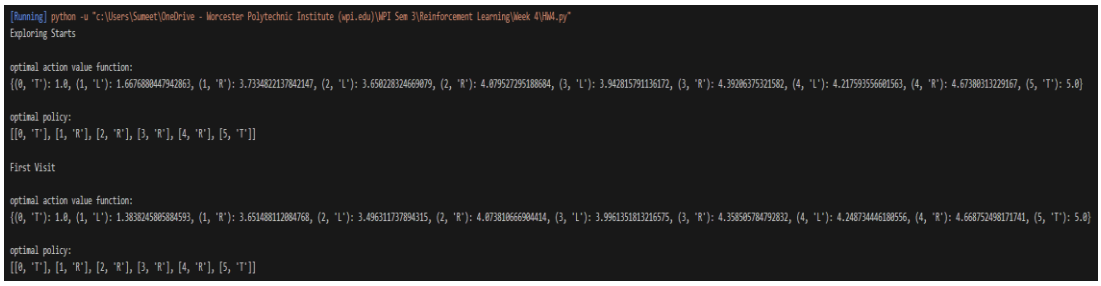
Sumeet Shanbhag

Student ID : 641020714

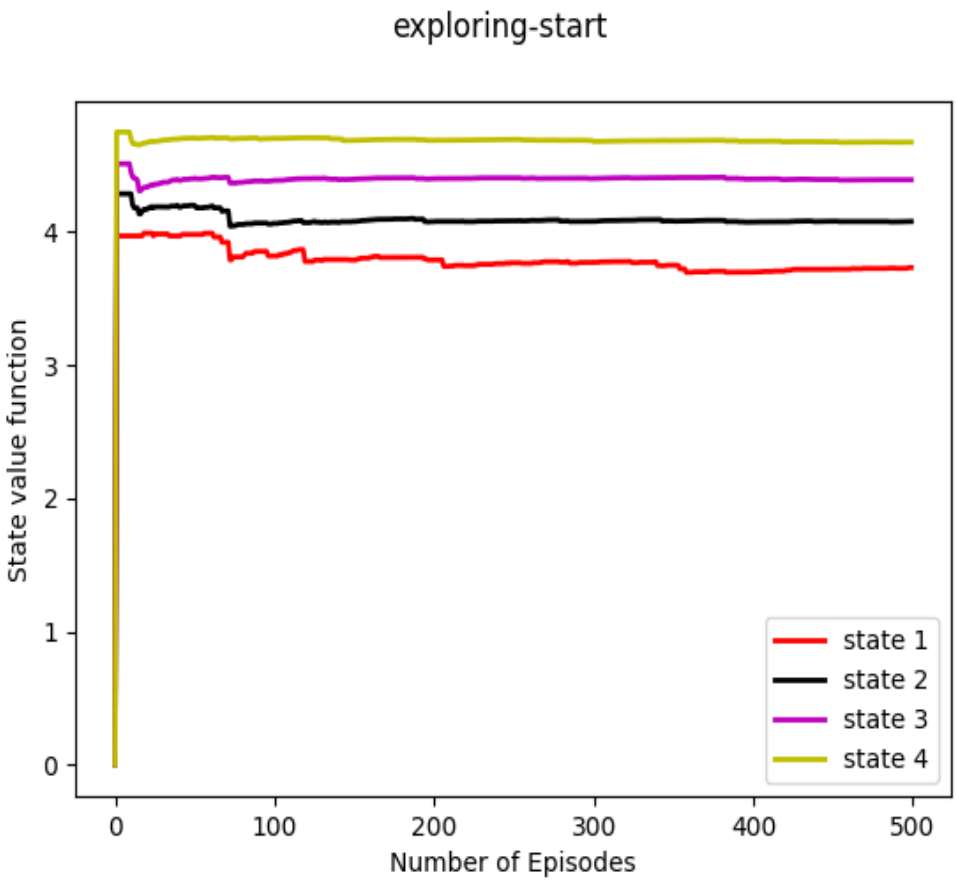
Reinforcement Learning Monte Carlo Programming Assignment

Output:

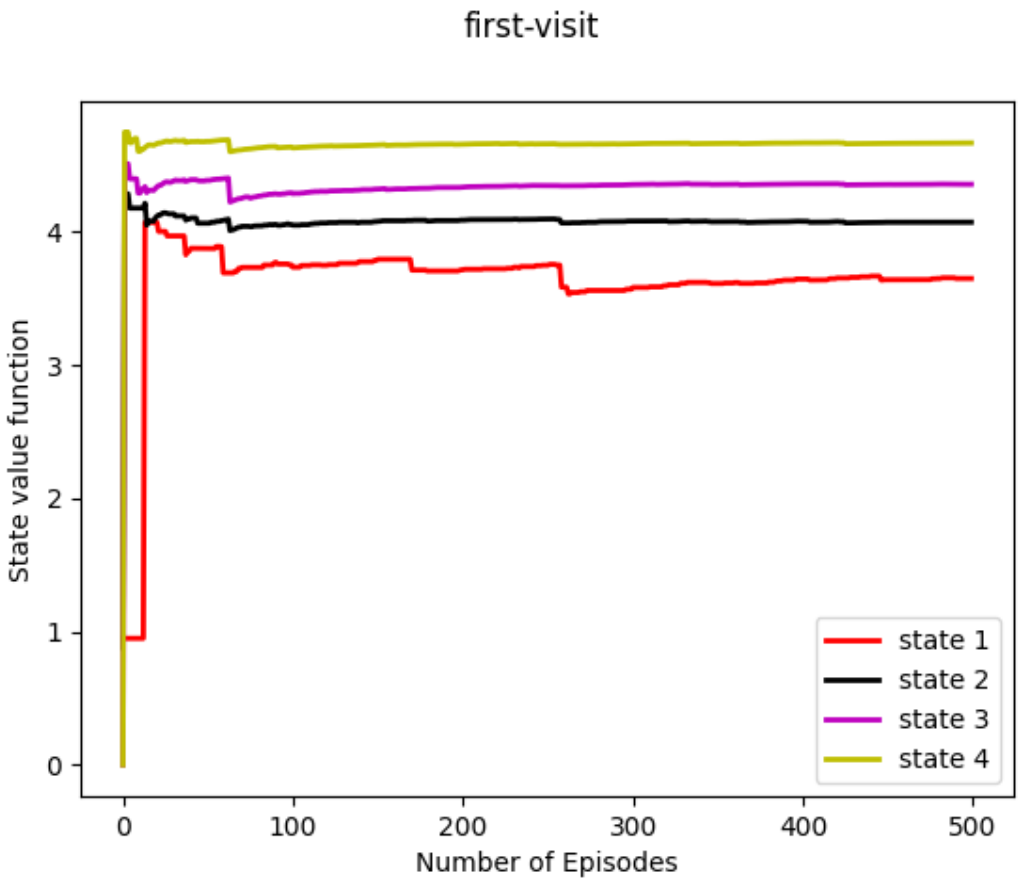
Below is the output for optimal action value functions and optimal policies using the exploring start and first visit methods after running for 500 episodes.



The figure below shows the estimation of the state value function vs the number of episodes it takes for Monte Carlo method of exploring starts



The figure below shows the estimation of the state value function vs the number of episodes it takes for Monte Carlo method of On-policy first visit



Convergence:

Exploring starts method ensures every state-action pair is sampled at least once. This guarantees complete exploration of the environment, allowing for faster and more accurate convergence of state value estimations, especially when compared to methods that might miss certain state-action pairs.

The first-visit Monte Carlo method estimates state values by averaging returns from the first occurrence of each state in episodes. While it provides a consistent estimate as the number of episodes increases, its convergence can be slower, especially for states that are infrequently visited.