

Assignment-6

2023-02-19

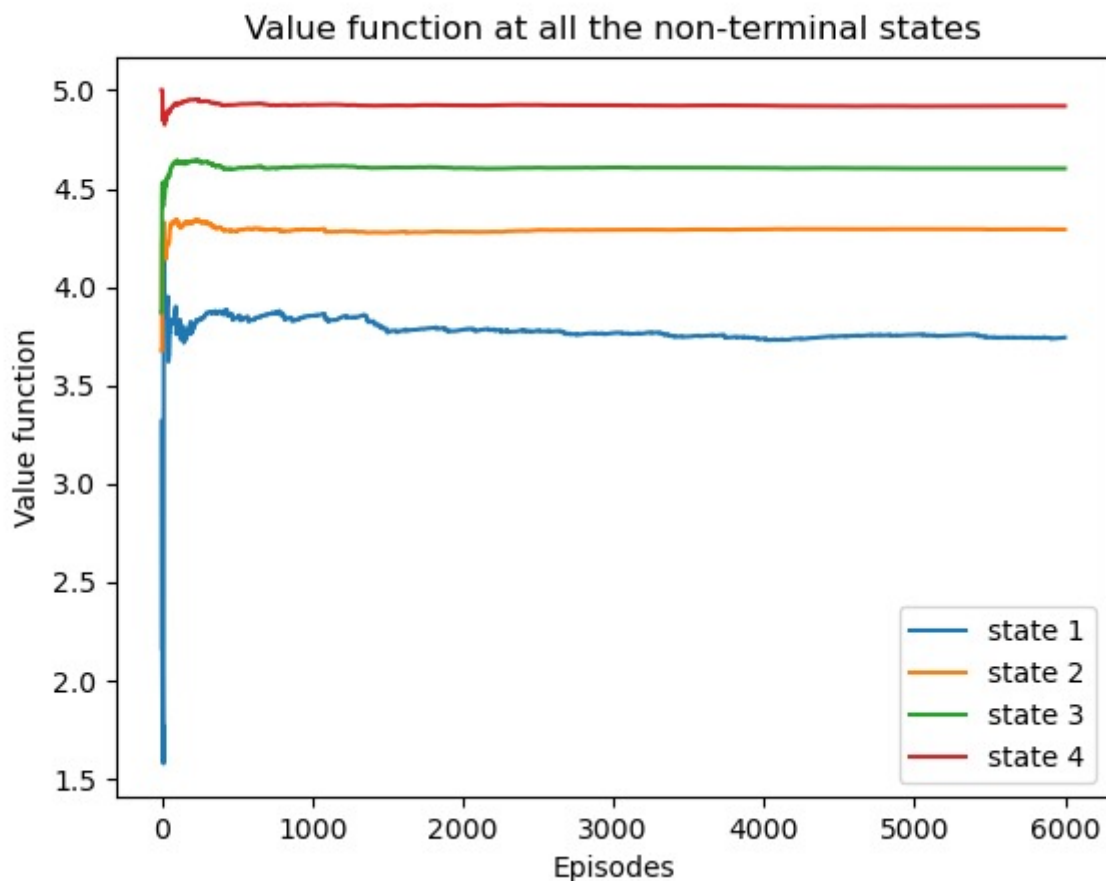
Name: Varun Gampa - 773686296, Rohin Siddhartha - 808068806

ID: vgampa, rvenkateswaran

Two methods Monte Carlo ES (Exploring Starts), and On-policy first-visit MC control were used to find the state action value function along with the optimal policy. Using Q and the policy, state value function is found.

Monte Carlo ES (Exploring Starts)

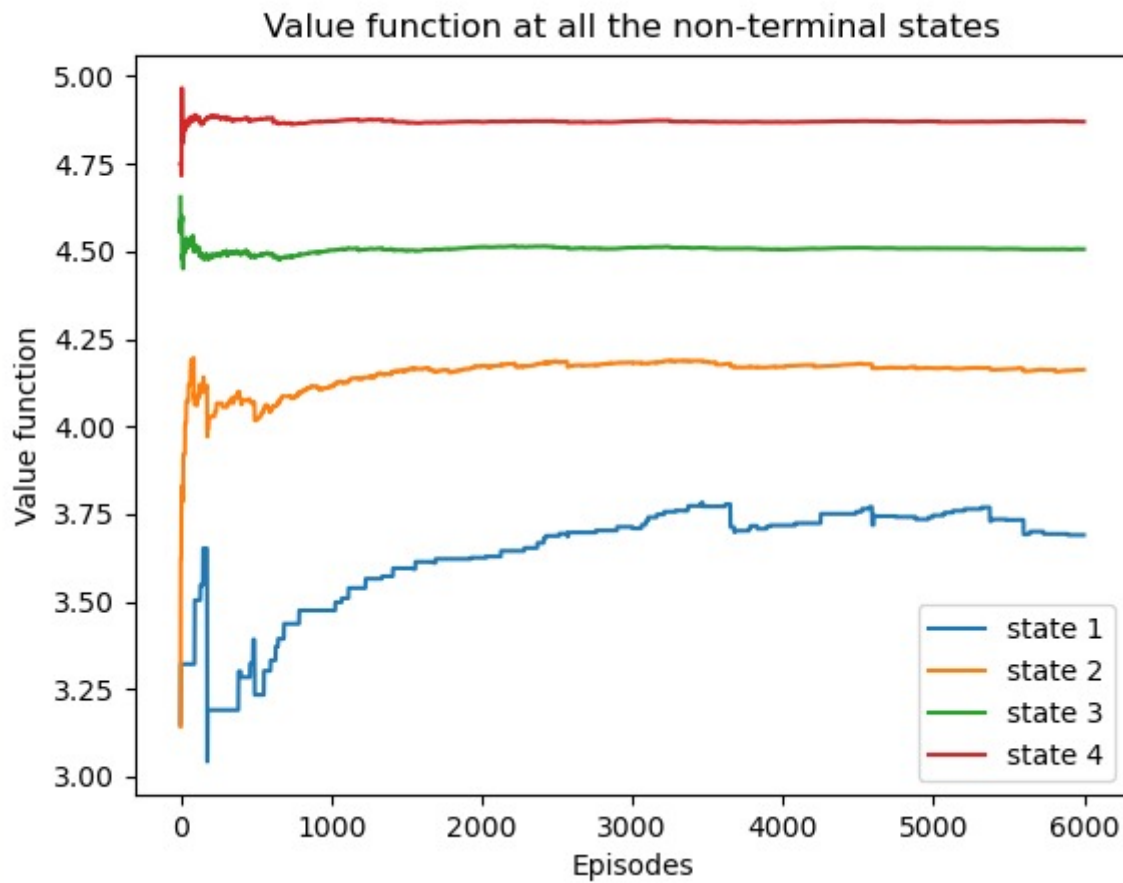
Policy is argmax of actions over state action value function for exploring starts, after each episode, as the state action value function is updated, policy is also updated. A total of 6000 episodes are run and state value function is recorded after each episode which is plotted.



On-policy first-visit MC

Policy is argmax of actions over state action value function with a probability of 1-epsilon or any action with a probability of epsilon. After each episode, as the state action value function is updated, policy is also updated.

A total of 6000 episodes are run and state value function is recorded after each episode which is plotted.



Note : The method for calculating V , used is:

$$v_*(s) = \sum_{a \in A(s)} \pi(a|s) \cdot q_\pi(s, a)$$