Assignment 3 - Reinforcement Learning

Laksh Bhatia **604561**

Solution 2

I implemented the SARSA algorithm for learning the policy for the provided example. I decided to implement this algorithm because it's a very simple implementation of the Temporal difference learning method. SARSA takes into account what action it would perform after the current step. Its like a one step look ahead kind of system which is faster if the number of actions that we can perform in every state is low.
For SARSA the learning rate that I have used is 1/t where t is the number of states visited in the current episode. I am using an epsilon greedy method to choose the actions.

Solution 3

Average reward for learning from n episodes.
n in our case is 10 ,100 ,1000 ,10000
10: -8.1294
100: -2.1678
1000: -0.2239
10000: -0.0367

Solution 4

The speed of learning in the value iteration was much faster than what I got for SARSA learning method, since value iteration broke as soon as convergence is met whereas the RL method has to run for the given number of episodes which makes it more time consuming.