Reinforcement Learning

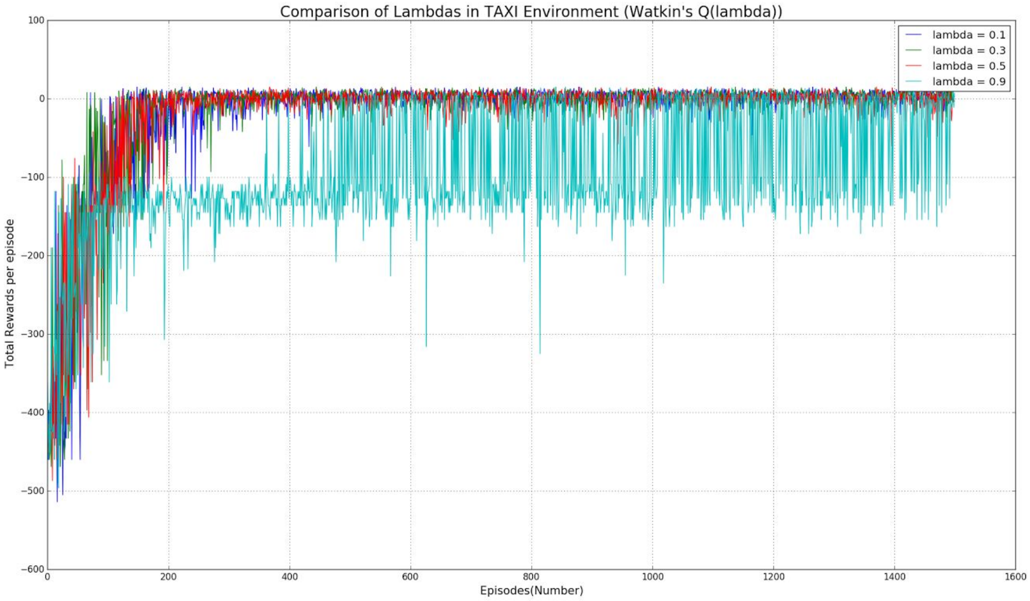
Project 3

TD Methods

Shivam Goel

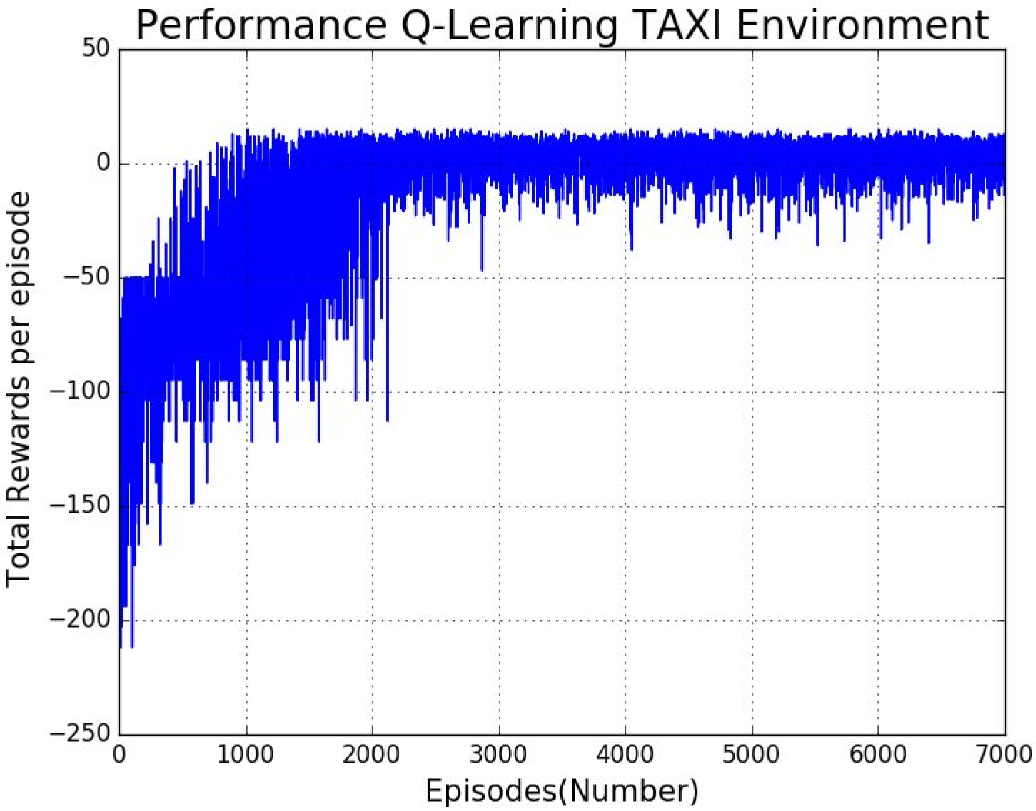
WSUID# 11483916

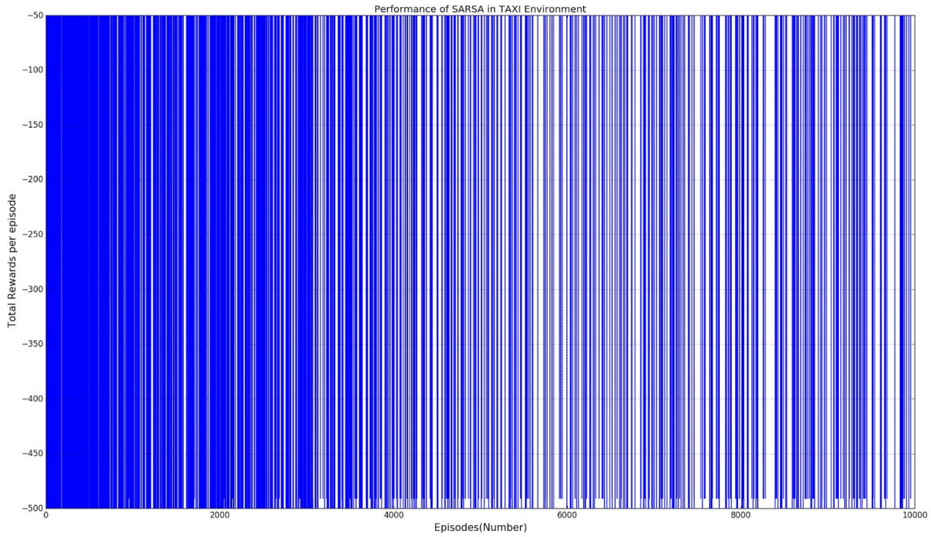
Implemented the Watkins Q(Lambda) algorithm (Eligibility Trace).
Tested the algorithm on Taxi (Deterministic)
Compared the various values of the Lambda, ranging from 0.2-0.
Implementation of the algorithm is in the code and that can be commented out to test in the various environments.
The graph is also saved in the folder for better viewing.
The graphs plotted are discussed below



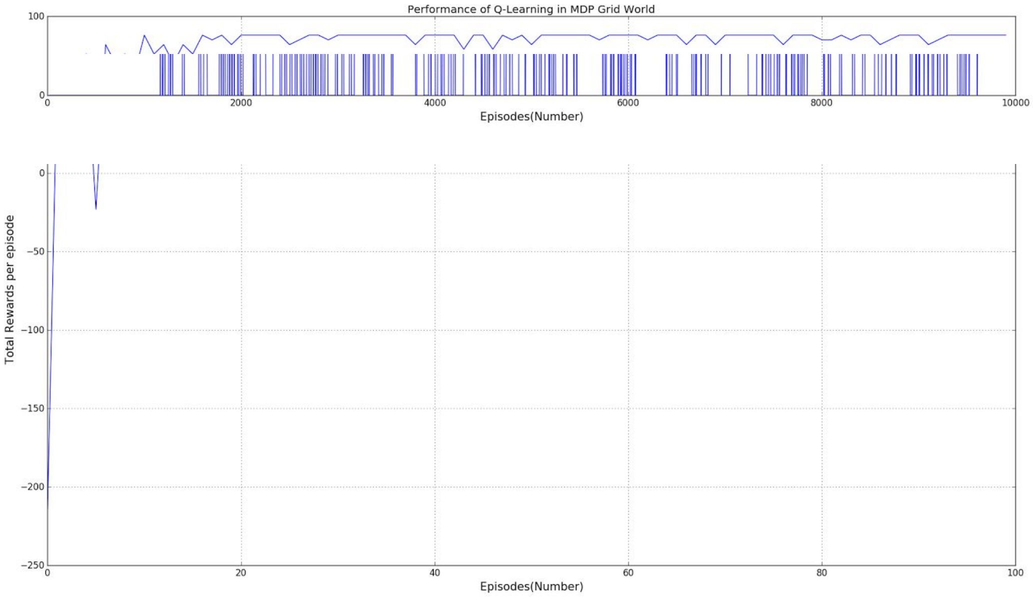
The above graphs shows the performance of the algorithm on the Taxi environment, the Average Rewards over the number of episodes have been plotted in the graph.. Also, various values of lambdas are compared, 0.1 (blue), 0.3 (green) and 0.5 (red) and 0.9 (cyan) to compare how well the algorithm performs on changing the lambda factor.It is seen that the algorithms performs better as we decrease the value of lambda, hence for higher lambda value it gives out poor average rewards as compared with lesser lambdas. Overall it was seen that the policy gets stable after 170 episodes

The two graph shows the performances of the Q learning and SARSA on same environment.
It shows that the Q learning learns a stable policy after close to 2000 episodes whereas in the SARSA the graph looks a little bit weird so it's not quite clearly understandable from it.

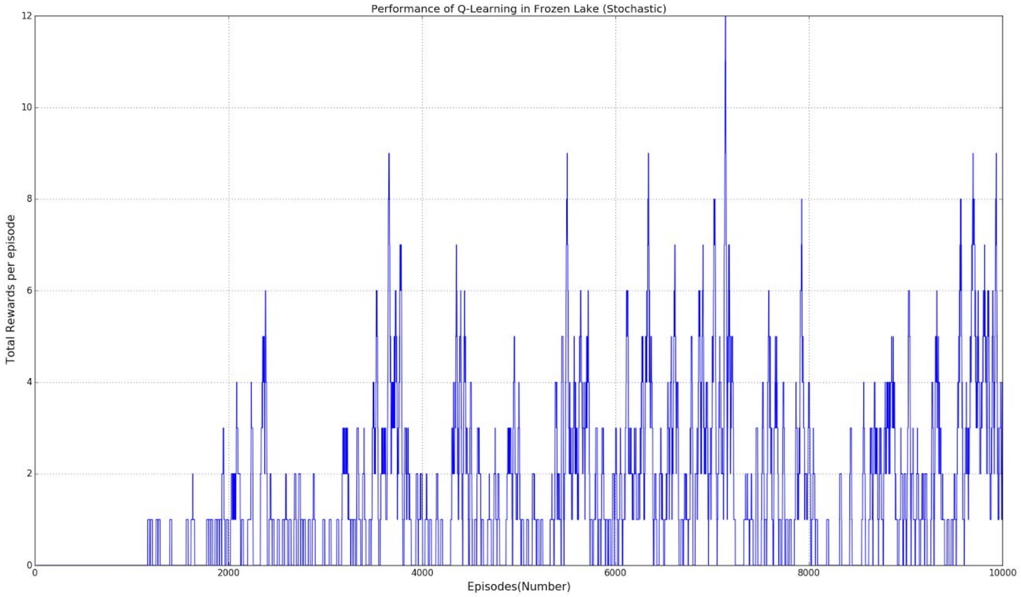




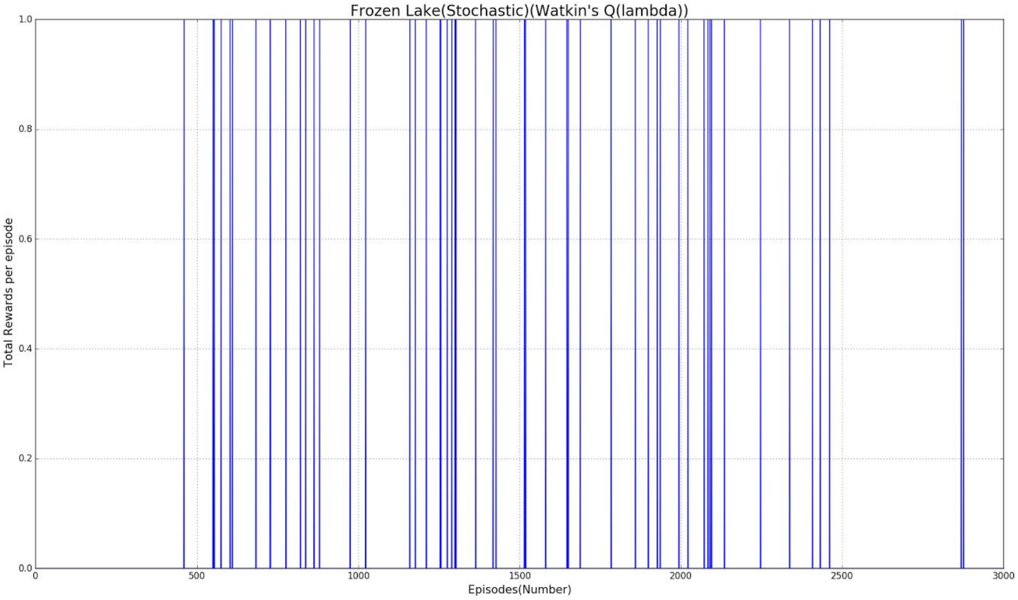
The graph below shows the performance of the Q Learning in the MDP Grid World and it gives a stable positive reward after 11 episodes.



The graph below shows the performance of the Q Learning in the Stochastic Environment Since the environment is stochastic so we get a list ups and downs in the graph but they are compensated by using a sliding window of size 15, and it shows positive rewards after 500 episodes



The graph below shows the performance of the Frozen Lake, which is a Stochastic Environment , using the Eligibility Trace algorithm. The algorithm starts showing positive rewards after almost 400 episodes.


Frozen Lake(Stochastic)(Watkin's Q(lambda))
Total Rewards per episode
Episodes(Number)
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000