

Reinforcement Learning

Homework 3 Report

Abhishek Agarwal

2016126

Ans 1)

*Attached below

Ans2)

*Attached below

Ans3)

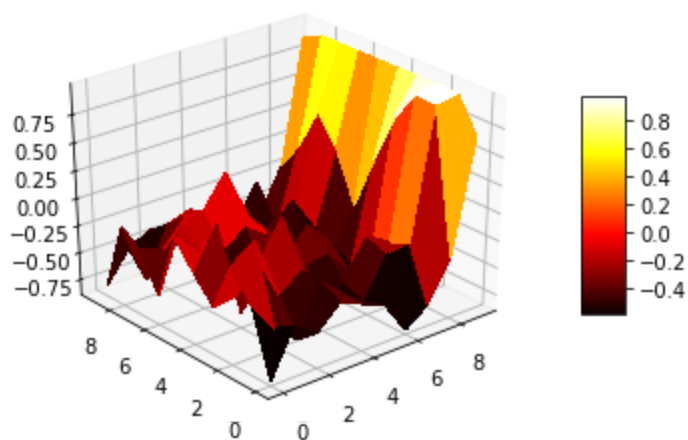
*Attached below

Ans4)

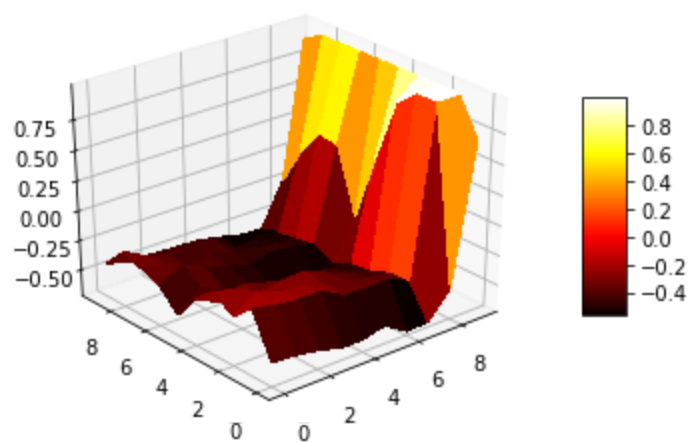
The plots generated are as follows:

All the graphs are shown in Jupyter Notebook.

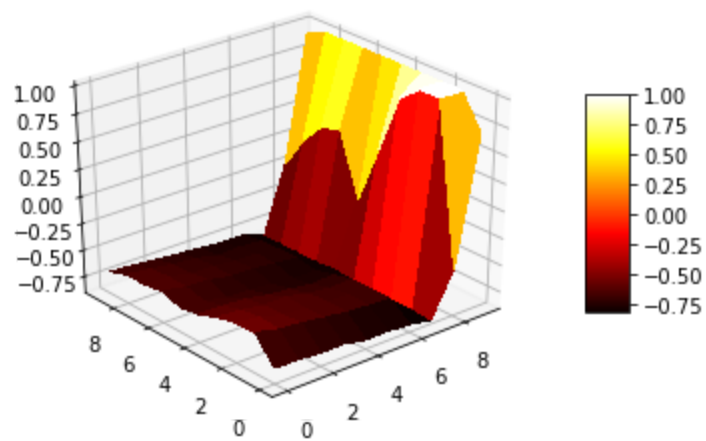
(a) 10k simulations with usable ace:



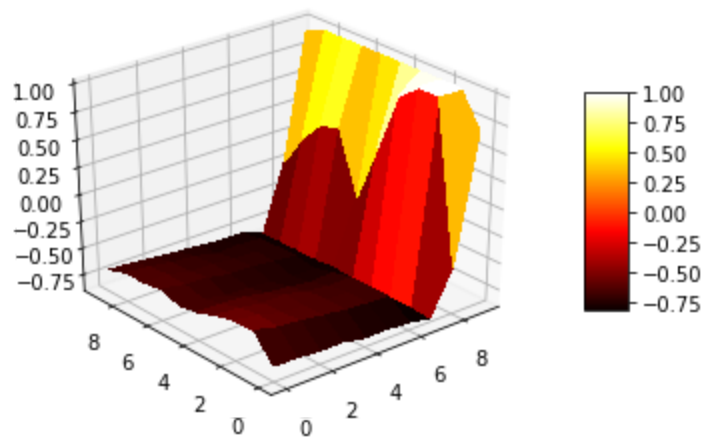
(b) 500k simulations with usable ace:



(c) 10k simulations with non usable ace:



(d) 500k simulations with non usable ace:



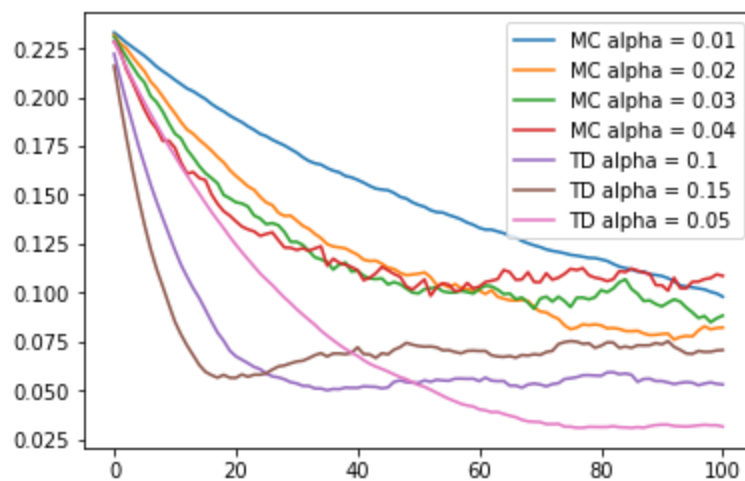
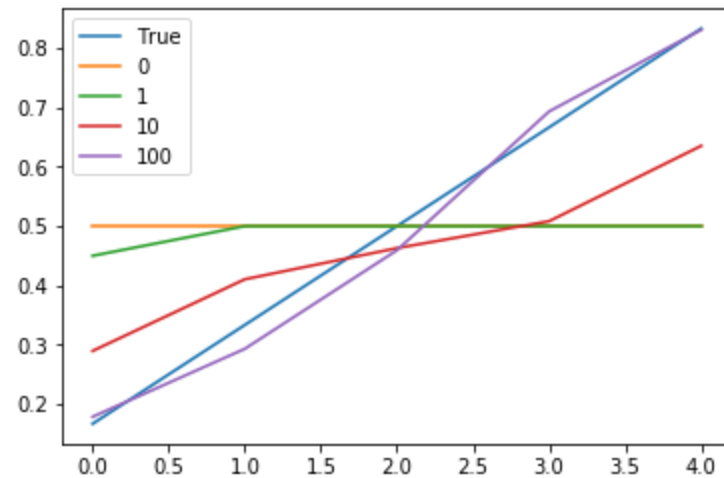
Ans5)

*Attached below

Ans6)

The generated plots are shown below:

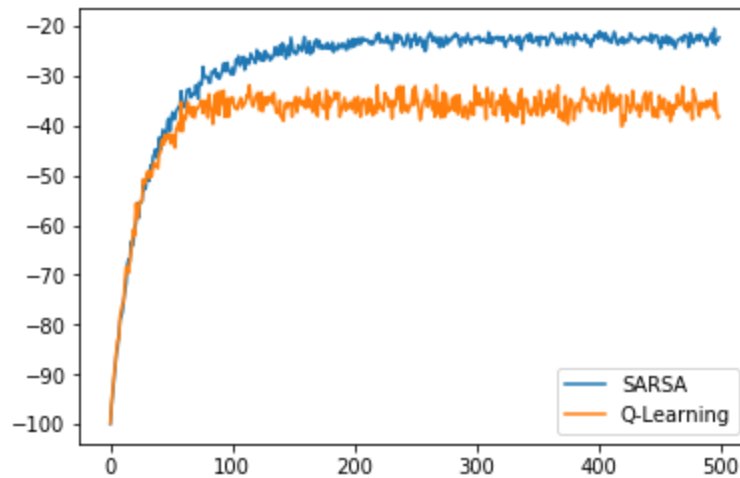
The values on 100 episodes are very close to the true values.



Ans7)

In the cliff example, we can see the difference in Q Learning and SARSA as shown below. The Q learning method chooses the greedy path which is difficult as it is risky,

whereas the SARSA method explores the different paths. The path selected by Q learning is near the cliff and hence can receive high negative reward on taking one wrong action.



Ans8)

*Attached below

Ans 6.3)

Since the rewards are 0 when we go to any non terminal state and we have initialized all the values of non terminal states as 0.5.

When we end up in the leftmost terminal state, only the value of state A is updated. The new value will be 0.45 i.e, there is a change of 0.05.