RL Assignment 3

Suraj Pandey
MT18025

Question 4 : -

Figures in Blackjack game

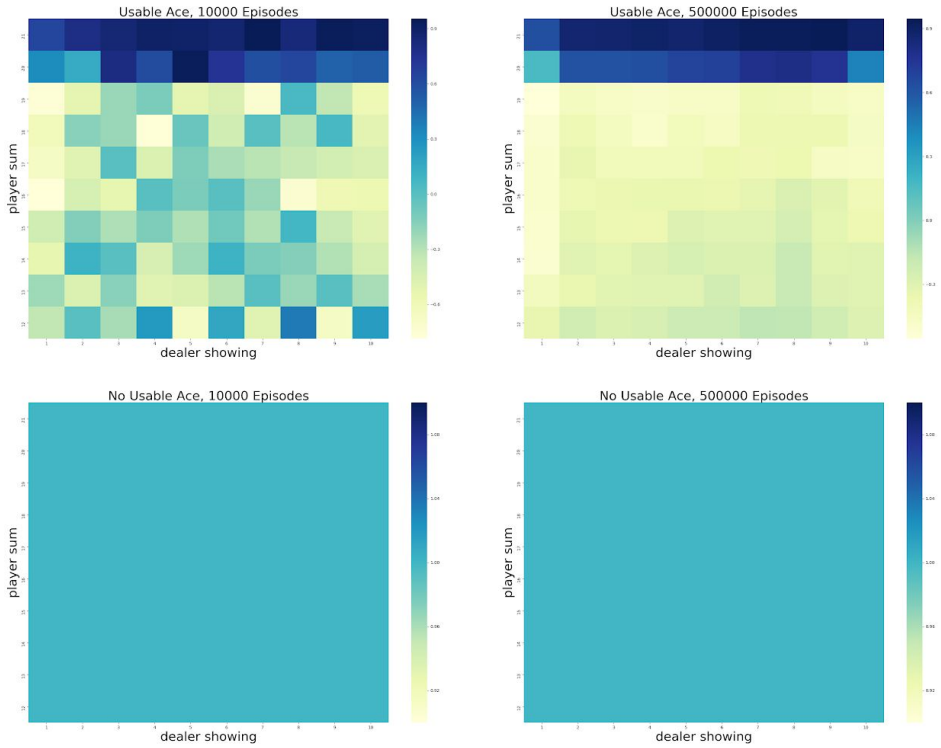


Figure 5.1

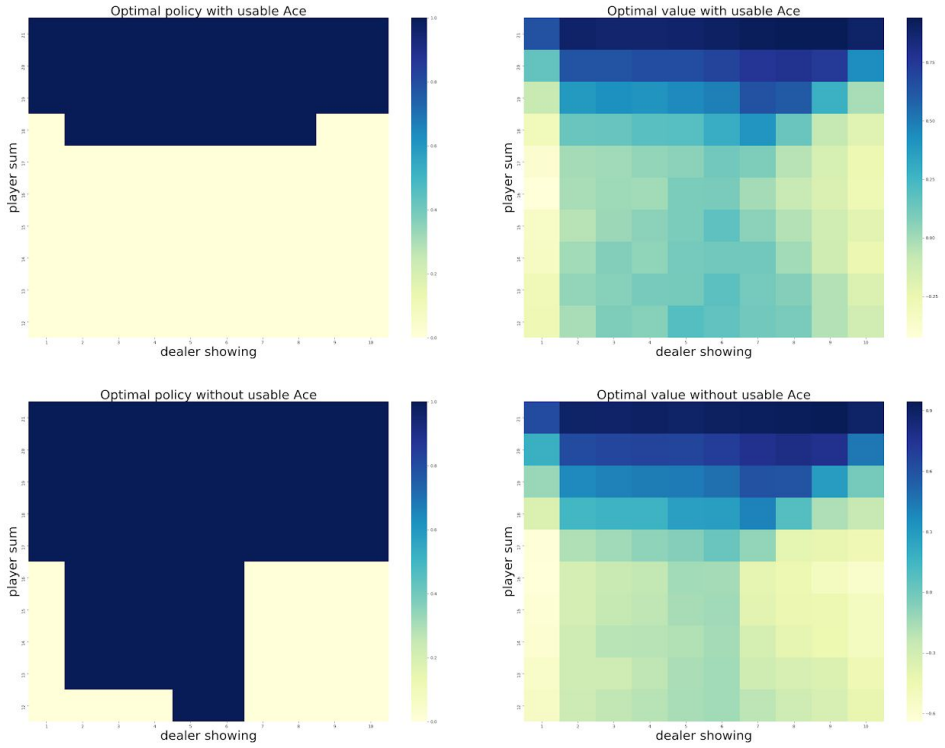


Figure 5.2

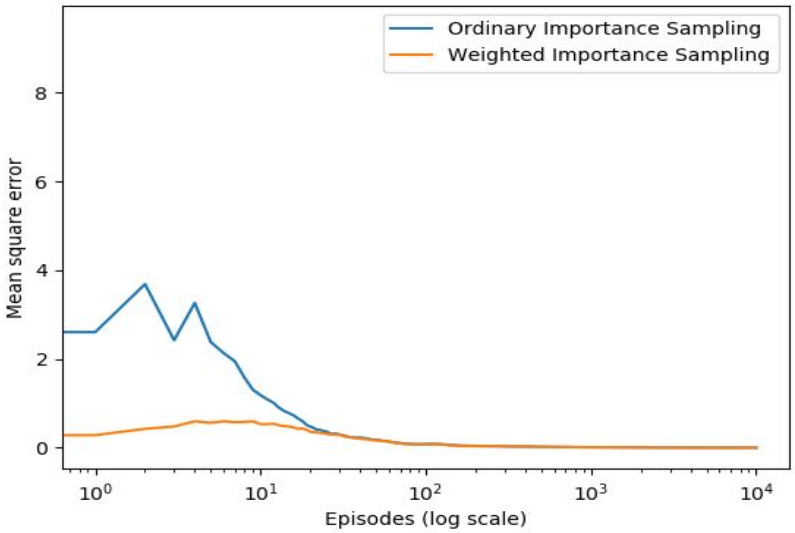
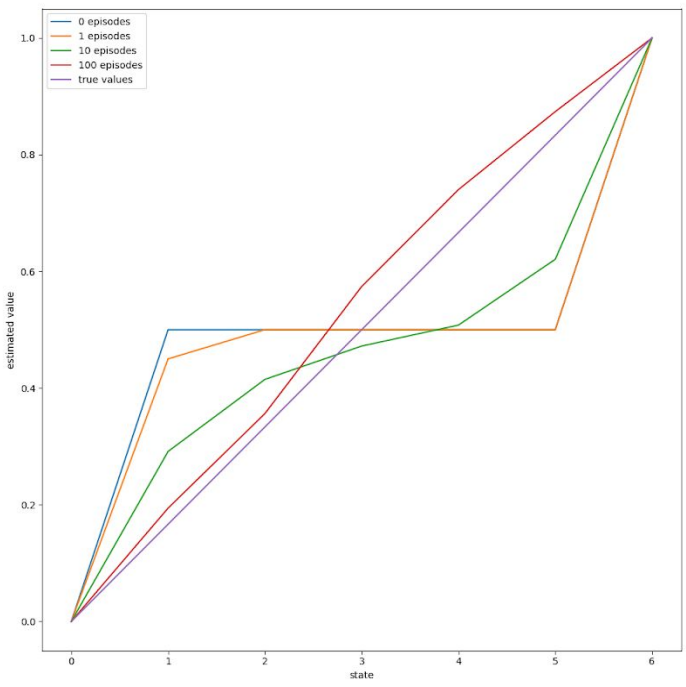
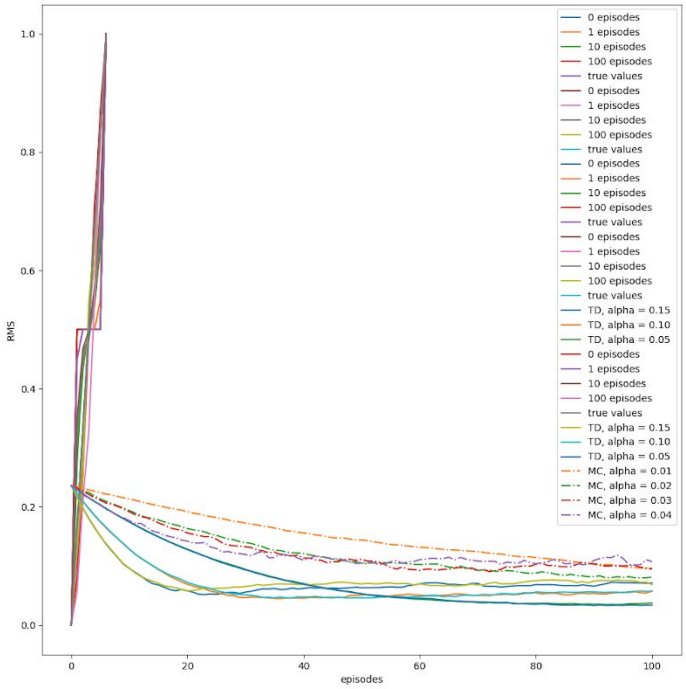


Figure 5.3

Question 6: -

Figure in Example 6.2





Question 7 :-

Environment used : Maze 5*5 open AI gym

Q- learning

Q(st,at)=Q(st,at)+learning_factor*(reward_t+discount_factor*max_on_a(Q(st+1,at+1))-Q(st,at))

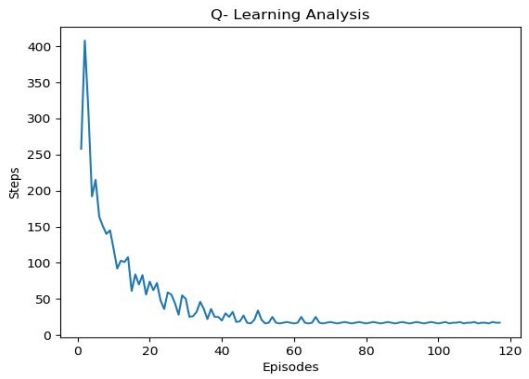
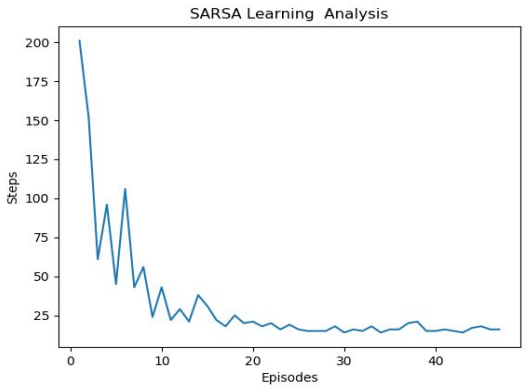
SARSA

Q(st,at)=Q(st,at)+learning_factor*(reward_t+discount_factor*Q(st+1,at+1)-Q(st,at))

Observation Table for Q-learning learning:

S. No.	Learning Rate	Discount Factor	Episodes Taken	Time taken
2	0.2	0.2	253	169
3	0.2	0.8	55	62
4	0.4	0.8	32	38

6	0.6	0.8	25	25
7	0.8	0.6	23	23

Comparison between Q-learning and SARSA	
	

Conclusion / Inferences:

- Q-learning directly learns the optimal policy, whereas SARSA learns a near-optimal policy while exploring
- Q-learning takes more time to converge than SARSA.
- From Observation, when discount factor is large then, the Q-learning algorithm will work more faster i.e, converges faster than SARSA.
- Number of Episodes taken to converge in SARSA will be small in comparison to Q-learning.And when the discount factor and exploration rate increases then, SARSA takes more number of episodes than Q-learning.