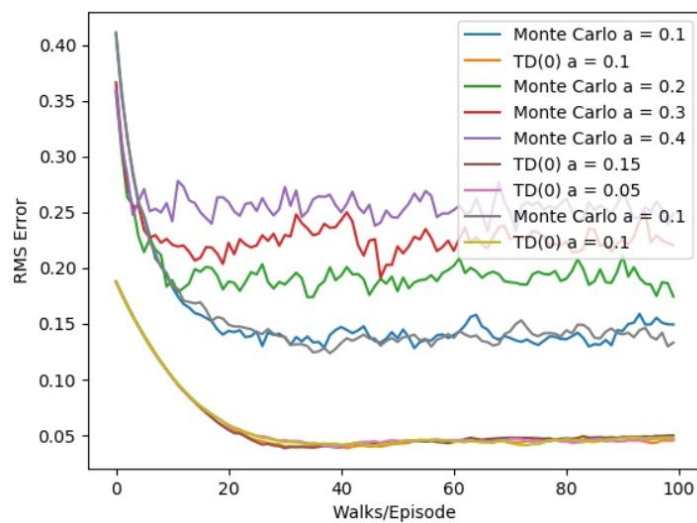
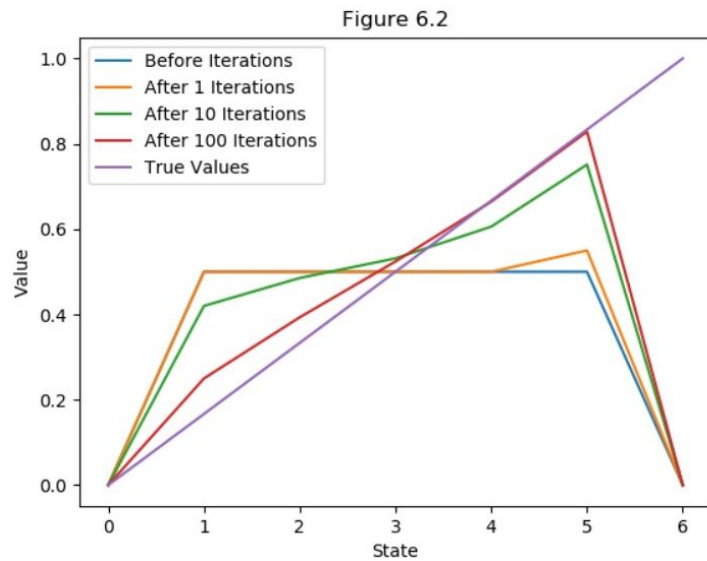


# Assignment - 3

## Abhishek Maiti

### Question 6

Estimation of the given policy

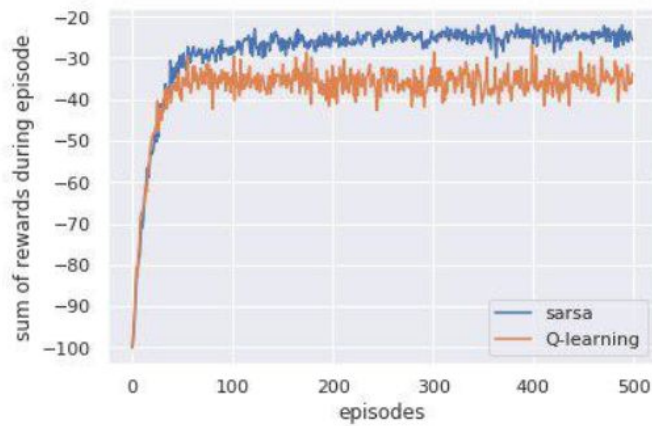


Ignore 0 and the 6th state. These are the terminal states.

TD(0) is performing better than Monte-Carlo for this particular task.

The top graph is Value Iteration for the different states, and the bottom is the error in estimation

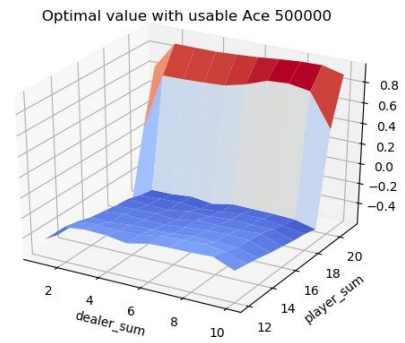
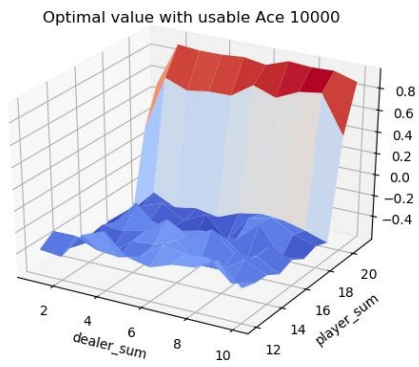
### Question 7



### Q-Learning Vs Sarsa

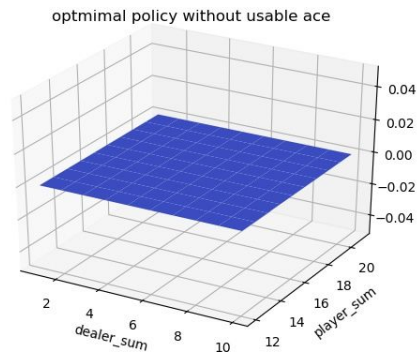
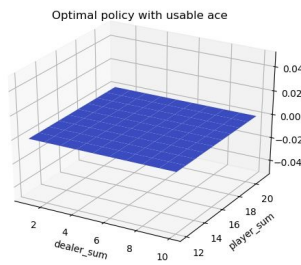
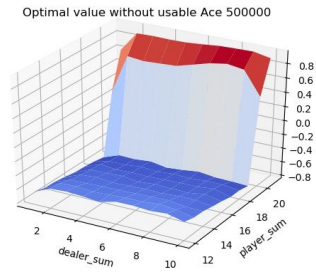
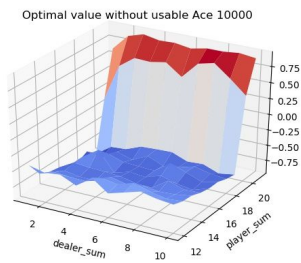
We can see that Sarsa is performing better than Q-learning.

### Question 4

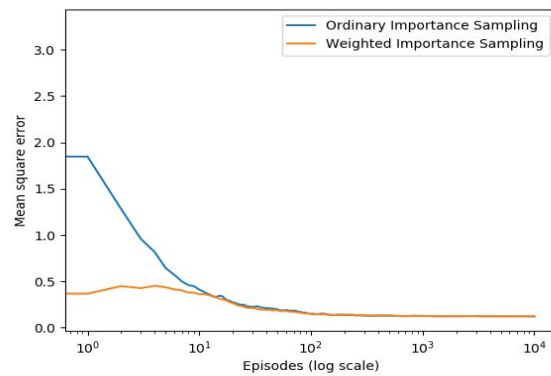
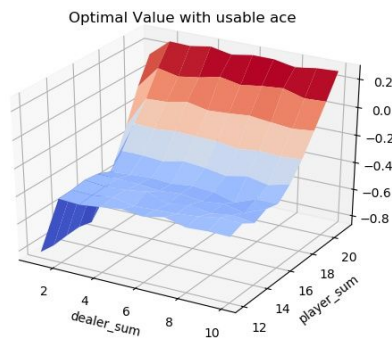


Optimal Values for 10000 and 500000 trials (with usable ace)

For without usable ace, turn to next page.



**Best Policy for given rewards. (Above Figures) and Value function below right**



**Weighted importance sampling produces lower error estimates of the value of a single blackjack state from off - policy episodes. (above right)**