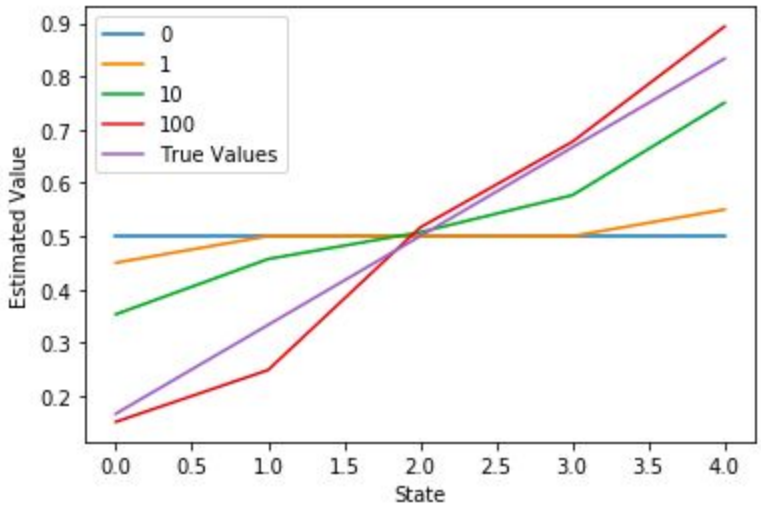
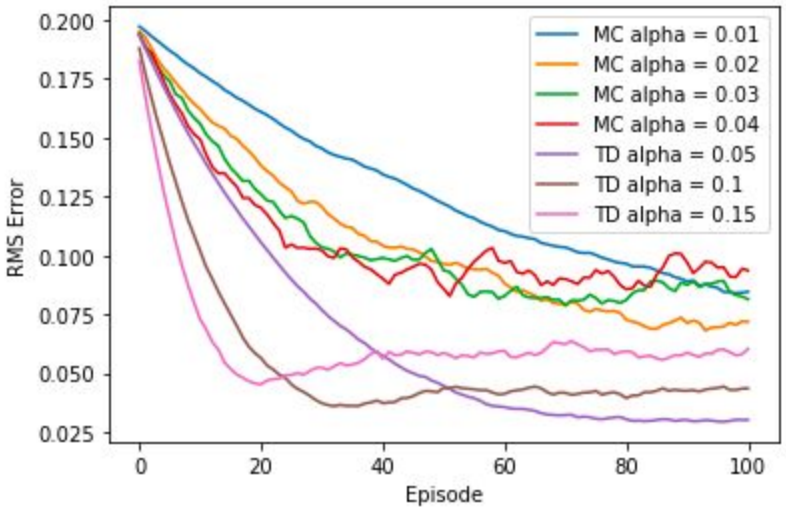
Question 6:

Estimated Value vs State



We can see from the graph that estimates after 100 episodes are almost equal to the true values and we can see that as the number of episodes increases, the estimates get better and better.

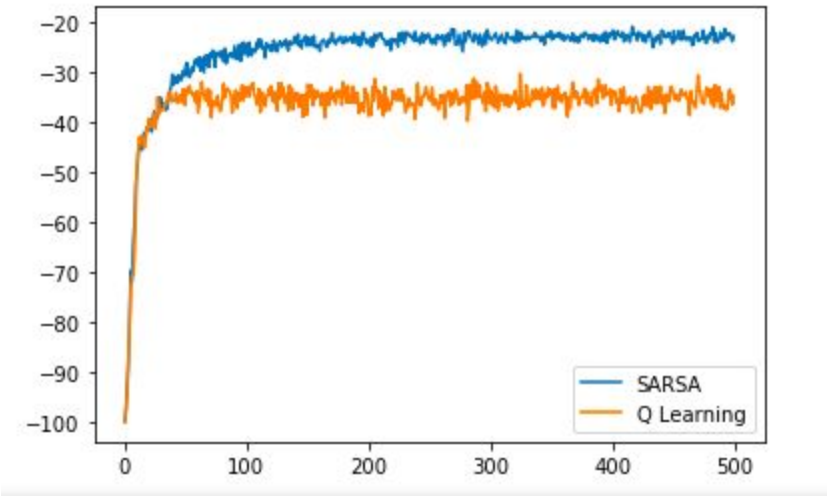
RMSE vs Episode



Just by comparing MC and TD as a whole, we can see that TD performs much better than MC consistently. A small value of alpha gives a smooth

curve for RMSE in case of both TD and MC because it is less sensitive to the rewards at a given episode. Larger values of alpha give a lower RMSE initially but later on increase.

Question 7:



Sarsa learns the safer path because it takes into consideration the action selection. Because of the safer path, its sum of rewards is higher than Q learning. However, the Q learning falls off the cliff much more than Sarsa hence the lower sum of rewards