MACHINE LEARNING NANODEGREE
PROJECT - 4
TRAIN A SMART CAB

Question 1 - Observe what you see with the agent's behavior as it takes random actions. Does the ***smartcab*** *eventually make it to the destination? Are there any other interesting observations to note?*

Answer – When the agent used random actions, it never reached destination. I was just seeing the inputs(giving the details about the traffic), deadline(decreasing and goes to -100), action(the random action that it was taking),reward(the reward it got after executing that action). Not any other interesting observation it was just executing the random actions.

Question 2-What states have you identified that are appropriate for modeling the ***smartcab*** *and environment? Why do you believe each of these states to be appropriate for this problem?*

Answer – I have choosen 5 states. All the inputs and the next waypoint. All the inputs are the lights at traffic signal , the traffic from all three directions and where to go next. All the mentioned states are needed for the smartcab to make good decisions like what's the traffic, what's the signal and where to go next. I haven't choosen deadline because with the help of next waypoint it's already taking perfect paths, so including deadline won't make any difference.

Question 3 - What changes do you notice in the agent's behavior when compared to the basic driving agent when random actions were always taken? Why is this behavior occurring?

Answer – Now agent's behavior chnaged now it's taking actions with respect to Q-Learning. By the use of Q-Learning and training it has learned which actions to take when you are in that particular state. Now it's taking actions with respect to policy not taking random actions as it was taking ppreviously and it was able to reach destination few times when I ran the code.

Question – 4 Report the different values for the parameters tuned in your basic implementation of Q-Learning. For which set of parameters does the agent perform best? How well does the final driving agent perform?

Answer – I will be using this notation in this answer. E is epsilon, L is learning rate, G is gamma. I have used E,L,G ={(0.3,0.3,0.4), (0.5,0.3,0.8) , (0.8,0.3,0.5) and (0.1,0.3,0.8)}. The best combination that works best is E,L,G = {0.3,0.3,0.8},but it doesn't give consistent results,sometimes we don't have expected success rate in the last 10 trials,so the combination we are using is, {0.1,0.3.0.8}, to get consistent results.Low value of epsilon means that we are more interested in exploiting rather than exploring. Using these paramters after tuning the Q-Learning algorithm, the agent got to the destination very quickly using very minimal number of steps and maximizing reward. The results of the hyperparameters are as follows:

Epsilon	Learning Rate	Gamma	Total reward
0.3	0.3	0.8	26
0.1	0.3	0.8	21.5
0.5	0.3	0.8	23.5
0.8	0.3	0.5	7.5

Question – 5 Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties? How would you describe an optimal policy for this problem?

Answer – Yes, the agent almost achieved optimal policy. Although it did incurr some penalties as total reward came down a bit but then starts increasing and maximizing, after few more trials it won't be incurring any penalties,and in later trials it was completing the task in half the number of available steps, hence getting to optimal policy. I would define optimal polcy as, an action that any agent must take in any state to maximize reward i.e. it must always choose an optimal action in any given state i.e. circumstances.

In last 5 trials, the mistakes agent made are:

Input	Action	Reward
{'light': 'green', 'oncoming': None, 'right': None, 'left': 'left'}	left	-0.5
{'light': 'red', 'oncoming': None, 'right': None, 'left': None}	left	-1.0
{'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}	left	-1.0
{'light': 'green', 'oncoming': None, 'right': None, 'left': None}	right	-0.5

Other than these mistakes, the agent performed very well and did all the right actions,hence we can say that the agent is getting to the optimal policy.