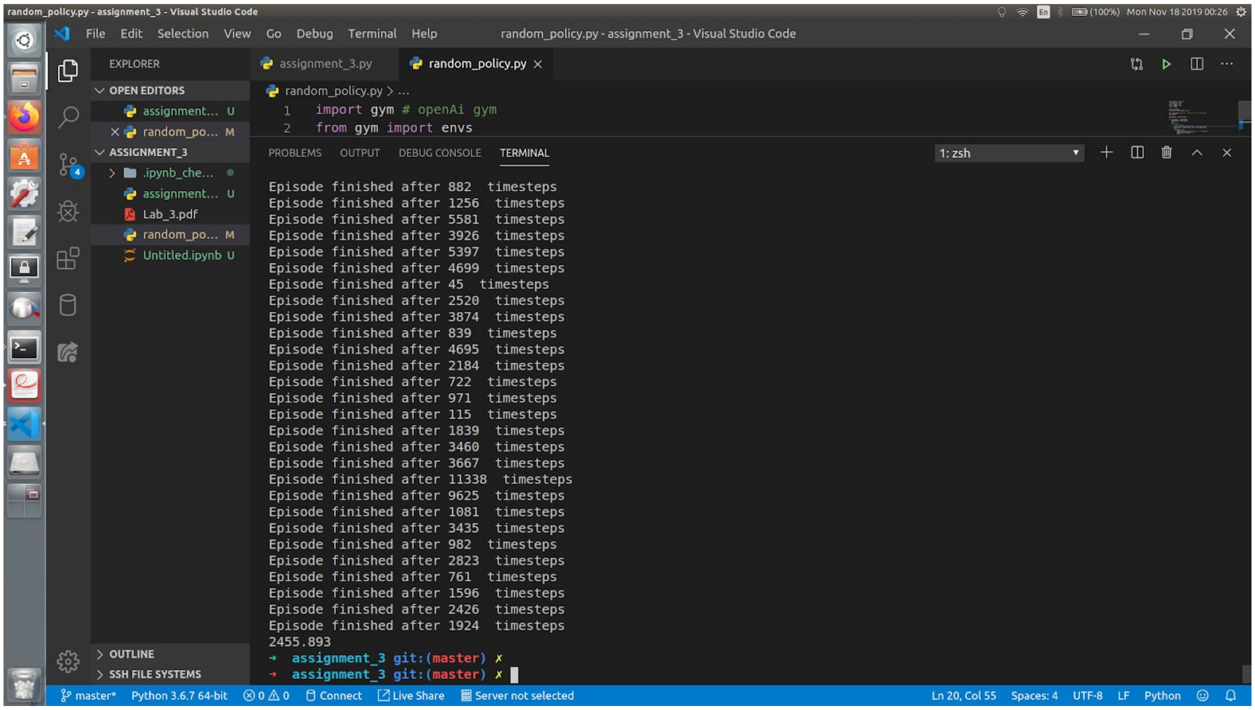
CS529– Applied Artificial Intelligence
Lab Assignment - 3

Name : Shah Jainam Mukeshbhai
Entry Number : 2017csb1107

Question 1 :

Chose an action randomly and when the reward was **20**, then reseted the environment. The average number of timesteps for 1000 episodes were **2455.893**(Around 2500).



```
import gym # openAi gym
from gym import envs
```

```
Episode finished after 882  timesteps
Episode finished after 1256  timesteps
Episode finished after 5581  timesteps
Episode finished after 3926  timesteps
Episode finished after 5397  timesteps
Episode finished after 4699  timesteps
Episode finished after 45  timesteps
Episode finished after 2520  timesteps
Episode finished after 3874  timesteps
Episode finished after 839  timesteps
Episode finished after 4695  timesteps
Episode finished after 2184  timesteps
Episode finished after 722  timesteps
Episode finished after 971  timesteps
Episode finished after 115  timesteps
Episode finished after 1839  timesteps
Episode finished after 3460  timesteps
Episode finished after 3667  timesteps
Episode finished after 11338  timesteps
Episode finished after 9625  timesteps
Episode finished after 1081  timesteps
Episode finished after 3435  timesteps
Episode finished after 982  timesteps
Episode finished after 2823  timesteps
Episode finished after 761  timesteps
Episode finished after 1596  timesteps
Episode finished after 2426  timesteps
Episode finished after 1924  timesteps
2455.893
➜  assignment_3 git:(master) ✗
➜  assignment_3 git:(master) ✗
```

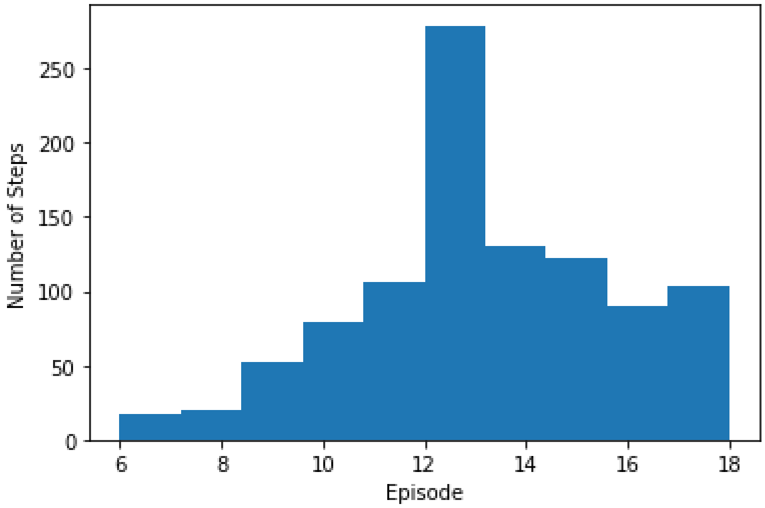
Question 2 :

Discount Factor	Policy Iteration	Value Iteration
0.99	12 * 1100	725
0.95	12 * 226	146
0.9	12 * 110	74

0.8	12 * 53	38
0.4	12 * 14	14

- For discount factor 0.4, I didn’t get the same policy
- Otherwise, Optimal policy is same for Value Iteration and Policy Iteration
- Compared to each other, value-iteration is computationally efficient even though it takes more number of iterations to converge, each iteration is less computationally expensive than policy-iteration.

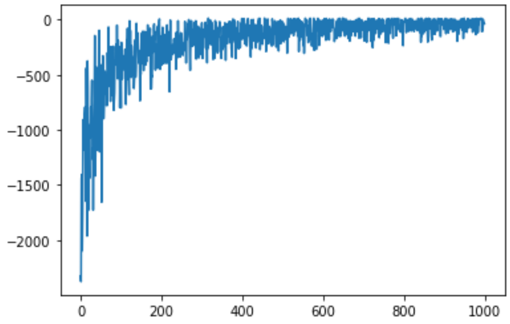
Distribution of Number of steps taken for 1000 episodes is as follows:



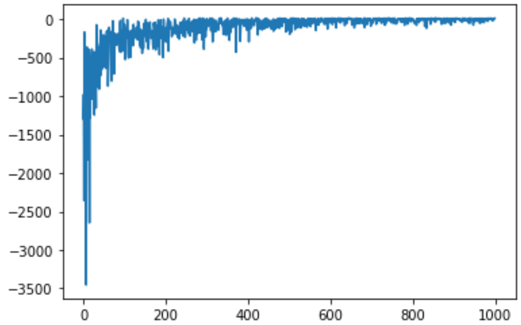
Average number of steps is 13.

Question 3 :

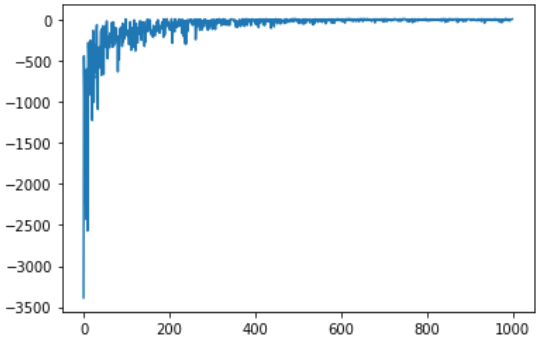
Alpha = 0.05, Convergence around greater than 1000 eps



Alpha = 0.1, Convergence around 600 eps



Alpha = 0.2, Convergence around 500 eps

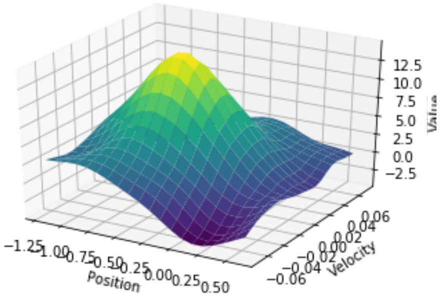


- On Y-axis, I kept total reward of the episode.
- On convergence, the change is total reward becomes constant. It becomes parallel to x-axis.

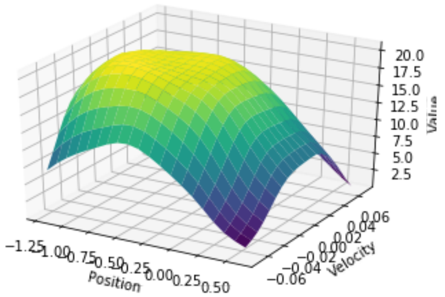
Question 4 :

The results were matching the graphs of book. Below are the plots of Mountain Car problem.

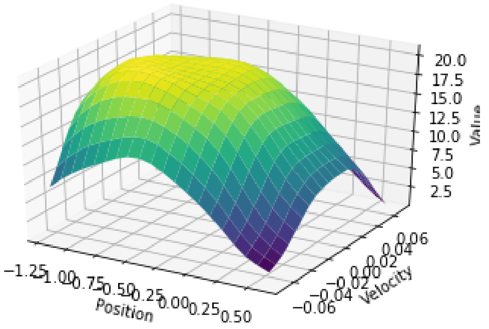
Episode 0



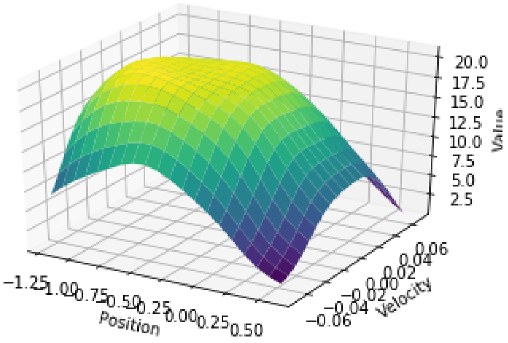
Episode 100



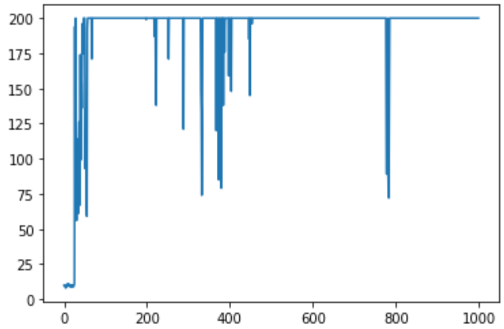
Episode 200



Episode 300



For Cart Pole, the plot of Total reward with num of episodes is :



References:
For Mountain Car:
https://github.com/SamKirkiles/mountain-car-SARSA-AC/blob/master/mountain_car.py
For Cart Pole: https://github.com/ceteke/RL/blob/master/Approximation/Linear%20Sarsa.ipynb