# Reinforcement Learning
# EX4 : Monte Carlo

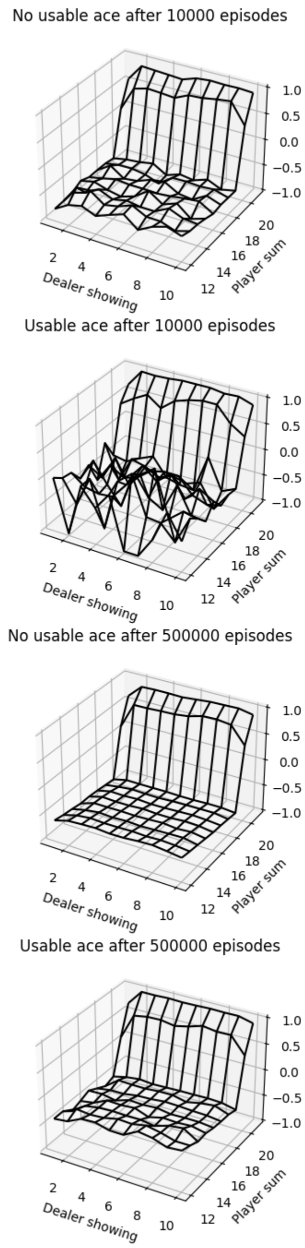
**NOTE –**

**I had to submit late by a few hours as I was getting so many errors due to gym 0.26 version. Once I downgraded it to gym 0.20 everything magically started working. Please consider this while grading my assignment. You have no idea how frustrating it is to solve unnecessary errors because you are using the latest version. I suggest you add this point in your instructions so that the next batch won't suffer.**

Q2 C)
I have attached the code but I was not able to produce the results.
Q Suppose that instead an every-visit MC method was used on the
same problem. Would the variance of the estimator still be infinite? Why or why not?
Answers →
I think, No, the variance of the estimator would no longer be infinite if an every-visit MC method was used on the same problem.
Reason → every-visit MC estimator averages over all returns following visits to a state, regardless of state is visited repeatedly. In short, every-visit MC estimator is not susceptible to the same kind of bias as the first-visit MC estimator, which only considers the return following the first visit to a state.

Q3)

a)



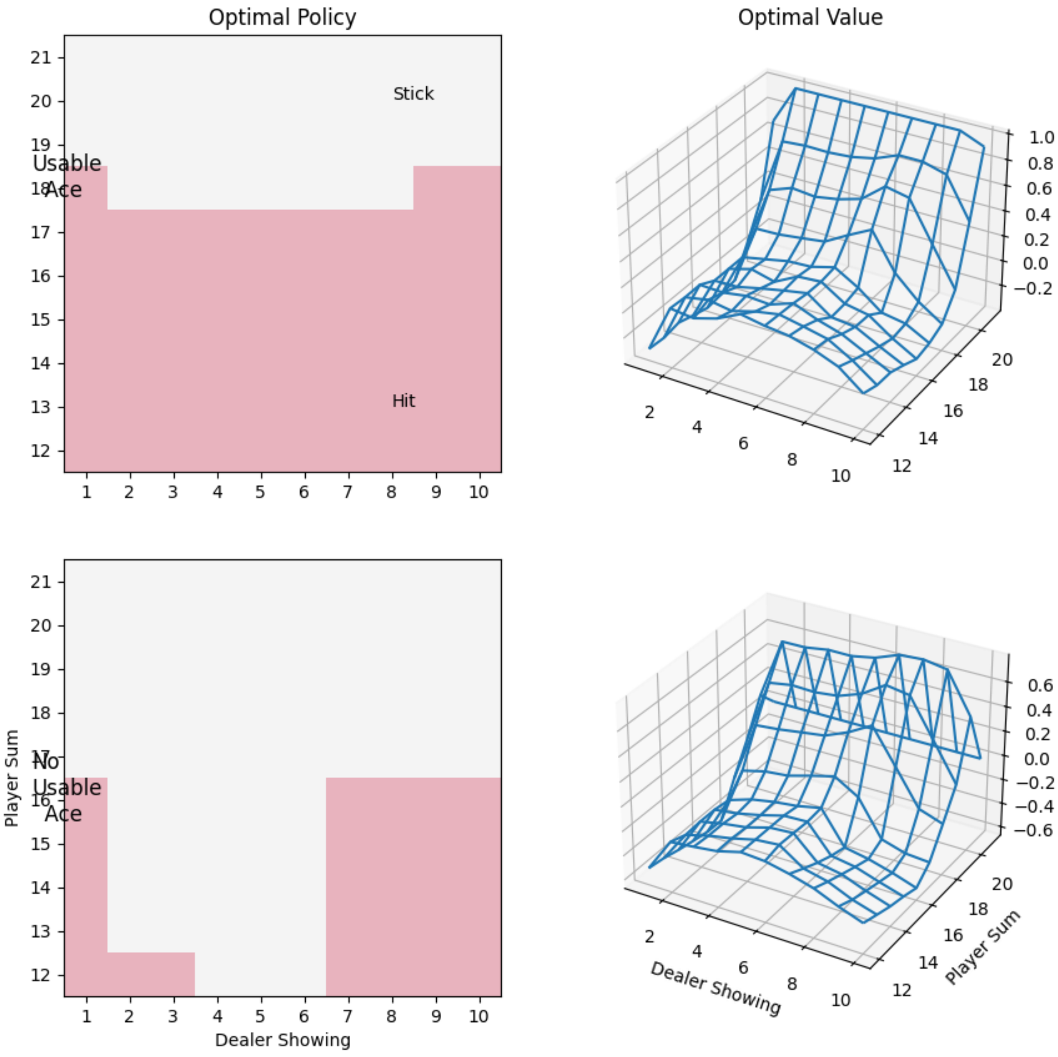
No usable ace after 10000 episodes

Usable ace after 10000 episodes

No usable ace after 500000 episodes

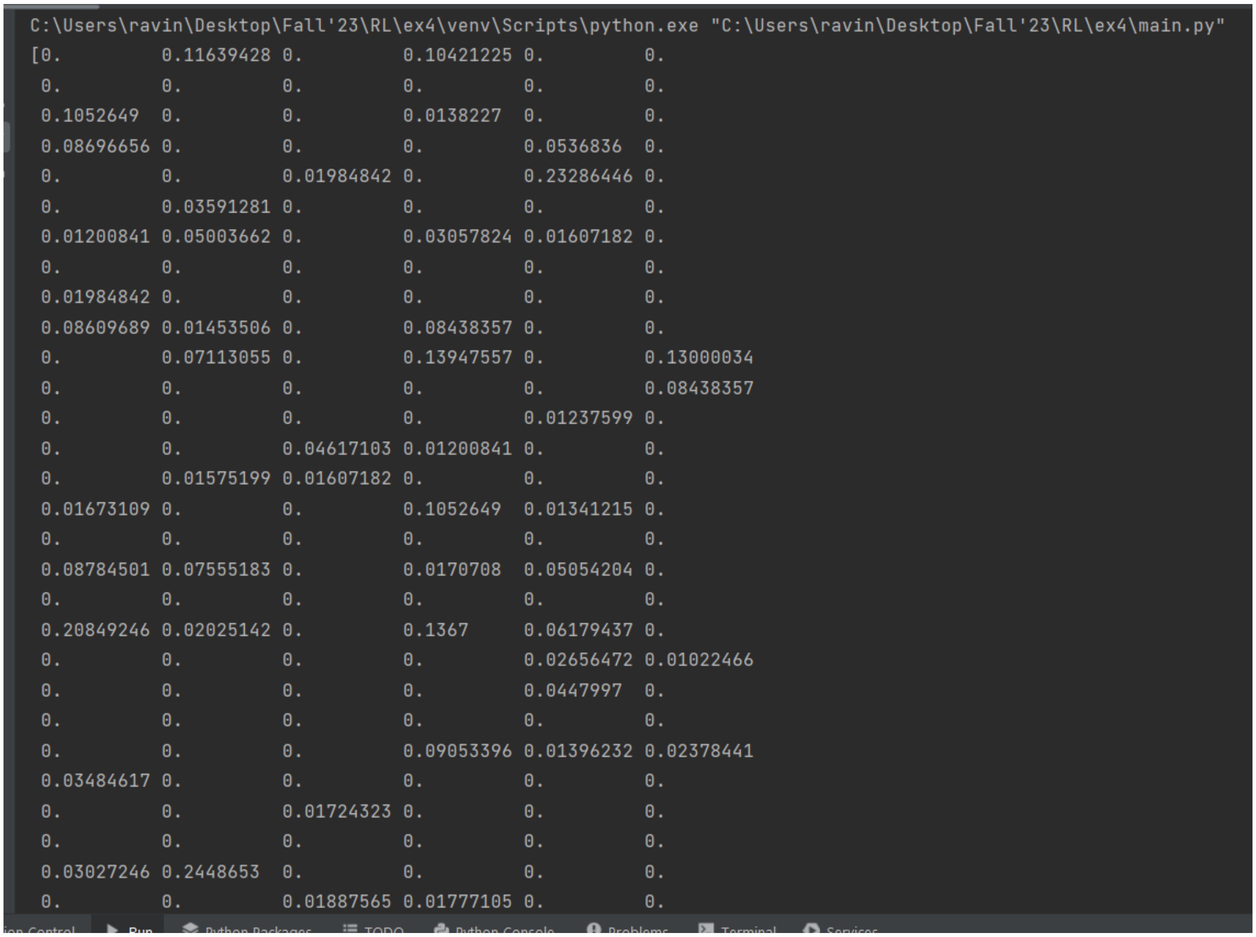
Usable ace after 500000 episodes

b) first visit Monte-Carlo with exploring starts



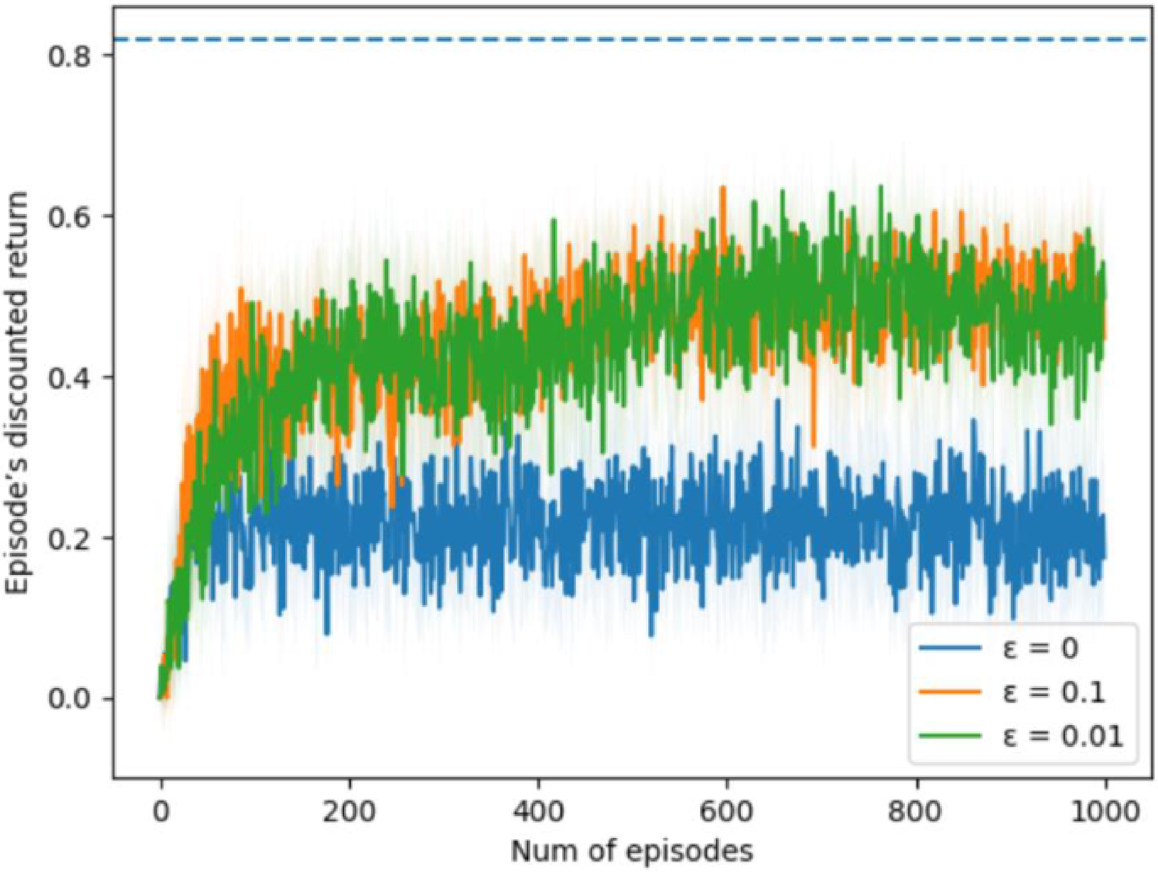
Q4 Epsilon Soft policy

a)



```
C:\Users\ravin\Desktop\Fall'23\RL\ex4\venv\Scripts\python.exe "C:\Users\ravin\Desktop\Fall'23\RL\ex4\main.py"
[0.         0.11639428 0.         0.10421225 0.         0.
 0.         0.         0.         0.         0.         0.
 0.1052649  0.         0.         0.0138227  0.         0.
 0.08696656 0.         0.         0.         0.0536836  0.
 0.         0.         0.01984842 0.         0.23286446 0.
 0.         0.03591281 0.         0.         0.         0.
 0.01200841 0.05003662 0.         0.03057824 0.01607182 0.
 0.         0.         0.         0.         0.         0.
 0.01984842 0.         0.         0.         0.         0.
 0.08609689 0.01453506 0.         0.08438357 0.         0.
 0.         0.07113055 0.         0.13947557 0.         0.13000034
 0.         0.         0.         0.         0.         0.08438357
 0.         0.         0.         0.         0.01237599 0.
 0.         0.         0.04617103 0.01200841 0.         0.
 0.         0.01575199 0.01607182 0.         0.         0.
 0.01673109 0.         0.         0.1052649  0.01341215 0.
 0.         0.         0.         0.         0.         0.
 0.08784501 0.07555183 0.         0.0170708  0.05054204 0.
 0.         0.         0.         0.         0.         0.
 0.20849246 0.02025142 0.         0.1367     0.06179437 0.
 0.         0.         0.         0.         0.02656472 0.01022466
 0.         0.         0.         0.         0.0447997  0.
 0.         0.         0.         0.         0.         0.
 0.         0.         0.         0.09053396 0.01396232 0.02378441
 0.03484617 0.         0.         0.         0.         0.
 0.         0.         0.01724323 0.         0.         0.
 0.         0.         0.         0.         0.         0.
 0.03027246 0.2448653  0.         0.         0.         0.
 0.         0.         0.01887565 0.01777105 0.         0.
```

b)



c) Explain how the result of the epsilon = 0 setting demonstrate the importance of doing ES in MC —>

1. ε = 0 means the algorithm always selects the best action  and refuses to explore or make new attempts.
2. Without exploration, the algorithm can get stuck in a local optimum
3. With ES, this is avoided, as it will try out other states too. And there is a chance that the policy may get improved.