Exercise 4

Kai Schneider

May 18, 2021

Task 1

a.) Advantages MC vs DP

- MC only considers values from one episode for it's updates, not from all previous backups
  Although this might increase variance, it drastically improves computation time
- For DP we need the transition matrix/distribution, which might be unrealistic for real scenarios.
  For MC we try to obtain our information by interacting with the environment.

b.) Example environment

MC methods are only suitable for repetitve scenarios, because we only update after an episode has ended. The most common examples are games like blackjack or backgammon where it is close to impossible to determine the value of the current state. Therefore DP isn't a viable option, but MC is.
A more exotic example environment might be the simulation of illumnation/reflection behaviours. Approximating the path of the light with MC is an efficient way to solve this problem because we don't have to solve the infinity number of possible situations (angle, reflectivity, intensity, etc) or create an complex simulation model of the system.

Task 2

a.) first visit MC

I: 10k episodes

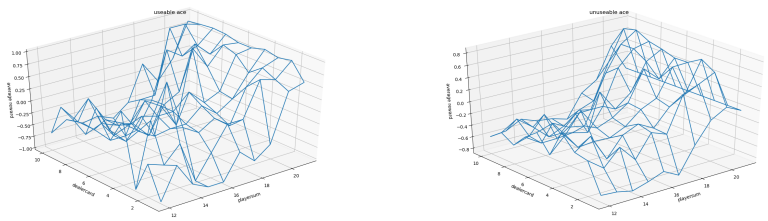

Figure 1: plots first-visit MC with 10k episodes (usable ace left, unusable right)

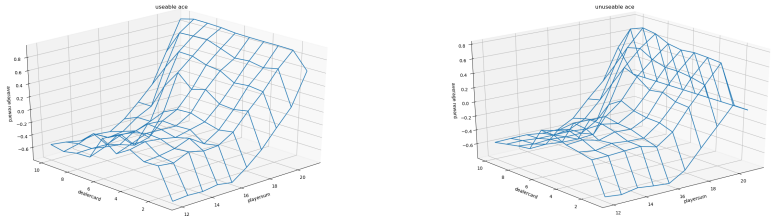

Figure 2: plots first-visit MC with 500k episodes (usable ace left, unusable right)

II: 500k episodes

b.) ES MC

For all plots:

- **Yellow:** 1.0 - hit
- **Purple:** 0.0 - stick

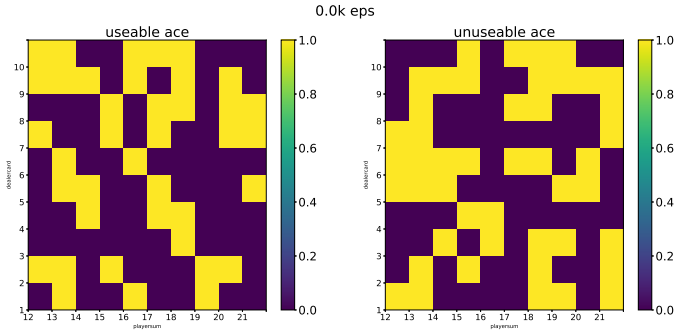

Figure 3: plots ES MC with 0k episodes (usable ace left, unusable right)

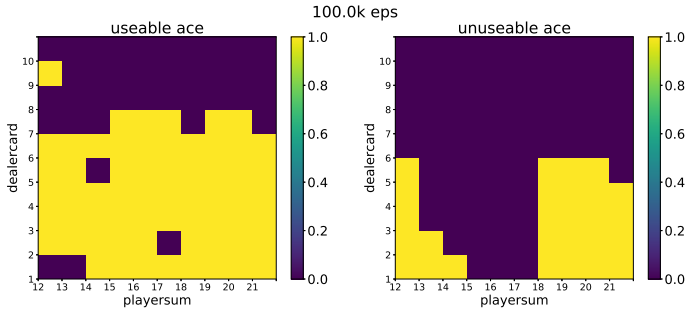

Figure 4: plots ES MC with 100k episodes (usable ace left, unusable right)

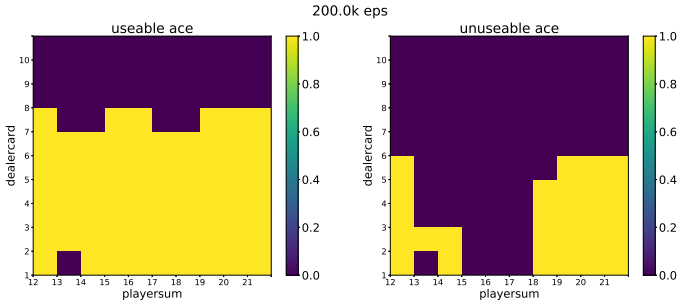

Figure 5: plots ES MC with 200k episodes (usable ace left, unusable right)

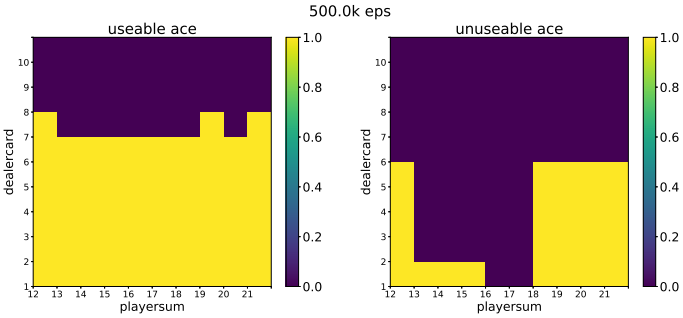

Figure 6: plots ES MC with 300k episodes (usable ace left, unusable right)

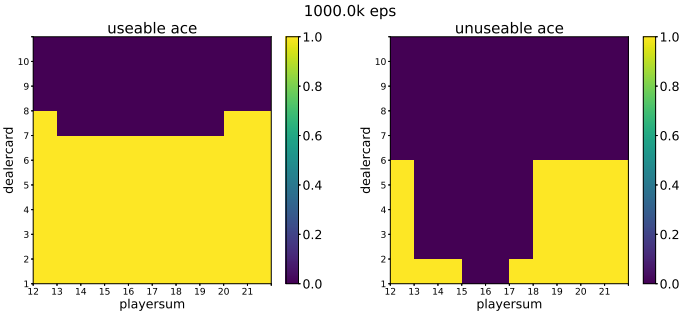

Figure 7: plots ES MC with 400k episodes (usable ace left, unusable right)