CS533-Assignment3-Report

Chunxiao Wang

1 Part I: Build a Planner

I choose the policy iteration to get the exact optimal policy and value function for MDP planning.

2 Part II: Run the Planner

The results below provide the value function and optimal policy derived from policy iteration for MDP1 and MDP2 with different β values. The value function and policy are n-dimensional vectors and the indices are states.

2.1 MDP1, $\beta = 0.1$

In this case, the optimal value function with four decimal is

```
0.1001
0.0090
0.0088
0.0090
1.0010
0.0068
0.0683
0.0086
0.0100
0.0896,
```

and the optimal policy is

```
3
3
2
0
0
0
1
2
1
3
```

2.2 MDP1, $\beta = 0.9$

In this case, the optimal value function with four decimal is

```
3.3210
2.9234
2.8914
2.9234
3.6900
2.8407
3.1564
2.9071
2.9889
3.2482,
```

and the optimal policy is

```
3
3
2
0
0
0
1
2
1
3
```

2.3 MDP2, $\beta = 0.1$

In this case, the optimal value function with four decimal is

```
0.0114
0.0100
0.5733
0.0015
0.0604
0.1010
1.0101
0.0080
0.0060
0.1010,
```

and the optimal policy is

```
3
0
0
1
```

```
2
2
2
3
1
2
```

2.4 MDP2, $\beta = 0.9$

In this case, the optimal value function with four decimal is

```
4.2632
4.2577
5.1885
3.8440
4.4169
4.7368
5.2632
3.8351
3.9752
4.7368,
```

and the optimal policy is

```
0
0
0
1
2
2
2
1
1
2
```

3 Part III: Parking Domain

3.1 Design

We specify our MDP using suggested structure. To make the location L more explicit, we further split it into column $C \in \{A, B\}$ and row $R \in \{1, 2, \ldots, n\}$ with n being the number of parking rows. Therefore each state is a 4-tuple (C, R, O, P). Here we set n as 10 and the discount factor $\beta = 0.9$, which is close to 1 to let the behavior be mainly influenced by the MDP dynamics and reward. We select two sets of parameters. The only difference is in the assumption of the negative reward that represents the cost of driving. For the first set, we use the -1 representing the cost of driving and for the second set, we use the -5

representing the cost of driving. The other parameters remain the same. The expectation is for the second setting, there should be more $PARK$ action and smaller value function compared with the first setting since the cost of driving gets larger from -1 to -5. The other qualitative characteristics are as follows,

- The reward for parking as a function of being in row $r \in \{2, 3, \ldots, n\}$ is $(n-(r-1)) \cdot 10$, where 10 is the reward coefficient which can be changed. The driver will get more reward if they park closer to the store except the handicap spots. The high cost for parking in the handicap spots is -1000 and for collision is -10000. The reward for the terminal state is 1. The terminal state has no transitions.

- After a DRIVE action, if the next spot is a handicap spot $(A[1], B[1])$, the probability that they are occupied is a small value fixed at 0.01. If the next spot is not a handicap spot, the probability that the next spot is occupied is a linear function of the row r: $(n-(r-1))/n$. As the row gets closer to the store, the occupied probability increases.

We implement the policy iteration to get the value function and optimal policy. For the first set with driving cost -1, the value function is

```
(A, 0, unoccupied, unparked) 63.16
(A, 0, unoccupied, parked) -999.10
(A, 0, occupied, unparked) 63.16
(A, 0, occupied, parked) -9999.10
(A, 1, unoccupied, unparked) 80.81
(A, 1, unoccupied, parked) 90.90
(A, 1, occupied, unparked) 55.84
(A, 1, occupied, parked) -9999.10
(A, 2, unoccupied, unparked) 71.81
(A, 2, unoccupied, parked) 80.90
(A, 2, occupied, unparked) 53.75
(A, 2, occupied, parked) -9999.10
(A, 3, unoccupied, unparked) 62.81
(A, 3, unoccupied, parked) 70.90
(A, 3, occupied, unparked) 52.25
(A, 3, occupied, parked) -9999.10
(A, 4, unoccupied, unparked) 53.81
(A, 4, unoccupied, parked) 60.90
(A, 4, occupied, unparked) 49.83
(A, 4, occupied, parked) -9999.10
(A, 5, unoccupied, unparked) 45.64
(A, 5, unoccupied, parked) 50.90
(A, 5, occupied, unparked) 45.64
(A, 5, occupied, parked) -9999.10
(A, 6, unoccupied, unparked) 40.07
(A, 6, unoccupied, parked) 40.90
```

```
(A, 6, occupied, unparked) 40.07
(A, 6, occupied, parked) -9999.10
(A, 7, unoccupied, unparked) 35.07
(A, 7, unoccupied, parked) 30.90
(A, 7, occupied, unparked) 35.07
(A, 7, occupied, parked) -9999.10
(A, 8, unoccupied, unparked) 30.56
(A, 8, unoccupied, parked) 20.90
(A, 8, occupied, unparked) 30.56
(A, 8, occupied, parked) -9999.10
(A, 9, unoccupied, unparked) 26.50
(A, 9, unoccupied, parked) 10.90
(A, 9, occupied, unparked) 26.50
(A, 9, occupied, parked) -9999.10
(B, 0, unoccupied, unparked) 71.29
(B, 0, unoccupied, parked) -999.10
(B, 0, occupied, unparked) 71.29
(B, 0, occupied, parked) -9999.10
(B, 1, unoccupied, unparked) 80.81
(B, 1, unoccupied, parked) 90.90
(B, 1, occupied, unparked) 32.00
(B, 1, occupied, parked) -9999.10
(B, 2, unoccupied, unparked) 71.81
(B, 2, unoccupied, parked) 80.90
(B, 2, occupied, unparked) 32.76
(B, 2, occupied, parked) -9999.10
(B, 3, unoccupied, unparked) 62.81
(B, 3, unoccupied, parked) 70.90
(B, 3, occupied, unparked) 31.18
(B, 3, occupied, parked) -9999.10
(B, 4, unoccupied, unparked) 53.81
(B, 4, unoccupied, parked) 60.90
(B, 4, occupied, unparked) 28.02
(B, 4, occupied, parked) -9999.10
(B, 5, unoccupied, unparked) 44.81
(B, 5, unoccupied, parked) 50.90
(B, 5, occupied, unparked) 23.87
(B, 5, occupied, parked) -9999.10
(B, 6, unoccupied, unparked) 35.81
(B, 6, unoccupied, parked) 40.90
(B, 6, occupied, unparked) 19.46
(B, 6, occupied, parked) -9999.10
(B, 7, unoccupied, unparked) 26.81
(B, 7, unoccupied, parked) 30.90
(B, 7, occupied, unparked) 16.61
(B, 7, occupied, parked) -9999.10
```

```
(B, 8, unoccupied, unparked) 19.57
(B, 8, unoccupied, parked) 20.90
(B, 8, occupied, unparked) 19.57
(B, 8, occupied, parked) -9999.10
(B, 9, unoccupied, unparked) 22.85
(B, 9, unoccupied, parked) 10.90
(B, 9, occupied, unparked) 22.85
(B, 9, occupied, parked) -9999.10
terminal_state 1.00,
```

and the optimal policy is

```
(A, 0, unoccupied, unparked) DRIVE
(A, 0, unoccupied, parked) EXIT
(A, 0, occupied, unparked) DRIVE
(A, 0, occupied, parked) EXIT
(A, 1, unoccupied, unparked) PARK
(A, 1, unoccupied, parked) EXIT
(A, 1, occupied, unparked) DRIVE
(A, 1, occupied, parked) EXIT
(A, 2, unoccupied, unparked) PARK
(A, 2, unoccupied, parked) EXIT
(A, 2, occupied, unparked) DRIVE
(A, 2, occupied, parked) EXIT
(A, 3, unoccupied, unparked) PARK
(A, 3, unoccupied, parked) EXIT
(A, 3, occupied, unparked) DRIVE
(A, 3, occupied, parked) EXIT
(A, 4, unoccupied, unparked) PARK
(A, 4, unoccupied, parked) EXIT
(A, 4, occupied, unparked) DRIVE
(A, 4, occupied, parked) EXIT
(A, 5, unoccupied, unparked) DRIVE
(A, 5, unoccupied, parked) EXIT
(A, 5, occupied, unparked) DRIVE
(A, 5, occupied, parked) EXIT
(A, 6, unoccupied, unparked) DRIVE
(A, 6, unoccupied, parked) EXIT
(A, 6, occupied, unparked) DRIVE
(A, 6, occupied, parked) EXIT
(A, 7, unoccupied, unparked) DRIVE
(A, 7, unoccupied, parked) EXIT
(A, 7, occupied, unparked) DRIVE
(A, 7, occupied, parked) EXIT
(A, 8, unoccupied, unparked) DRIVE
(A, 8, unoccupied, parked) EXIT
(A, 8, occupied, unparked) DRIVE
```

```
(A, 8, occupied, parked) EXIT
(A, 9, unoccupied, unparked) DRIVE
(A, 9, unoccupied, parked) EXIT
(A, 9, occupied, unparked) DRIVE
(A, 9, occupied, parked) EXIT
(B, 0, unoccupied, unparked) DRIVE
(B, 0, unoccupied, parked) EXIT
(B, 0, occupied, unparked) DRIVE
(B, 0, occupied, parked) EXIT
(B, 1, unoccupied, unparked) PARK
(B, 1, unoccupied, parked) EXIT
(B, 1, occupied, unparked) DRIVE
(B, 1, occupied, parked) EXIT
(B, 2, unoccupied, unparked) PARK
(B, 2, unoccupied, parked) EXIT
(B, 2, occupied, unparked) DRIVE
(B, 2, occupied, parked) EXIT
(B, 3, unoccupied, unparked) PARK
(B, 3, unoccupied, parked) EXIT
(B, 3, occupied, unparked) DRIVE
(B, 3, occupied, parked) EXIT
(B, 4, unoccupied, unparked) PARK
(B, 4, unoccupied, parked) EXIT
(B, 4, occupied, unparked) DRIVE
(B, 4, occupied, parked) EXIT
(B, 5, unoccupied, unparked) PARK
(B, 5, unoccupied, parked) EXIT
(B, 5, occupied, unparked) DRIVE
(B, 5, occupied, parked) EXIT
(B, 6, unoccupied, unparked) PARK
(B, 6, unoccupied, parked) EXIT
(B, 6, occupied, unparked) DRIVE
(B, 6, occupied, parked) EXIT
(B, 7, unoccupied, unparked) PARK
(B, 7, unoccupied, parked) EXIT
(B, 7, occupied, unparked) DRIVE
(B, 7, occupied, parked) EXIT
(B, 8, unoccupied, unparked) DRIVE
(B, 8, unoccupied, parked) EXIT
(B, 8, occupied, unparked) DRIVE
(B, 8, occupied, parked) EXIT
(B, 9, unoccupied, unparked) DRIVE
(B, 9, unoccupied, parked) EXIT
(B, 9, occupied, unparked) DRIVE
(B, 9, occupied, parked) EXIT
terminal_state NA.
```

Both value function and optimal policy make sense. The value function are negative for parking in the occupied states(causing collision) and the handicap spots, and the collision caused more negative reward than parking in the handicap spots. The highest value is 90.90 corresponding to states where the driver parks in spot in A1 or B1 when they are unoccupied and the values decrease as the spots getting farther away from store. This makes sense since the reward for parking farther from the store is less. From the optimal policy, when the agent parked, the next action is EXIT. When not parked, the action could be DRIVE or PARK. When not parked, if the spot is occupied, the optimal action is DRIVE to avoid collision which makes sense to avoid high cost of collision(-10000). When not parked and at a handicap spot, the optimal action is also DRIVE to avoid the high cost (-1000). When not parked, the unoccupied spots with optimal action PARK are A1, A2, A3, A4, B1, B2, B3, B4, B5, B6, B7. There are more PARK action in unoccupied spots from row B than from row A. It might be caused by the clockwise circular path of the driving from row A to row B.

For the second setting with driving cost -5, the value function is

```
(A, 0, unoccupied, unparked) 52.23
(A, 0, unoccupied, parked) -999.10
(A, 0, occupied, unparked) 52.23
(A, 0, occupied, parked) -9999.10
(A, 1, unoccupied, unparked) 76.81
(A, 1, unoccupied, parked) 90.90
(A, 1, occupied, unparked) 42.01
(A, 1, occupied, parked) -9999.10
(A, 2, unoccupied, unparked) 67.81
(A, 2, unoccupied, parked) 80.90
(A, 2, occupied, unparked) 39.07
(A, 2, occupied, parked) -9999.10
(A, 3, unoccupied, unparked) 58.81
(A, 3, unoccupied, parked) 70.90
(A, 3, occupied, unparked) 37.92
(A, 3, occupied, parked) -9999.10
(A, 4, unoccupied, unparked) 49.81
(A, 4, unoccupied, parked) 60.90
(A, 4, occupied, unparked) 36.65
(A, 4, occupied, parked) -9999.10
(A, 5, unoccupied, unparked) 40.81
(A, 5, unoccupied, parked) 50.90
(A, 5, occupied, unparked) 33.91
(A, 5, occupied, parked) -9999.10
(A, 6, unoccupied, unparked) 31.81
(A, 6, unoccupied, parked) 40.90
(A, 6, occupied, unparked) 29.24
(A, 6, occupied, parked) -9999.10
```

(A, 7, unoccupied, unparked) 22.94
(A, 7, unoccupied, parked) 30.90
(A, 7, occupied, unparked) 22.94
(A, 7, occupied, parked) -9999.10
(A, 8, unoccupied, unparked) 15.64
(A, 8, unoccupied, parked) 20.90
(A, 8, occupied, unparked) 15.64
(A, 8, occupied, parked) -9999.10
(A, 9, unoccupied, unparked) 9.08
(A, 9, unoccupied, parked) 10.90
(A, 9, occupied, unparked) 9.08
(A, 9, occupied, parked) -9999.10
(B, 0, unoccupied, unparked) 63.59
(B, 0, unoccupied, parked) -999.10
(B, 0, occupied, unparked) 63.59
(B, 0, occupied, parked) -9999.10
(B, 1, unoccupied, unparked) 76.81
(B, 1, unoccupied, parked) 90.90
(B, 1, occupied, unparked) 16.73
(B, 1, occupied, parked) -9999.10
(B, 2, unoccupied, unparked) 67.81
(B, 2, unoccupied, parked) 80.90
(B, 2, occupied, unparked) 19.29
(B, 2, occupied, parked) -9999.10
(B, 3, unoccupied, unparked) 58.81
(B, 3, unoccupied, parked) 70.90
(B, 3, occupied, unparked) 19.03
(B, 3, occupied, parked) -9999.10
(B, 4, unoccupied, unparked) 49.81
(B, 4, unoccupied, parked) 60.90
(B, 4, occupied, unparked) 16.80
(B, 4, occupied, parked) -9999.10
(B, 5, unoccupied, unparked) 40.81
(B, 5, unoccupied, parked) 50.90
(B, 5, occupied, unparked) 13.16
(B, 5, occupied, parked) -9999.10
(B, 6, unoccupied, unparked) 31.81
(B, 6, unoccupied, parked) 40.90
(B, 6, occupied, unparked) 8.56
(B, 6, occupied, parked) -9999.10
(B, 7, unoccupied, unparked) 22.81
(B, 7, unoccupied, parked) 30.90
(B, 7, occupied, unparked) 3.44
(B, 7, occupied, parked) -9999.10
(B, 8, unoccupied, unparked) 13.81
(B, 8, unoccupied, parked) 20.90

```
(B, 8, occupied, unparked) -0.97
(B, 8, occupied, parked) -9999.10
(B, 9, unoccupied, unparked) 4.81
(B, 9, unoccupied, parked) 10.90
(B, 9, occupied, unparked) 3.17
(B, 9, occupied, parked) -9999.10
terminal_state 1.00,
```

and the optimal policy is

```
(A, 0, unoccupied, unparked) DRIVE
(A, 0, unoccupied, parked) EXIT
(A, 0, occupied, unparked) DRIVE
(A, 0, occupied, parked) EXIT
(A, 1, unoccupied, unparked) PARK
(A, 1, unoccupied, parked) EXIT
(A, 1, occupied, unparked) DRIVE
(A, 1, occupied, parked) EXIT
(A, 2, unoccupied, unparked) PARK
(A, 2, unoccupied, parked) EXIT
(A, 2, occupied, unparked) DRIVE
(A, 2, occupied, parked) EXIT
(A, 3, unoccupied, unparked) PARK
(A, 3, unoccupied, parked) EXIT
(A, 3, occupied, unparked) DRIVE
(A, 3, occupied, parked) EXIT
(A, 4, unoccupied, unparked) PARK
(A, 4, unoccupied, parked) EXIT
(A, 4, occupied, unparked) DRIVE
(A, 4, occupied, parked) EXIT
(A, 5, unoccupied, unparked) PARK
(A, 5, unoccupied, parked) EXIT
(A, 5, occupied, unparked) DRIVE
(A, 5, occupied, parked) EXIT
(A, 6, unoccupied, unparked) PARK
(A, 6, unoccupied, parked) EXIT
(A, 6, occupied, unparked) DRIVE
(A, 6, occupied, parked) EXIT
(A, 7, unoccupied, unparked) DRIVE
(A, 7, unoccupied, parked) EXIT
(A, 7, occupied, unparked) DRIVE
(A, 7, occupied, parked) EXIT
(A, 8, unoccupied, unparked) DRIVE
(A, 8, unoccupied, parked) EXIT
(A, 8, occupied, unparked) DRIVE
(A, 8, occupied, parked) EXIT
(A, 9, unoccupied, unparked) DRIVE
```

```
(A, 9, unoccupied, parked) EXIT
(A, 9, occupied, unparked) DRIVE
(A, 9, occupied, parked) EXIT
(B, 0, unoccupied, unparked) DRIVE
(B, 0, unoccupied, parked) EXIT
(B, 0, occupied, unparked) DRIVE
(B, 0, occupied, parked) EXIT
(B, 1, unoccupied, unparked) PARK
(B, 1, unoccupied, parked) EXIT
(B, 1, occupied, unparked) DRIVE
(B, 1, occupied, parked) EXIT
(B, 2, unoccupied, unparked) PARK
(B, 2, unoccupied, parked) EXIT
(B, 2, occupied, unparked) DRIVE
(B, 2, occupied, parked) EXIT
(B, 3, unoccupied, unparked) PARK
(B, 3, unoccupied, parked) EXIT
(B, 3, occupied, unparked) DRIVE
(B, 3, occupied, parked) EXIT
(B, 4, unoccupied, unparked) PARK
(B, 4, unoccupied, parked) EXIT
(B, 4, occupied, unparked) DRIVE
(B, 4, occupied, parked) EXIT
(B, 5, unoccupied, unparked) PARK
(B, 5, unoccupied, parked) EXIT
(B, 5, occupied, unparked) DRIVE
(B, 5, occupied, parked) EXIT
(B, 6, unoccupied, unparked) PARK
(B, 6, unoccupied, parked) EXIT
(B, 6, occupied, unparked) DRIVE
(B, 6, occupied, parked) EXIT
(B, 7, unoccupied, unparked) PARK
(B, 7, unoccupied, parked) EXIT
(B, 7, occupied, unparked) DRIVE
(B, 7, occupied, parked) EXIT
(B, 8, unoccupied, unparked) PARK
(B, 8, unoccupied, parked) EXIT
(B, 8, occupied, unparked) DRIVE
(B, 8, occupied, parked) EXIT
(B, 9, unoccupied, unparked) PARK
(B, 9, unoccupied, parked) EXIT
(B, 9, occupied, unparked) DRIVE
(B, 9, occupied, parked) EXIT
terminal_state NA.
```

For the same reason stated in the first setting, both the value function and the optimal policy make sense. Because the driving cost here is higher, the value function is no larger than the value function derived from the first setting which makes sense, and when not parked, the unoccupied spots with optimal action PARK are A1, A2, A3, A4, A5, A6, B1, B2, B3, B4, B5, B6, B7, B8, B9. Compared with the optimal policy in the first setting, there are more unoccupied spots with action $PARK$. The change in both the value function and the optimal policy match our expectation.