

## First Visit Monte Carlo Prediction

Optimal policy, discount = 1 , 20.000 iterations

```
Policy:
3 ['→', '→', '→', None]
2 ['↑', '↑', '↑', '↑']
1 ['→', '↑', '←', '←']
0 [None, '↑', '↑', '↑']
    0,    1,    2,    3

value_grid:
[38.0, 39.0, 40.0, 0]
[37.0, 38.0, 39.0, 40.0]
[36.0, 37.0, 36.0, 35.0]
[0, 36.0, 35.0, 34.0]
```

Optimal policy, discount = 0.9, 20.000 iterations

```
Policy:
3 ['→', '→', '→', None]
2 ['↑', '↑', '↑', '↑']
1 ['→', '↑', '←', '↑']
0 [None, '↑', '↑', '↑']
    0,    1,    2,    3

value_grid:
[30.5, 35.0, 40.0, 0]
[26.45, 30.5, 35.0, 40.0]
[22.81, 26.45, 22.8, 19.52]
[0, 22.81, 19.52, 16.57]
```

Random policy, discount = 1, 20.000 iterations

```
Policy:
3 ['→', '→', '→', None]
2 ['↑', '↑', '↑', '↑']
1 ['↓', '←', '←', '↑']
0 [None, '←', '←', '↑']
    0,    1,    2,    3

value_grid:
[12.31, 19.19, 40.0, 0]
[9.11, 9.1, 10.35, 40.0]
[10.0, 1.79, -11.22, -16.22]
[0, 10.0, -16.22, -18.28]
```

Random policy, discount = 0.9, 20.000 iterations

```
Policy:
3 ['→', '→', '→', None]
2 ['↓', '↑', '↑', '↑']
1 ['↓', '←', '↑', '↑']
0 [None, '←', '←', '←']
    0,    1,    2,    3

value_grid:
[5.61, 13.4, 40.0, 0]
[4.63, 5.07, 11.92, 40.0]
[10.0, 2.38, -3.87, -5.94]
[0, 10.0, -5.87, -6.66]
```