Lab 11 Analysis

Daniel “NP-Complete” Nanetti-Palacios, *Box 4426*
Tyler “Dew-while loop” Dewey, *Box 3426*

Problem 5: *Value Iteration*

Tests:

```
$ ./value_iteration .99 .01 4x3.mdp
0.776184
0.716627
0.650616
0.843935
-0.040000
0.592541
0.905096
0.641327
0.560027
1.000000
-1.000000
0.337952

$ ./value_iteration .99999 .001 4x3.mdp
0.811522
0.761512
0.705252
0.867784
-0.040000
0.655243
0.917795
0.660255
0.611348
1.000000
-1.000000
0.387860
```

Predict:

We expect to see a direct path to a +1 terminal state that avoids coming close to -1 terminal states to have a greater cumulative value than both the shortest path and longer paths. In all instances distance dulls the effect of the terminal state's reward. States closer to -1 terminal states have a lower value than states farther away, and similarly, states closer to +1 terminal states have a higher value. We expect the state adjacent to both a +1 and a -1 (12,0) to have a value near 0.5. In the tests we ran, the blank state had a permanent value of -0.04, and we expect the same for all the blank states in this grid.

Experiment:

With blanks:

```
  |   0    |  1    |  2    |  3    |  4    |  5    |  6    |  7
  +--------------------------------------------------------------
0 | -0.232 -0.182 -0.132 -0.091 -0.040  0.340  0.396  0.446   Wrapped
1 | -0.040 -0.040 -0.081 -0.036 -0.040  0.290 -0.040 -0.040     to
2 | -0.040 -0.104 -0.036  0.019 -0.040  0.240 -0.040 -0.040     next
3 | -1.000 -0.248 -0.040  0.076  0.134  0.184 -0.040 -1.000     line

  |   8    |  9    |  10   |  11   |  12   |  13   |  14   |  15
  +--------------------------------------------------------------
0 | 0.496  0.446 -0.040 -1.000  0.944  1.000 -0.040 -3999.9
1 | 0.559 -0.040  0.674 -0.040  0.894 -0.040 -0.040 -0.040
2 | 0.609  0.669  0.724  0.787  0.837 -0.040 -0.040  1.000
3 | 0.578  0.624  0.669 -0.040  0.799  0.844  0.894  0.944
```

Without blanks:

```
  |   0    |  1    |  2    |  3    |  4    |  5    |  6    |  7
  +--------------------------------------------------------------
0 | -0.232 -0.182 -0.132 -0.091 [redacted]  0.340  0.396  0.446   Wrapped
1 | [redacted] [redacted] -0.081 -0.036 [redacted]  0.290 [redacted] [redacted]     to
2 | [redacted] -0.104 -0.036  0.019 [redacted]  0.240 [redacted] [redacted]     next
3 | -1.000 -0.248 [redacted]  0.076  0.134  0.184 [redacted] -1.000     line


  |   8    |  9    |  10   |  11   |  12   |  13   |  14   |  15
  +--------------------------------------------------------------
0 | 0.496  0.446 [redacted] -1.000  0.944  1.000 [redacted] -3999.9
1 | 0.559 [redacted]  0.674 [redacted]  0.894 [redacted] [redacted] [redacted]
2 | 0.609  0.669  0.724  0.787  0.837 [redacted] [redacted]  1.000
3 | 0.578  0.624  0.669 [redacted]  0.799  0.844  0.894  0.944
```

Reflect:

Our prediction was correct that states closer to -1 terminals have lower value than those farther away (reverse for +1 terminal states). We were also correct in assuming that blank squares would have a value of -0.04 (their reward). We were very wrong about the state (12, 0), which has a value the same as the the other next-to-plus-1-terminal state. This does make sense in hindsight, because there is no way you could end up in (11,0) (the -1 terminal) from (12,0) if you move towards (13,0) (the +1 terminal). We were also surprised that if you follow the path of maximum increase that you end up going to (12, 0) rather than (15, 2), with the large difference between (12, 1) = 0,894 and (12, 3) = 0.799. In hindsight, this also makes sense (especially with our previous observation) and is consistent with our rule that states farther away from a +1 terminal have a lower value. We were also surprised and amused that (15, 0) had such a negative value.

Problem 6: Policy Iteration

Test:

```
    $ ./policy_iteration .99 .01 4x3.mdp
    3
    0
    0               → → → ↑
    3               ↑ ↑ ↑ ↑
    0               ↑ ← ↑ ←
    2
    3               → → → ↑
    0               ↑ ■ ↑ ↑
    0               ↑ ← ↑ ←
    0
    0
    2

    $ ./policy_iteration .999 .001 4x3.mdp
    3
    0
    0                → → → ↑
    3                ↑ ↑ ↑ ↑
    0                ↑ ← ← ←
    2
    3
    0                → → → +
    2                ↑ ■ ↑ -
    0                ↑ ← ← ←
    0
    2
```

Predict:

We observed in our tests that the policy is very risk-averse. When placed in the bottom left corner, rather than just going back up, the policy for 4x3 takes an agent all the way back to (0,2) (the bottom left corner). We assume that the policy generated for 16x4 would be similar, taking great lengths to find the "safest" path to a +1, avoiding paths that could (because of the action's unpredictable result) put the agent in or near a -1 terminal state. We predict that there will be a main path that never gets closer than it must to -1 states, where the states nearer to the -1 point towards that path. We predict that the policy will favor the +1 state farther away from the -1, not the one at (13, 0) that is separated from a -1 only by 1 square.

Results:

With blank spaces:

```
→ → ↓ ↓ ↑ → → → ↓ ← ↑ ↑ → ↑ ↑ ↑
↑ ↑ ↓ ↓ ↑ ↑ ↑ ↑ ↓ ↑ ↓ ↑ ↑ ↑ ↑ ↑
↑ → → ↓ ↑ ↑ ↑ ↑ → → → → ↑ ↑ ↑ ↑
↑ ↑ ↑ → → ↑ ↑ ↑ → → ↑ ↑ → → → ↑
```

Without blank spaces:

```
→ → ↓ ↓ ■ → → → ↓ ← ■ ↑ → ↑ ■ ↑
■ ■ ↓ ↓ ■ ↑ ■ ■ ↓ ■ ↓ ■ ↑ ■ ■ ■
■ → → ↓ ■ ↑ ■ ■ → → → → ↑ ■ ■ ↑
↑ ↑ ■ → → ↑ ■ ↑ → → ↑ ■ → → → ↑
```

W/o blanks & with terminal states:

```
→ → ↓ ↓ ■ → → → ↓ ← ■ − → + ■ ↑
■ ■ ↓ ↓ ■ ↑ ■ ■ ↓ ■ ↓ ■ ↑ ■ ■ ■
■ → → ↓ ■ ↑ ■ ■ → → → → ↑ ■ ■ +
− ↑ ■ → → ↑ ■ − → → ↑ ■ → → → ↑
```

Reflect:

We were correct to some degree about the agent following the safest path. In each state near -1 terminals, the policy says to move away, which we predicted. Similar to our results in Value Iteration, our predictions were also wrong when dealing with the space (12, 0) that lies between a -1 terminal and +1 terminal states. We believed the policy of that state would direct the agent away from this area to follow a safer path to the +1 terminal at (15, 2), but we were wrong since there would be no way it would end up in the negative terminal if it takes the rational action of going towards the +1 terminal.