CSE 150 Homework 5 Report

Eli Eshel-Krohn A11283400
Joshua Anthony A14281769
Andrew Hwang A11570188

1 - Policy Improvement

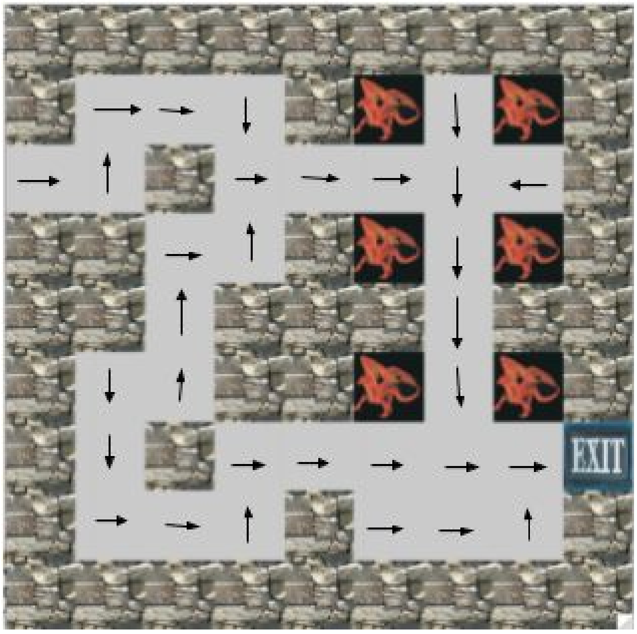
See included pdfs for individual turnins.

2 - Value and Policy Iteration

a/b)

```
(3, 69.2373929193, EAST)
(11, 70.7782298269, EAST)
(12, 69.993466006, NORTH)
(15, 84.2378190458, SOUTH)
(16, 85.5318438299, SOUTH)
(17, 86.5008261413, EAST)
(20, 71.5723387025, EAST)
(22, 71.5726248949, EAST)
(23, 70.7889450429, NORTH)
(24, 70.394384476, NORTH)
(26, 87.4716048185, EAST)
(29, 72.3753718247, SOUTH)
(30, 73.1874152733, EAST)
(31, 72.3753792769, NORTH)
(34, 89.4454661306, EAST)
(35, 88.4530329598, NORTH)
(39, 74.0305166864, EAST)
(43, 90.4490341156, EAST)
(48, 74.8617172617, EAST)
(52, 91.4638620149, EAST)
(53, 93.0926052336, EAST)
(56, 74.8828115598, SOUTH)
(57, 79.8023062776, SOUTH)
(58, 82.4793299449, SOUTH)
(59, 87.9647533329, SOUTH)
(60, 89.073584086, SOUTH)
(61, 94.9450793787, EAST)
(62, 94.1531975863, EAST)
(66, 22.8558703901, EAST)
(70, 96.2560589106, EAST)
```

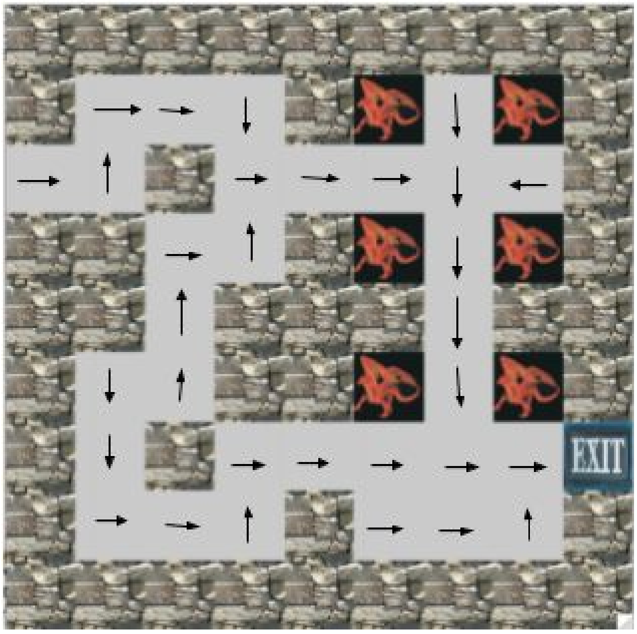
```
(71, 95.1886139797, NORTH)
(79, 100.0, NORTH)
```



Value Iteration Algorithm:

The algorithm iterates through each state and determines the optimal directional movement and value weight, given the current utility function which starts as a vector of zeros. At each state, all four possible actions are checked and the maximum value and action are stored. The algorithm would terminate once the utility vector converges and it would return the policy and the calculated utility function.

c)



The resulting map is the same as in part b.

Policy Iteration Algorithm:

This algorithm was similar to the value iteration algorithm, but it updates the policy vector until it converges, rather than utility vector, starting with a random policy. During the policy evaluation stage, the utility for a given policy is updated until it converges. After this, the algorithm checks, for each state, if the best move for the new utility is the same as the move provided by the current policy. If the new utility’s argmax is greater than the policy then the policy is changed and the utility is recalculated. This is done until the policy doesn’t need to be updated as it has converged. Since the policy converges faster than the utility, this algorithm will result in a speedup.

d) Josh - I contributed to the code for parts 2a, 2b, and 2c.
Eli - I contributed to parts b and c and the report.
Andrew - I contributed to parts b and c and the report.