

Reinforcement Learning – HW 3

Alon Ressler , 201547510, alonress@gmail.com
Eliran Shabat, 201602877, shabat.eliran@gmail.com

Programing Question 1

We implemented the off-policy model-based learning by following the given instructions.

Path to execution directory:

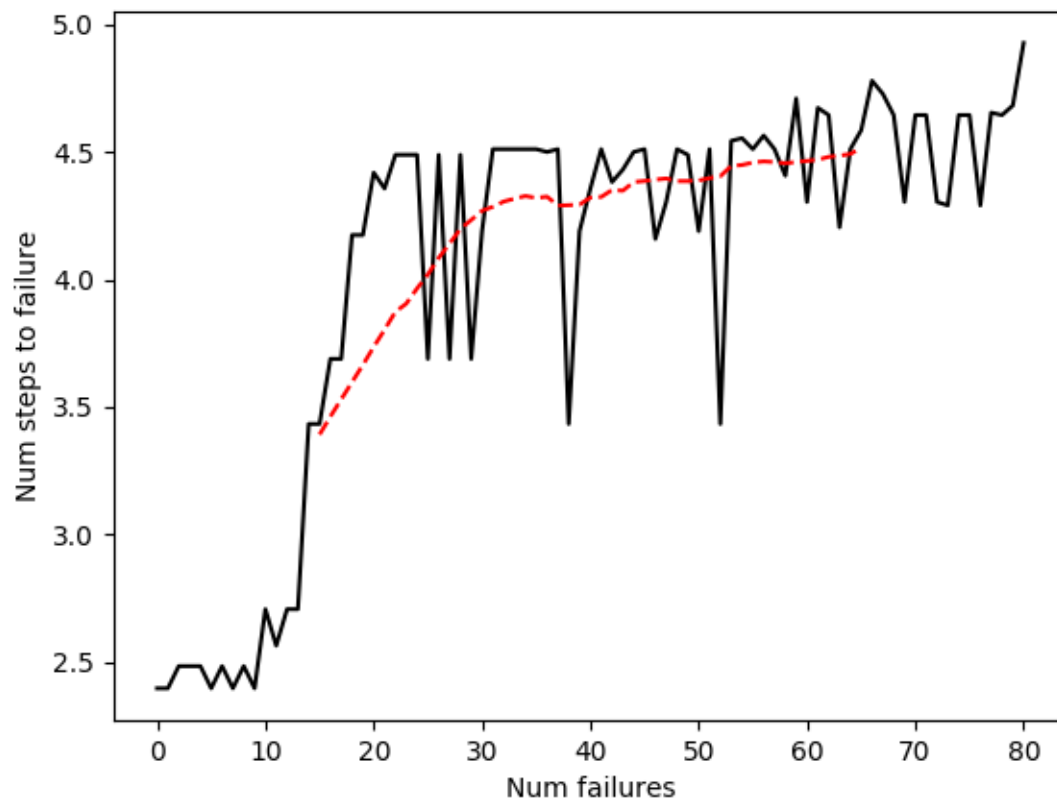
/specific/a/home/cc/students/cs/eliranshabat/courses/msc/RL/hw3

The code is implemented in the file:

/specific/a/home/cc/students/cs/eliranshabat/courses/msc/RL/hw3/control.py

The algorithm converged after 81 trials.

Learning curve:



Programing Question 2

We implemented Q-learning for the FrozenLake environment using a tabular and function-approximation methods.

Path to execution directory:

/specific/a/home/cc/students/cs/eliranshabat/courses/msc/RL/hw3

The code is implemented in the following files:

/specific/a/home/cc/students/cs/eliranshabat/courses/msc/RL/tabular_Q.py

/specific/a/home/cc/students/cs/eliranshabat/courses/msc/RL/network_Q.py

For the tabular case, we received a score of 0.2135 and the following Q-function:

```
[[1.66773351e-01 6.74114362e-02 8.76232402e-02 1.29524059e-01]
 [1.68149072e-04 1.10859524e-03 5.46226793e-04 1.09788639e-01]
 [6.43360602e-04 1.31912990e-02 8.36529762e-03 5.35260040e-03]
 [4.23242725e-03 2.92972347e-03 6.01371014e-04 6.48914148e-02]
 [1.71976554e-01 7.91620353e-02 2.41553043e-03 3.49254573e-02]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [5.95170746e-01 3.20803632e-06 7.68882310e-07 7.63933325e-06]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [6.30565673e-03 5.71717603e-02 8.35091431e-02 3.91005079e-01]
 [1.10947271e-01 5.71858642e-01 1.34448082e-01 1.52485667e-02]
 [8.73417089e-01 1.18821246e-03 1.18281904e-03 2.18672163e-03]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [1.52912333e-02 1.71805960e-01 6.46930216e-01 1.28222412e-01]
 [5.65964732e-01 8.51173973e-01 4.84896638e-01 2.56591228e-01]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]
```

Using the function-approximation method, we received a score of 0.0375.

The score we got by using function approximation is lower then the score received using the tabular approach. It seems reasonable because function approximation is useful when dealing with large state space. In our case, the state space is small, so using a table we can get a good approximation for the MDP.