

Ahmet Akman 2442366  
*Middle East Technical University*  
*Electrical and Electronics Engineering*

May 24, 2024

## **HOMEWORK 3 — Report**

# 1 Questions

1.1 Agent:

1.2 Environment:

1.3 Reward:

1.4 Policy:

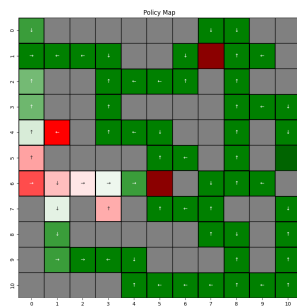
1.5 Exploration:

1.6 Exploitation:

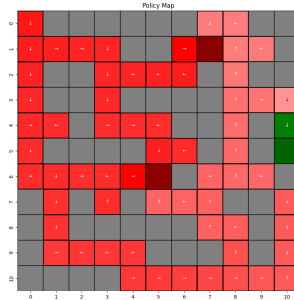
# 2 Experimental Work

2.1 Temporal Difference Learning Default Parameters

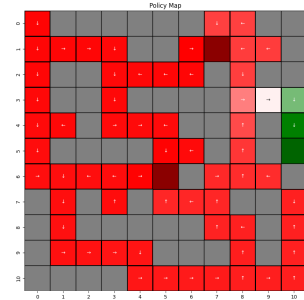
2.2 Q-Learning Default Parameters



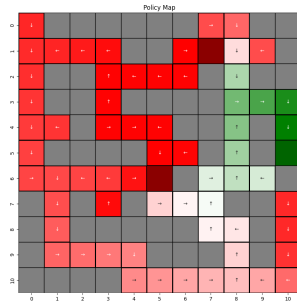
(a) Episode 1.



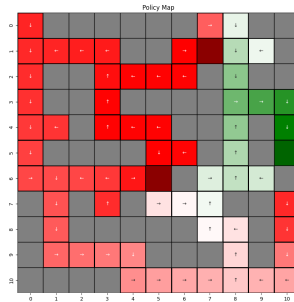
(b) Episode 50



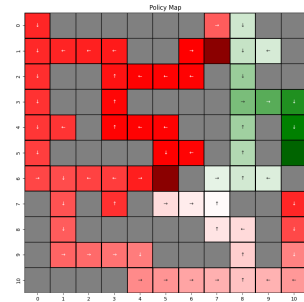
(c) Episode 100



(d) Episode 1000.

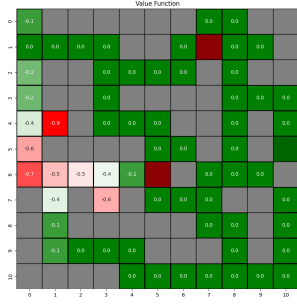


(e) Episode 5000

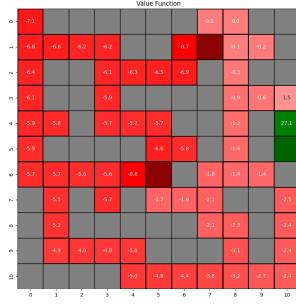


(f) Episode 10000

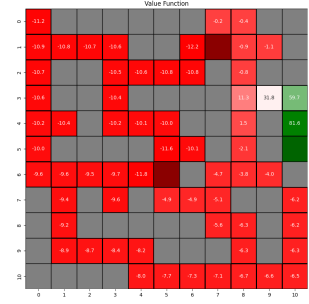
Figure 1: Evolution of policy maps throughout episodes.



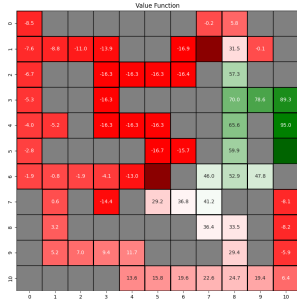
(a) Episode 1.



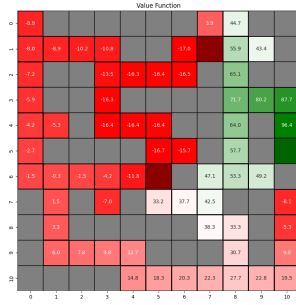
(b) Episode 50



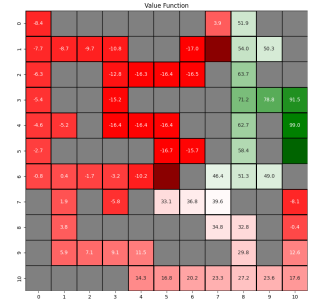
(c) Episode 100



(d) Episode 1000.



(e) Episode 5000



(f) Episode 10000

Figure 2: Evolution of value function throughout episodes.

- 2.3 Effect of Alpha in Temporal Difference Learning
- 2.4 Effect of Alpha in Q-Learning
- 2.5 Effect of Gamma in Temporal Difference Learning
- 2.6 Effect of Gamma in Q-Learning
- 2.7 Effect of Epsilon in Temporal Difference Learning
- 2.8 Effect of Epsilon in Q-Learning

### 3 Discussions

- 3.1 Q1
- 3.2 Q2
- 3.3 Q3
- 3.4 Q4
- 3.5 Q5
- 3.6 Q6
- 3.7 Q7
- 3.8 Q8
- 3.9 Q9
- 3.10 Q10

## Appendix

The code set used throughout this homework is provided as follows.