Warsaw University of Technology
Faculty of Electronics and Information Technology
Specialization: Computer Science
Course: Introduction to Artificial Intelligence

PROJECT DOCUMENTATION

Q-learning Algorithm

Author:
Monika Jung

Student ID:
331384

1 Introduction

This project implements and analyzes the Q-learning algorithm to solve the Cliff Walking environment from the Gymnasium Toolkit. The main objectives are:

- Develop a universal Q-learning implementation for discrete state-action spaces
- Investigate algorithm performance under different hyperparameters

2 Q-learning Implementation

The core implementation features:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \cdot \max Q(s',a') - Q(s,a) \right]$$

where:

- α – learning rate,
- γ – discount factor,
- r – reward received after taking action a in state s,
- s' – resulting next state.

3 Experimental Setup

The Cliff Walking environment (4×12 grid) features:

- Start state: (3,0)
- Goal state: (3,11)
- Cliff states: (3,1) to (3,10) with -100 reward
- Step reward: -1

3.1 Tested Parameters

- Learning rates (α): 0.01, 0.1, 0.5, 1.0
- Exploration probabilities (ϵ): 0.01, 0.05, 0.1, 0.5
- Training episodes: 100, 500, 1000, 2000

For each setting was recorded:

- Success rate (percentage of episodes finishing at goal),
- Cliff fail rate (episodes ending in cliff),
- Average number of steps per episode,
- Average reward in the last 10 episodes.

Results for Exploration Probability = 0.01

Learning rate	Episodes	Success rate	Cliff fail rate	Avg steps	Avg last 10 rewards
0.01	100	6.0%	94.0%	227.4	-66.7
0.01	500	57.4%	42.6%	112.9	-55.5
0.01	1000	76.7%	23.3%	78.7	-36.2
0.01	2000	87.2%	12.8%	53.2	-21.2
0.1	100	79.1%	20.9%	79.3	-37.7
0.1	500	93.4%	6.6%	30.5	-13.2
0.1	1000	95.4%	4.6%	22	-13.2
0.1	2000	96.2%	3.8%	17.7	-13.1
0.5	100	94.2%	5.8%	31.4	-13.3
0.5	500	96.7%	3.3%	17	-13.2
0.5	1000	96.9%	3.1%	15.2	-13.2
0.5	2000	97.2%	2.8%	14.3	-13.1
1	100	96.0%	4.0%	23.3	-13.3
1	500	97.3%	2.7%	15.3	-13.1
1	1000	97.0%	3.0%	14.3	-13.1
1	2000	97.3%	2.7%	13.8	-13.1

Results for Exploration Probability = 0.05

Learning rate	Episodes	Success rate	Cliff fail rate	Avg steps	Avg last 10 rewards
0.01	100	8.3%	91.7%	236.8	-70
0.01	500	52.6%	47.4%	115.5	-52.1
0.01	1000	69.2%	30.8%	80.3	-33.8
0.01	2000	79.3%	20.7%	54.2	-21.1
0.1	100	69.7%	30.3%	81.4	-36.1
0.1	500	84.6%	15.4%	31.7	-13.8
0.1	1000	85.8%	14.2%	23.3	-14.1
0.1	2000	86.2%	13.8%	19.1	-13.8
0.5	100	85.2%	14.8%	32.9	-13.9
0.5	500	86.4%	13.6%	18.5	-13.7
0.5	1000	86.7%	13.3%	16.6	-13.7
0.5	2000	86.5%	13.5%	15.7	-13.8
1	100	87.4%	12.6%	24.7	-13.6
1	500	88.0%	12.0%	16.7	-13.6
1	1000	87.6%	12.4%	15.7	-13.7
1	2000	87.5%	12.5%	15.2	-13.8

Results for Exploration Probability = 0.1

Learning rate	Episodes	Success rate	Cliff fail rate	Avg steps	Avg last 10 rewards
0.01	100	9.5%	90.5%	251.4	-68.9
0.01	500	46.7%	53.3%	119.2	-47.2
0.01	1000	60.4%	39.6%	82.4	-32
0.01	2000	69.6%	30.4%	55.8	-21.1
0.1	100	60.7%	39.3%	83.8	-35.3
0.1	500	74.7%	25.3%	33.5	-14.7
0.1	1000	74.3%	25.7%	25.3	-14.4
0.1	2000	74.7%	25.3%	21.2	-14.4
0.5	100	73.8%	26.2%	35.1	-14.8
0.5	500	75.1%	24.9%	20.6	-14.8
0.5	1000	74.7%	25.3%	18.8	-14.2
0.5	2000	74.8%	25.2%	17.8	-14.5
1	100	76.8%	23.2%	26.7	-14.5
1	500	77.7%	22.3%	18.5	-14.5
1	1000	76.9%	23.1%	17.6	-14.6
1	2000	77.2%	22.8%	17	-14.7

Results for Exploration Probability = 0.5

Learning rate	Episodes	Success rate	Cliff fail rate	Avg steps	Avg last 10 rewards
0.01	100	5.7%	94.3%	384.1	-26.7
0.01	500	16.5%	83.5%	161.8	-33.1
0.01	1000	20.5%	79.5%	113	-26.2
0.01	2000	22.4%	77.6%	84.9	-25
0.1	100	21.4%	78.6%	122.2	-27.9
0.1	500	21.7%	78.3%	71.5	-22
0.1	1000	20.9%	79.1%	66.2	-21.8
0.1	2000	24.0%	76.0%	58	-23.7
0.5	100	21.8%	78.2%	76.3	-23.6
0.5	500	25.7%	74.3%	55	-26.2
0.5	1000	28.6%	71.4%	49.2	-25
0.5	2000	29.8%	70.2%	46.4	-23.2
1	100	31.0%	69.0%	56.2	-23.7
1	500	31.0%	69.0%	46.3	-26.8
1	1000	31.9%	68.1%	44.9	-26.3
1	2000	31.2%	68.8%	44.3	-25

4 Results and Analysis

The best configuration achieved near-optimal performance:

- $\alpha = 1$, $\epsilon = 0.01$, 2000 episodes
- Success rate: 97.3%
- Average steps: 13.8
- Final rewards: -13.1 (theoretical minimum: -13)

Key observations:

- Low $\alpha = 0.01$ requires more episodes to reach a higher success rate.
- For bigger α, the number of episodes does not play such a significant role.
- The bigger the exploration probability, the worse the outcomes.
- The higher the learning rate, the better the success rate for all exploration probabilities.
- Cliff Walking heavily penalizes random exploration