EARIN Lab

Task 6

Variant 3

Group 3

By: Krzysztof Kotowski and Juan Manuel Aristizabal Henao

Task description:

Creation of an implementation of the Q-Learning algorithm to solve an RL problem usign the Gymnasium Library. For this implementation the **CliffWalking** environment was used.

In this environment the goal is to cross a 4x12 gridwold from a starting position [3, 0] to an ending position [3,11], while avoiding falling to a cliff. The starting and ending positions are in opposite positions of the cliff, so the algorithm has to learn how to go around the cliff without falling into it.

Action Space:

One value tuple containing:

- 0 : For move **up**.
- 1 : For move **right**.
- 2 : For move **down**.
- 3: For move **left**.

Observation Space:

There are 3 x 12 + 1 possible states. Which does not include the cliff and the ending position since it results in the end of the episode.

Rewards:

Each time step has a –1 reward, if the player steps on the cliff it gets a –100 reward.

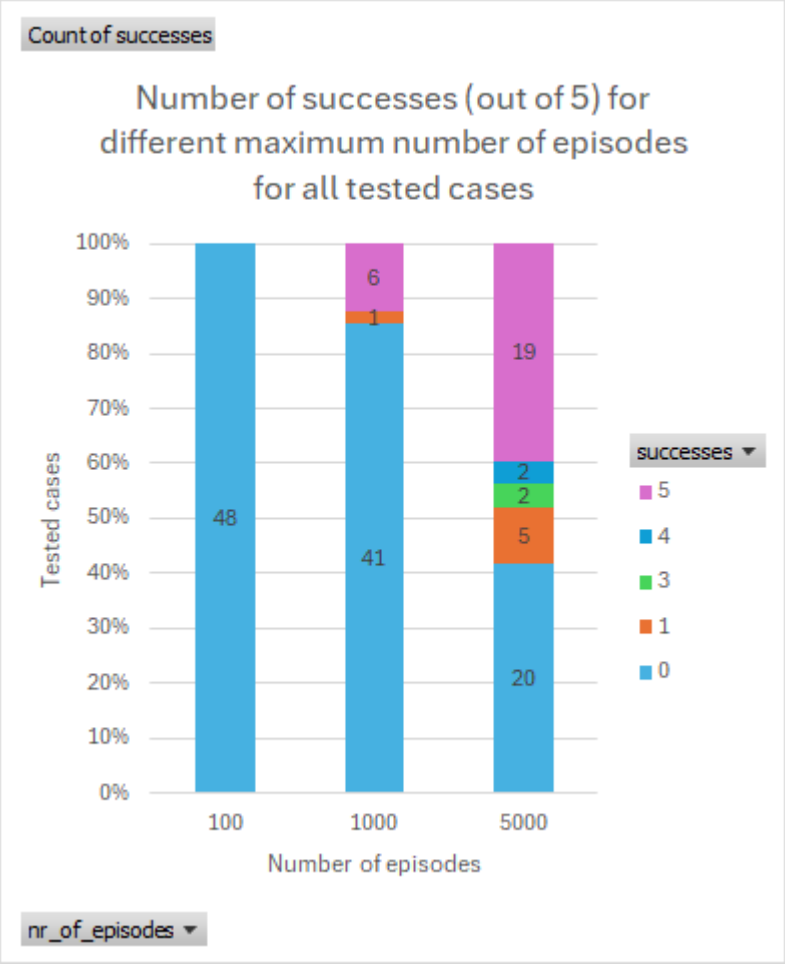
Model training:

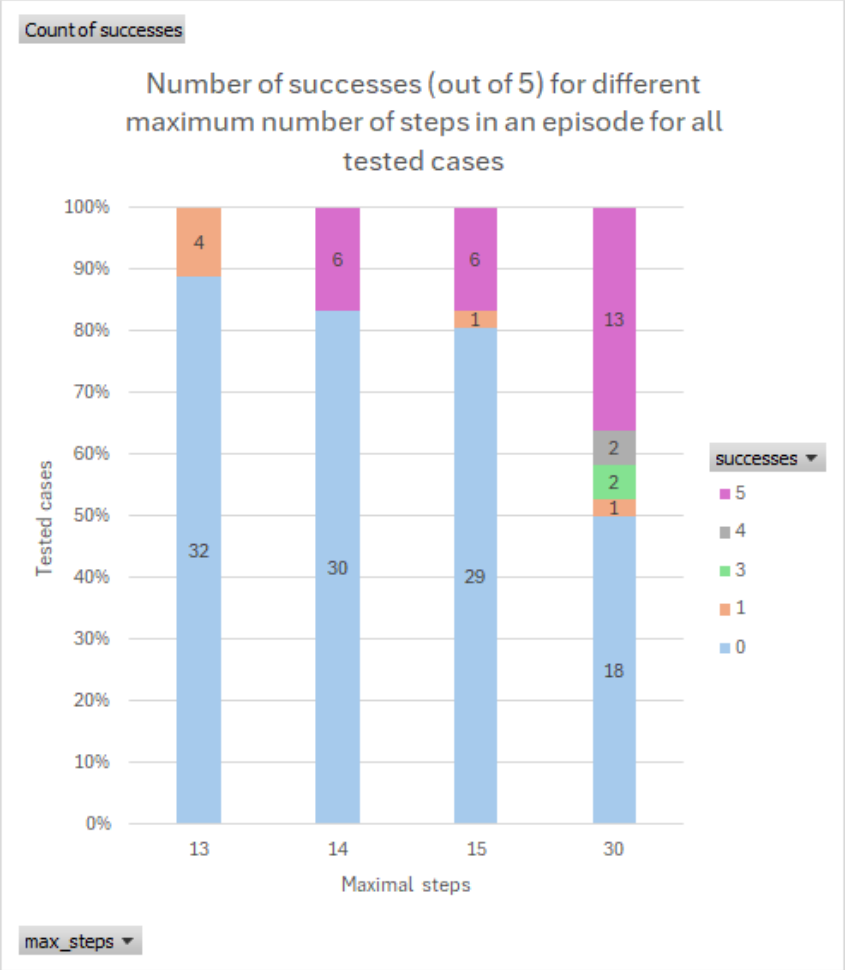
I trained the model on all combinations of these parameters:

- Nr_of_episodes: [100, 1000, 5000]
- Max_steps: [13, 14, 15, 30]
    - **Note: 13 is the minimal number of steps required to go from start to finish
- Learning_rate: [0.01, 0.1]
- Gamma: [0.99]
- Epsilon: [0.1, 0.9, 1.0]
- Epsilon_decay: [0.999, 1.0]

For all of those cases I trained model 5 times to account for randomness in training process (“flipping a coin” when choosing exploration vs exploitation strategy)

Results:





Interestingly, when the model was trained on maximum number of steps being equal to number of steps in perfect completion, models were able to complete the path only 1 out of 5 times for the same parameters. It probably stemmed from the fact that the model started by going immediately to the right, left or down. For 14 max steps it suddenly grows to 5 out of 5 completions and more models were successful

	13	14	15	30	Max steps
100	0,00%	0,00%	0,00%	0,00%	
1000	0,00%	0,00%	1,67%	50,00%	
5000	6,67%	50,00%	50,00%	83,33%	
Episodes					

Effect of learning rate and epsilon:

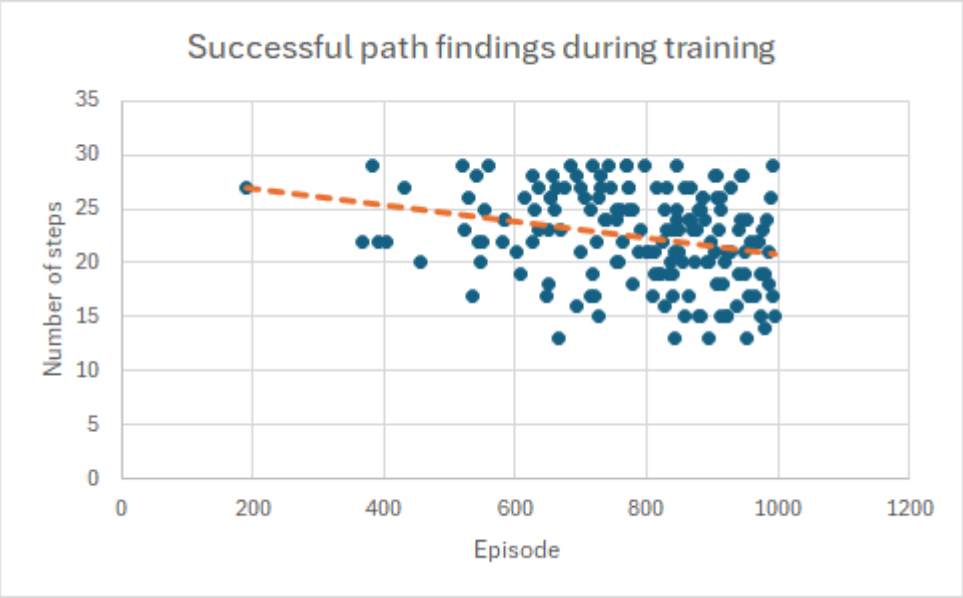
Epsilon	Percentage of successes
0,1	17,4%
0,9	16,7%
1	16,3%

Learning rate	Percentage of successes
0,01	4,6%
0,1	28,9%

Exploration vs exploitation choosing had some measurable effect on model efficiency, but the learning rate was a very fundamental parameter

Single training case study:

- nr_of_episodes = 1000
- max_steps = 30
- learning_rate = 0.1
- gamma = 0.99
- epsilon = 0.9
- epsilon_decay = 0.999



Episodes that found the final state: 165 out of 1000 = 16,5%