

Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

Wstęp do Sztucznej Inteligencji
Semestr 24L
Sprawozdanie z ćwiczenia nr 6

Q-Learning

Mikołaj Wewiór

Warszawa,
4 VI 2024

1. Opis problemu

Celem ćwiczenia było zaimplementowanie uczenia funkcji Q jako tablicy dla odpowiedniego środowiska - w tym przypadku było to **FrozenLake-v1** z biblioteki **gym**. Środowiskiem tym, jest mapa o wymiarze 8 na 8, po której porusza się agent. Może on wykonywać ruchy w 4 kierunkach. Dlatego tablica Q ma wymiary 64 na 4. Ważnym aspektem zadania był atrybut **is_slippery**, który wprowadzał do gry poślizg, przez co agent przy wykonaniu danego ruchu miał 33% na zamierzony ruch oraz po 33% na ruch w sąsiednich kierunkach. Ta nieduża zmiana znacznie zmienia omawiany problem.

2. Wstępne parametry algorytmu

Do wyboru akcji zbadano dwa podejścia. Pierwszym była strategia ϵ -zachłanna, a drugim strategia oparta na rozkładzie Boltzmann.

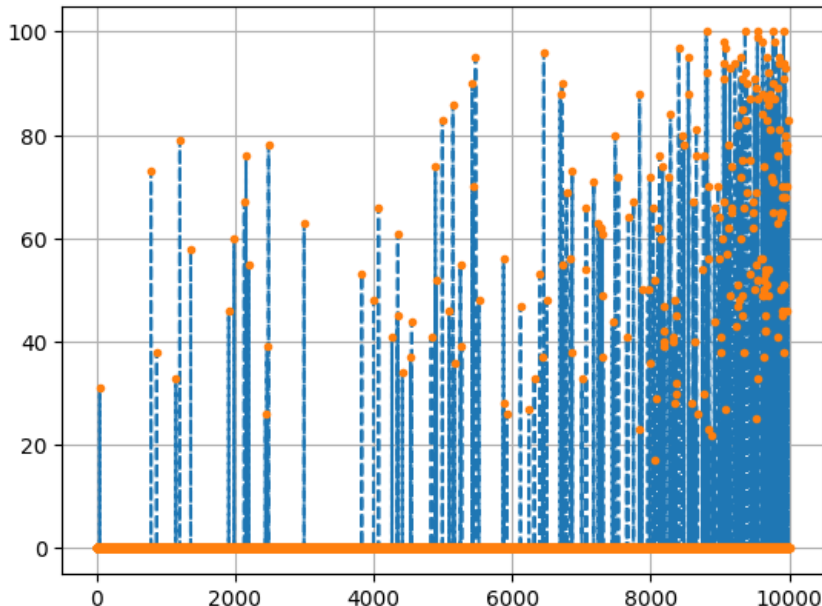
Po próbach znalezienia przyzwoitych nastaw badania rozpoczęto od następujących wartości:

- α : 0.9 (parametr uczenia),
- γ : 0.95 (parametr dyskontowania),
- ϵ : 0.95 (w trakcie uczenia),
- T : 1 (w trakcie uczenia).

Badania przeprowadzono dla 5 różnych ziaren (seedów) generatora liczb losowych o następujących wartościach: 318407, 4062024, 19122020, 27112002, 99815612. Uczenie zostało przeprowadzone dla 10000 epok, w każdej z nich maksymalnie po 192 ruchy.

Takie ustawienia dały następujące średnie wyniki:

- procentowe osiągnięcie nagrody: 35, 31%,
- procentowe osiągnięcie nagrody w ostatnich 100 krokach: 42.2%,
- średnia liczba kroków potrzebna do osiągnięcia celu: 67.85,
- odchylenie standardowe liczby kroków: 19.06.

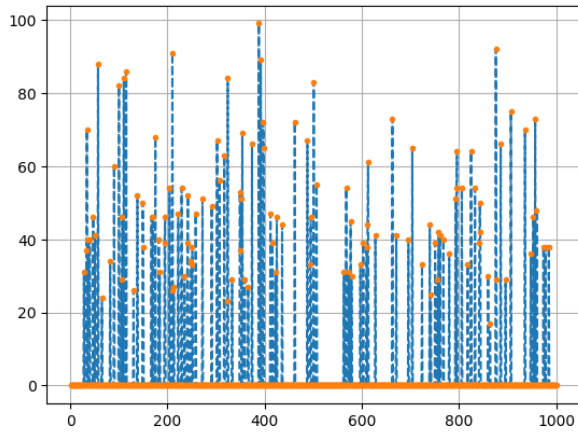
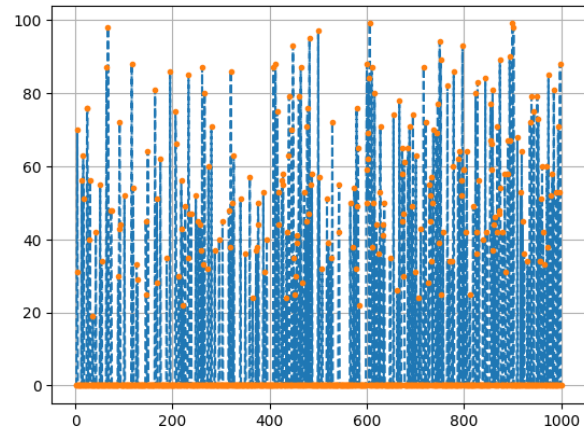
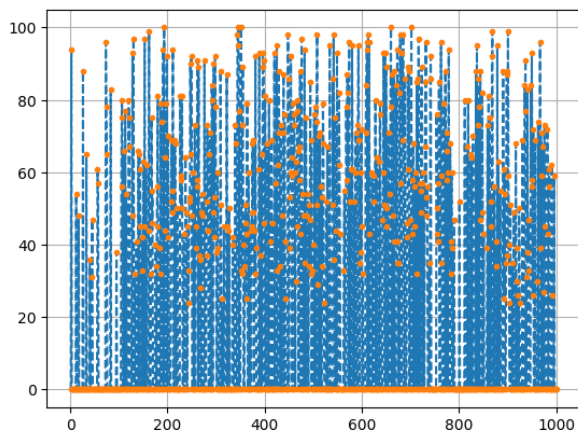
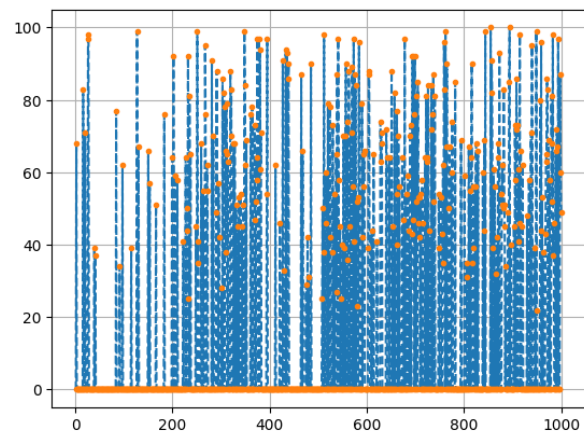


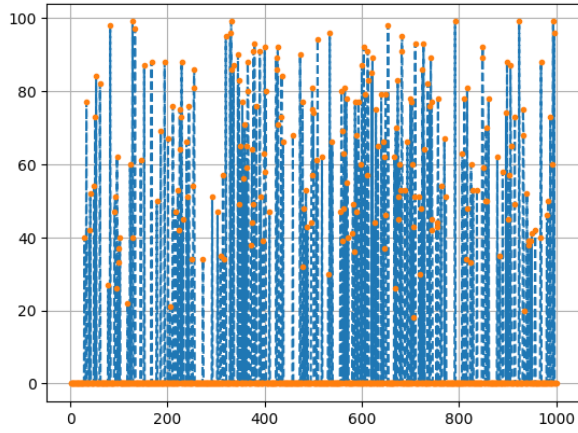
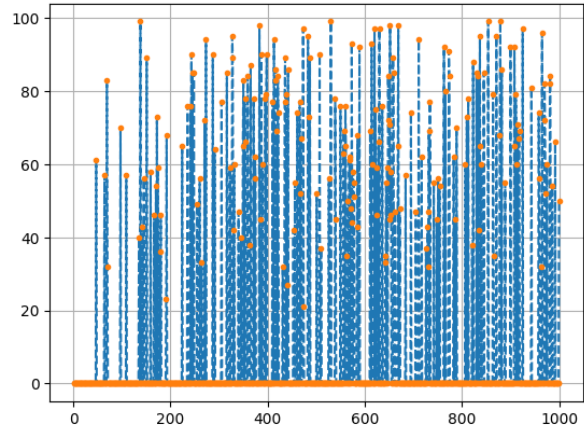
Rys. 1: liczba kroków potrzebnych do osiągnięcia nagrody. $\alpha = 0.9$, $\gamma = 0.95$, $\epsilon = 0.95$

3. Parametr uczenia α

α	średnia liczba osiągnięć nagrody w 100 próbach	liczba kroków do nagrody	
		średnia	odchylenie
0.9	42.2	67.9	19.1
0.8	39.4	64.4	20.8
0.7	42.2	63.4	19.4
0.6	45.6	66.5	20.7
0.5	49.6	61.8	20.3
0.4	50.0	63.2	20.3
0.3	50.6	60.0	19.2
0.2	55.2	61.3	19.9
0.1	55.2	61.3	19.5
0.05	40.0	60.5	19.2
0.02	18.2	56.4	21.1

Tab. 1

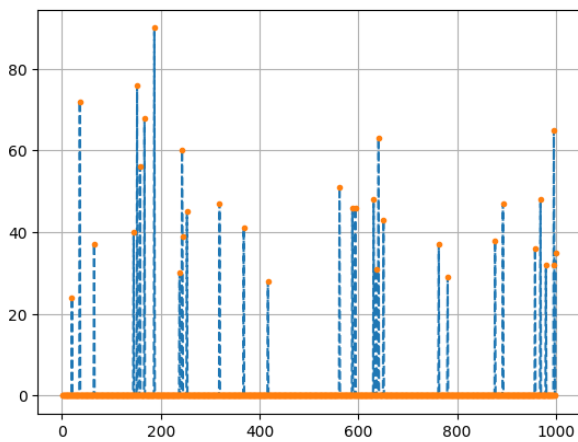
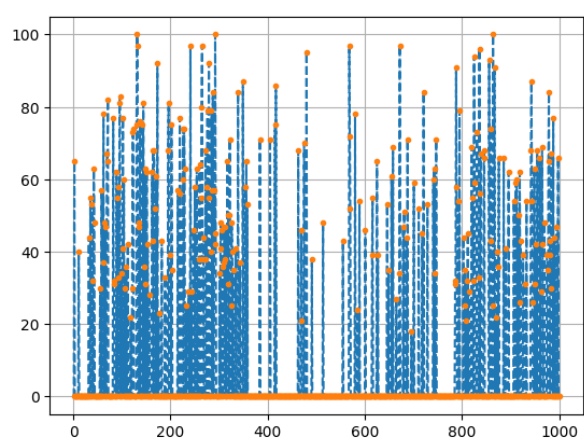
(a) $\alpha = 0.02$, $\gamma = 0.95$, $\epsilon = 0.95$ (b) $\alpha = 0.05$, $\gamma = 0.95$, $\epsilon = 0.95$ (a) $\alpha = 0.1$, $\gamma = 0.95$, $\epsilon = 0.95$ (b) $\alpha = 0.4$, $\gamma = 0.95$, $\epsilon = 0.95$

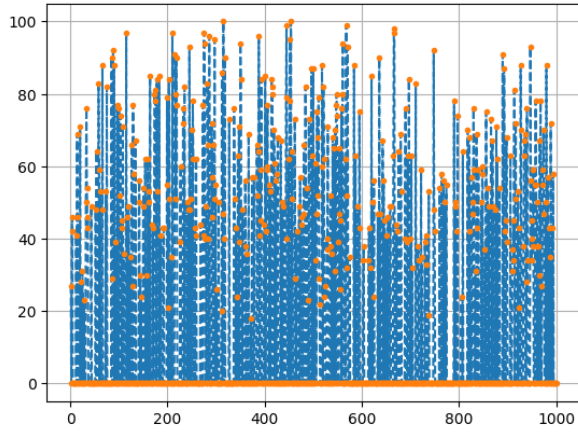
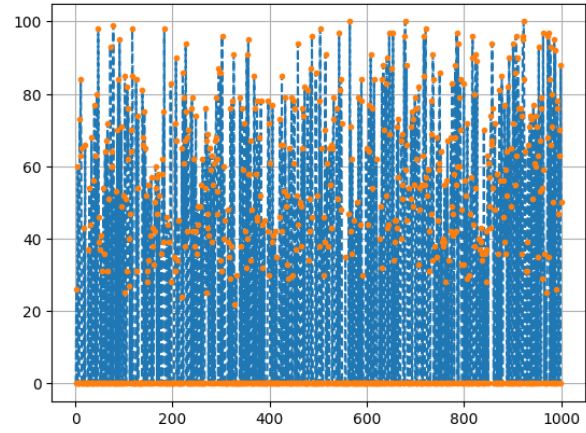
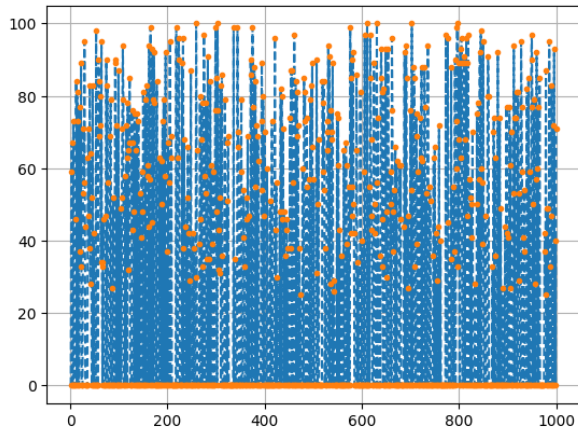
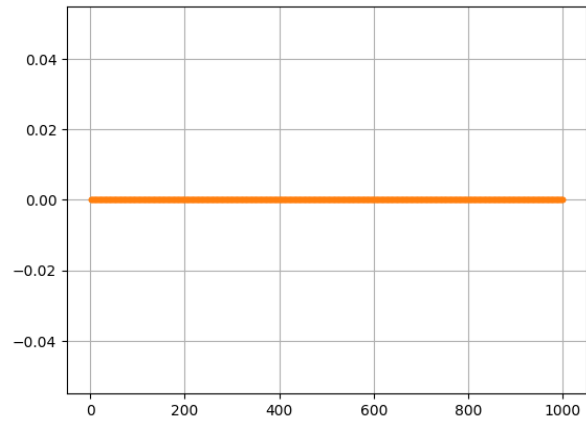
(a) $\alpha = 0.7, \gamma = 0.95, \epsilon = 0.95$ (b) $\alpha = 0.9, \gamma = 0.95, \epsilon = 0.95$

4. Parametr dyskontowania γ

γ	średnia liczba osiągnięć nagrody w 100 próbach	liczba kroków do nagrody	
		średnia	odchylenie
1.1	0.0	-	-
1.0	49.2	65.1	20.1
0.95	55.2	61.3	19.5
0.9	49.0	56.6	18.6
0.8	36.2	57.9	20.1
0.7	32.6	55.9	19.2
0.6	15.6	47.1	17.0

Tab. 2

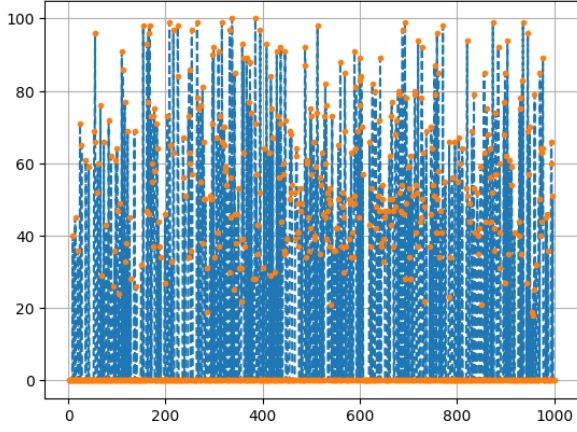
(a) $\alpha = 0.1, \gamma = 0.6, \epsilon = 0.95$ (b) $\alpha = 0.1, \gamma = 0.8, \epsilon = 0.95$

(a) $\alpha = 0.1, \gamma = 0.9, \epsilon = 0.95$ (b) $\alpha = 0.1, \gamma = 0.095, \epsilon = 0.95$ (a) $\alpha = 0.1, \gamma = 1.0, \epsilon = 0.95$ (b) $\alpha = 0.1, \gamma = 1.1, \epsilon = 0.95$

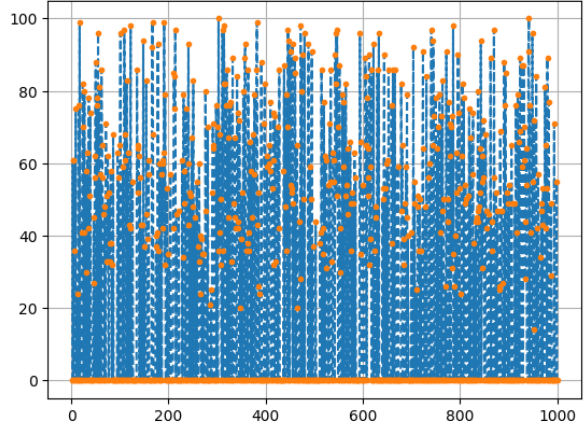
5. Parametr eksploracji ϵ

ϵ	średnia liczba osiągnięć nagrody w 100 próbach	liczba kroków do nagrody	
		średnia	odchylenie
0.99	53.6	59.6	19.3
0.95	55.2	61.3	19.5
0.9	55.6	59.6	19.5
0.8	57.0	58.3	19.8
0.7	43.0	59.5	19.9
0.6	10.0	63.1	19.6

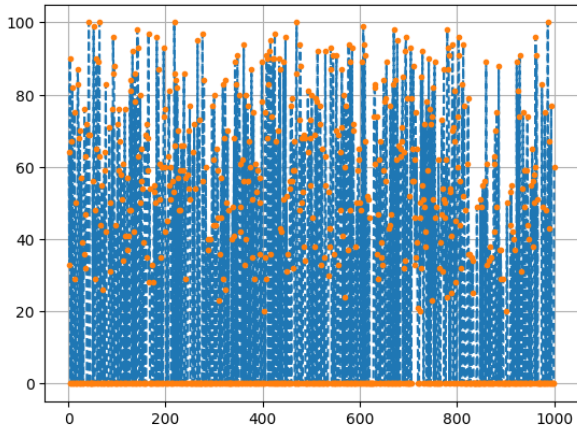
Tab. 3



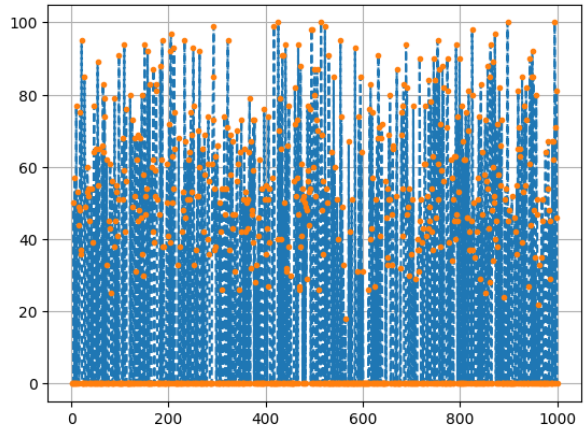
(a) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.99$



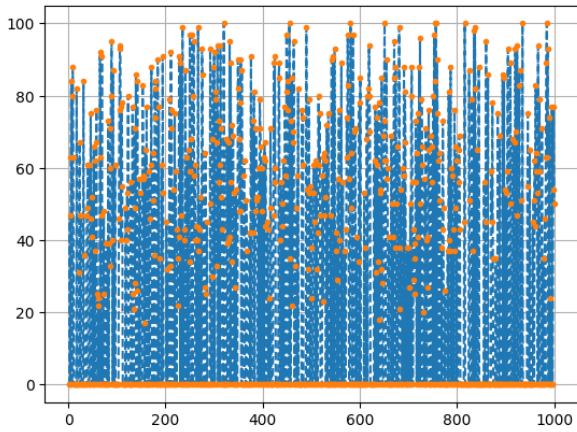
(b) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.95$



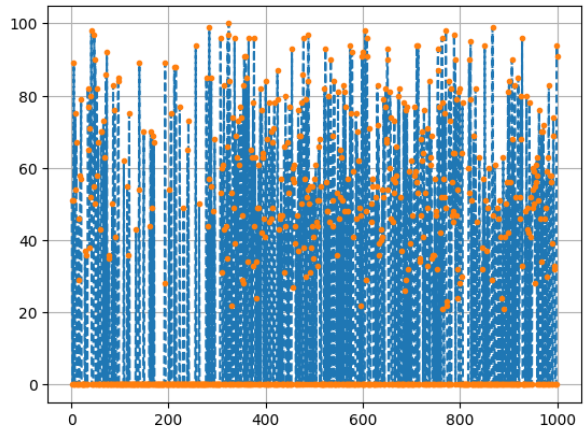
(a) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.9$



(b) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.8$



(a) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.7$



(b) $\alpha = 0.1, \gamma = 0.95, \epsilon = 0.6$

6. Podsumowanie