

Sieci Samouczące Się

Laboratorium 3 – Sprawozdanie

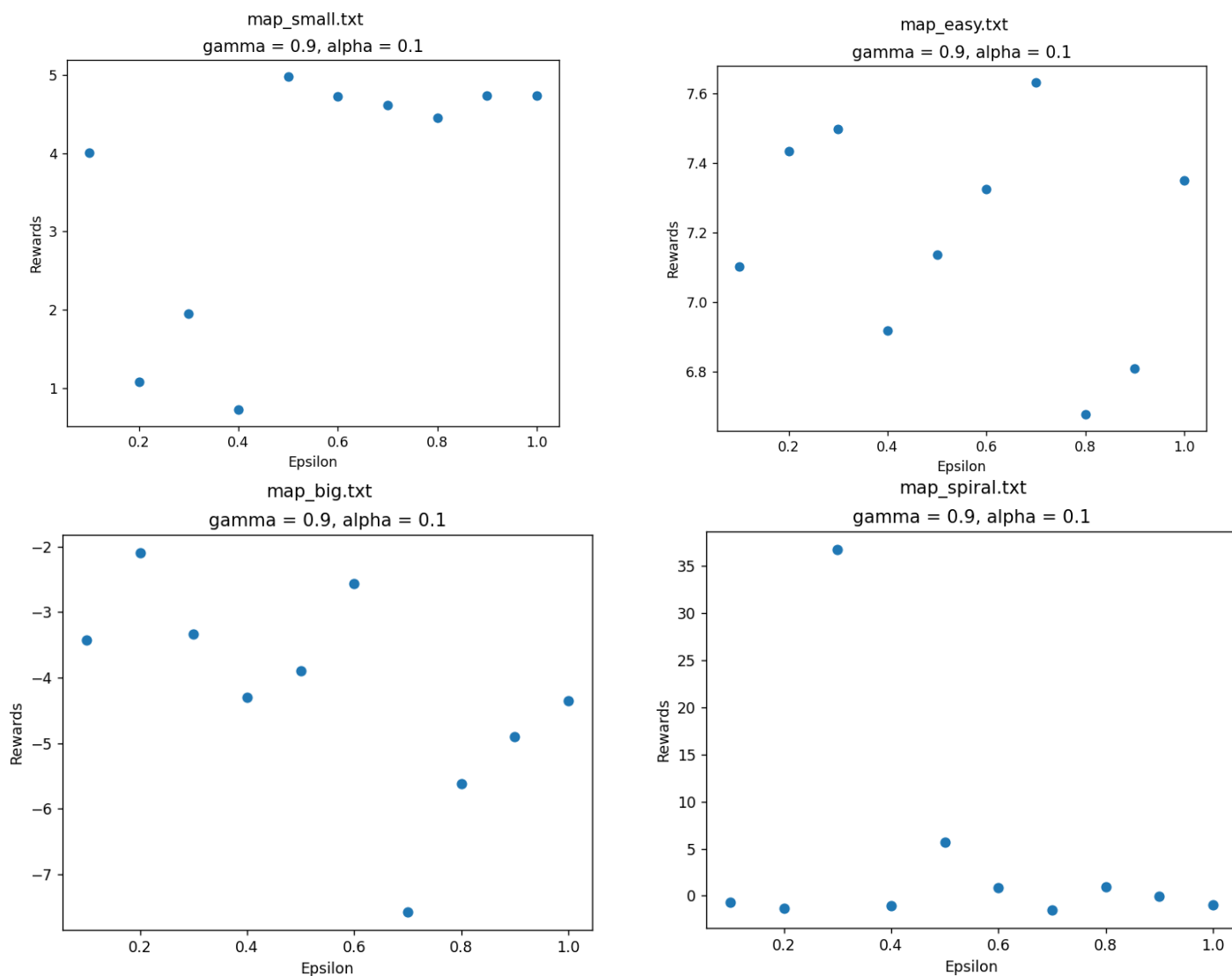
Metoda Q-learning – zaimplementowana eksploracja E – zachłanna, zestaw V1

Algorytm został zaimplementowany według slajdów wykładowych z ustalonymi uproszczeniami: stały epsilon (nie zmieniający się w pętli uczenia), brak implementacji eksploracji bolzmana.

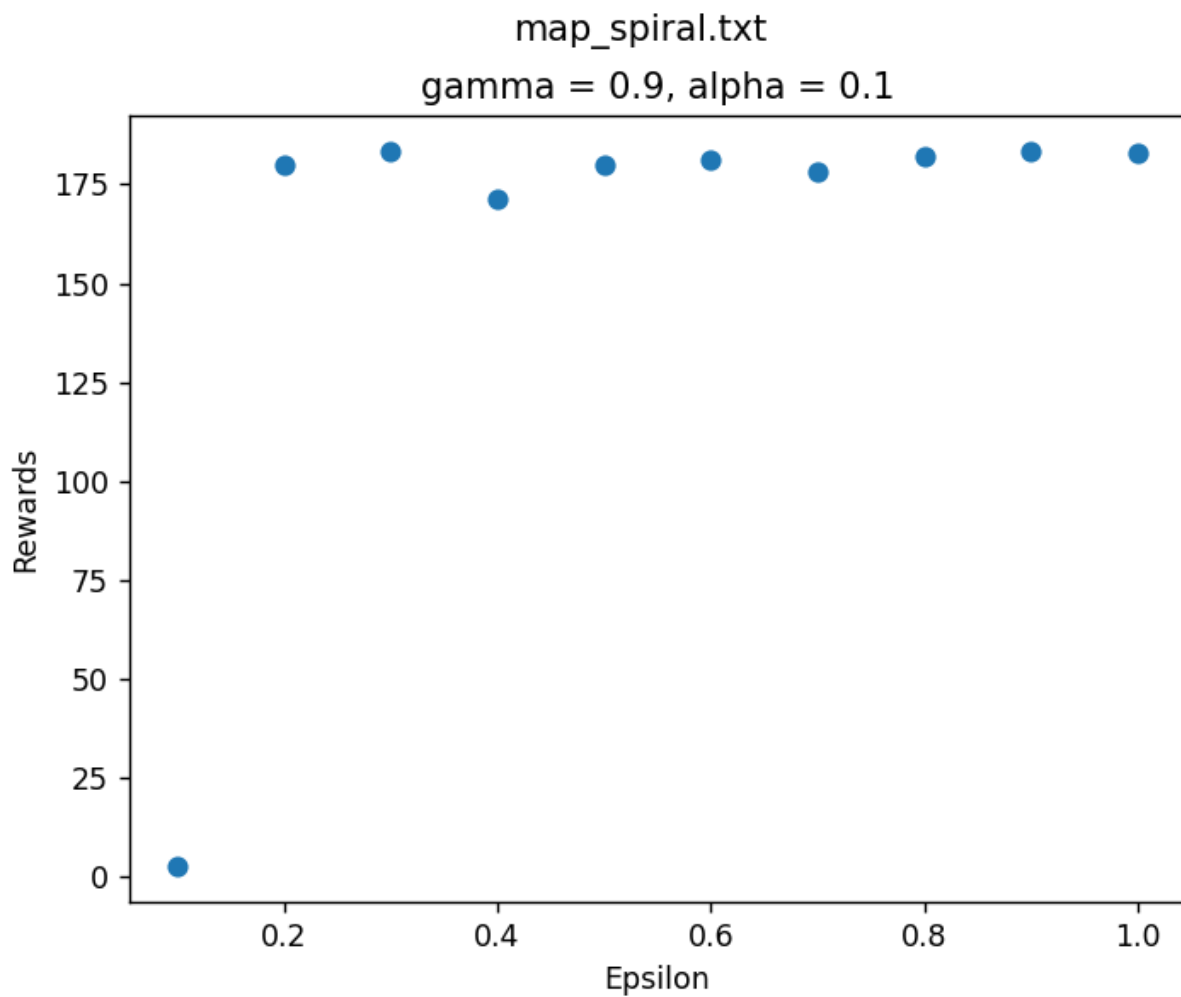
Działanie algorytmu:

W ustalonej pętli epizodów inicjujemy stan początkowy, następnie w wewnętrznej pętli dopóki nie osiągniemy stanu końcowego wykonujemy akcję według ustalonego algorytmu eksploracji. Zapisujemy nagrodę, następny stan i aktualizujemy tablicę użyteczności akcji Q.

Wyniki dla liczby epizodów równej 1000:



Wyniki dla map_spiral przy ustawieniu 10 000 epizodów:



Implementacja eksploracji E-zachłannej:

```
def epsilon_greedy(state):  
    if random.uniform(0,1) < epsilon:  
        # print("epsilon")  
        return random.randint(1,4)  
    else:  
        # print("max")  
        return (np.argmax(Q[state[0], state[1], :])+1)
```

Im większy epsilon, tym większa szansa na wybór losowej akcji.

Wnioski: Widać, że epsilon, w aktualnej implementacji ma znikomy wpływ na środowisko. Jednak najlepiej widać wpływ na ostatnim wykresie – bardzo mały epsilon oznacza słabe wyniki.