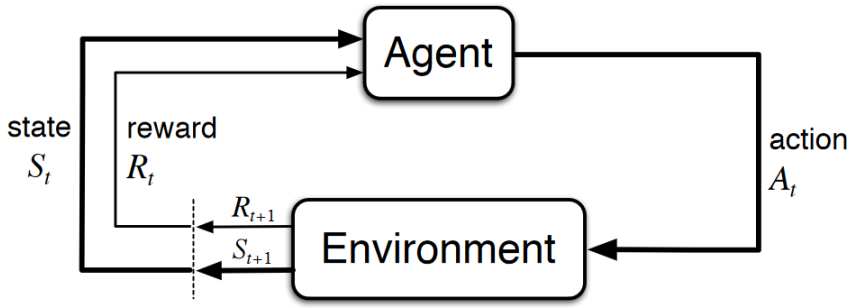
Uczenie ze wzmocnieniem (Reinforcement Learning)

Q-learning



Proces decyzyjny markowa (MDP)

- Cel: podejmowanie najlepszych decyzji w danym momencie (stanie)
- Tablica Q[stan][akcja] - wartość “jakości” decyzji. Im większa wartość, tym lepsza decyzja
- Aktualizacja tablicy Q, aby zapamiętać czy akcja w danym stanie była dobra czy nie

Initialized

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	327	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	499	0	0	0	0	0	0

Training

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	328	-2.30108105	-1.97092096	-2.30357004	-2.20591839	-10.3607344	-8.5583017
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	499	9.96984239	4.02706992	12.96022777	29	3.32877873	3.38230603

$$Q(state, action) \leftarrow (1-\alpha)Q(state, action) + \alpha\left(reward + \gamma \max_a Q(next\ state, all\ actions)\right)$$

***Równanie Bellmana**, gdzie α - learning rate (współczynnik szybkości aktualizacji Q-values), γ - (discount factor, jak ważne są przyszłe nagrody od aktualnego stanu – czy mają duży wpływ na aktualną decyzję?)*

Exploration vs Exploitation

- *Epsilon Greedy Strategy* – w każdym kroku uczenia musimy wybrać akcję, którą podejmiemy. Gdy sugerujemy się tylko akcjami z Q-table możemy popaść w lokalne minima. Aby pozwolić agentowi “eksplorować” środowisko, zgodnie z losową wartościa oraz parametrem ϵ pozwolimy mu podejmować losowe decyzje. Wartość ϵ z czasem się zmniejsza, aby kłaść nacisk na nauczone wartości

Zadanie:

1. Zapoznaj się z:
    a. Informacjami o RL: https://github.com/dennybritz/reinforcement-learning/tree/master/Introduction
    b. Środowiskiem OpenAI Gym - http://gym.openai.com/docs/
    c. Implementacją Q-learning'u w problemie Taxi – **self_driving_taxi.*py***
2. Zaimplementuj algorytm Q-learning dla wybranego środowiska OpenAI Gym (za wyłączeniem cartpole). Implementację oraz wyniki przedstaw w sprawozdaniu.