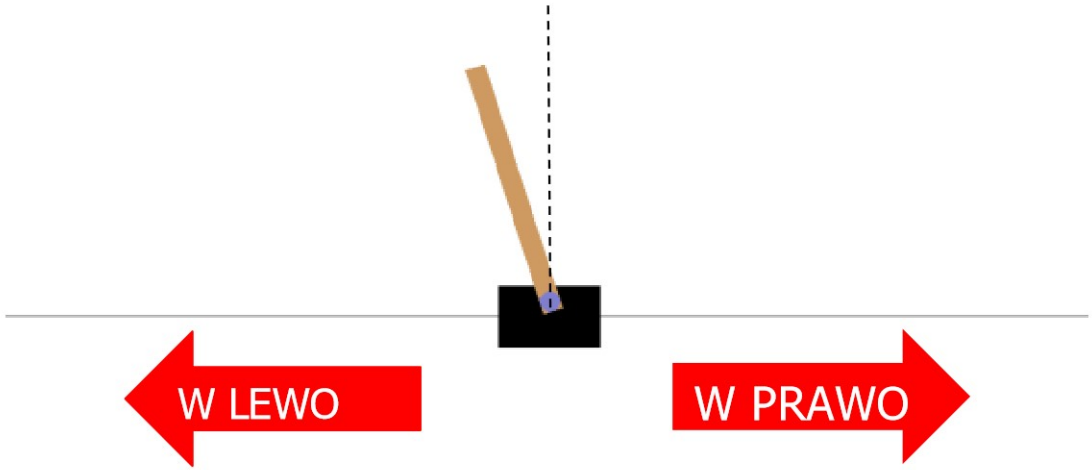
Wprowadzenie do uczenia ze wzmocnieniem

część 4

Cart Pole



Poruszamy wózkiem w prawo lub w lewo, tak aby jak najdłużej utrzymać drążek w przedziale pewnych kątów (-12 i 12 stopni) liczonych od położenia pionowego.

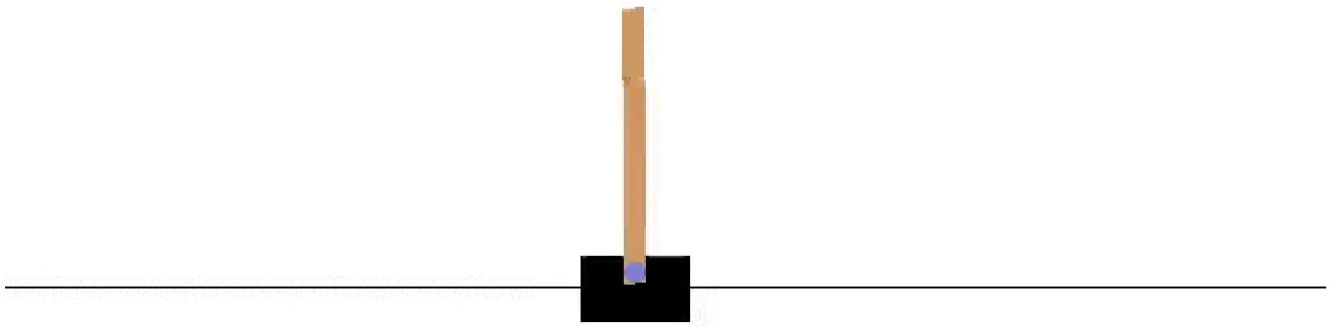
Koniec następuje w momencie przekroczenia wartości granicznych kątów lub wyjechania wózka poza dozwolony obszar.

Cart Pole

Przykładowa implementacja sprzętowa:

https://www.youtube.com/watch?v=5Q14EjnOJZc

Cart Pole



W przypadku środowiska Cart Pole stan opisywany jest za pomocą 4 parametrów:

- położenie (position)
- prędkość (velocity)
- odchylenie od pionu (pole angle)
- prędkość kątowa (pole velocity)

Cart Pole

Wartości minimalne i maksymalne dla 4 parametrów:

```
Observation:
    Type: Box(4)
    Num Observation                Min            Max
    0   Cart Position             -4.8            4.8
    1   Cart Velocity             -Inf            Inf
    2   Pole Angle                 -24 deg        24 deg
    3   Pole Velocity At Tip      -Inf            Inf
```

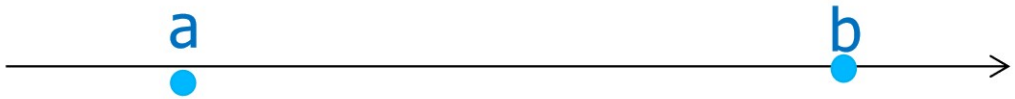
A zatem przestrzeń stanów jest 4 wymiarowa i ciągła, a nie dyskretna jak w przypadku środowiska Frozen Lake.

Ponieważ przedziały w jakich zawarte są 4 parametry są nieskończone zatem zbiór stanów jest zatem nieskończony!!!

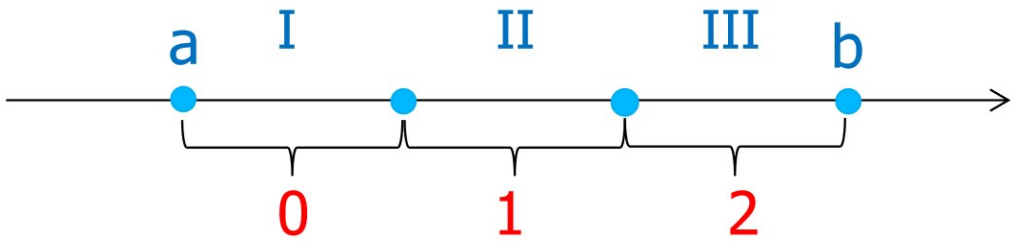
Cart Pole

Nieskończoną przestrzeń stanów możemy zdyskretyzować.

Rozważmy jeden parametr przyjmujący wartości z przedziału [a,b].



Przedział ten możemy podzielić na skończoną ilość odcinków (np. 3) oznaczonych I, II i III:



Jeżeli parametr przyjmuje wartość z odcinka I wówczas jego dyskretna wartość wynosi 0, jeżeli przyjmuje wartość z odcinka II wówczas jego dyskretna wartość wynosi 1 itd.

Cart Pole

Propozycja dyskretyzacji dla środowiska Cart Pole umieszczona jest w pliku CartPole_random_action_discretization.py

- Zmienna Cart Position – przedział [-4.8,4.8] dzielony jest na n_s[0] części (n_s to tablica zdefiniowana w powyższym pliku).
- Zmienna Cart Velocity – ograniczona do wartości z przedziału [-1.0,1.0], który dzielony jest na n_s[1] części.
- Zmienna Pole Angle – przedział [-24°,24°] dzielony jest na n_s[2] części.
- Zmienna Pole Velocity – ograniczona do wartości z przedziału [-1.0,1.0], który dzielony jest na n_s[3] części.

Cart Pole

W praktyce dyskrtetyzację przestrzeni stanów przeprowadzamy następująco:

```
# położenie, prędkość, kąt, prędkość kątowa
n_s = np.array([10,10,10,10])

#tablica zwierająca granice przedziałów
s_bounds =
np.array(list(zip(env.observation_space.low,env.ob
servation_space.high)))
s_bounds[1] = (-1.0,1.0)
s_bounds[3] = (-1.0,1.0)

#konieczna konwersja typu
s_bounds = np.dtype('float64').type(s_bounds)
```

Cart Pole

Tablica **s_bounds**:

```
[[-4.80000019  4.80000019]   ← position
 [-1.          1.        ]   ← velocity
 [-0.41887903  0.41887903]   ← angle
 [-1.          1.        ]]  ← angular velocity
```

Dyskretyzacja "w akcji":

```
for _ in range(1000):
    env.render()
    action = env.action_space.sample()
    obs, reward, done, info = env.step(action)
    state_new = discretize_state(obs,s_bounds,n_s)
    print(state_new)
    if done == True:
        break
```

Cart Pole

Po uruchomieniu (CartPole_random_action_discretization.py) otrzymujemy kolejne stany układu:

```
[4 4 5 6]
[4 5 5 5]
[4 6 5 4]
[4 5 5 5]
[4 6 5 4]
[4 6 5 2]
[4 7 5 1]
[5 8 5 0]
[5 9 4 0]
[5 9 4 0]
[5 9 4 0]
[5 9 3 0]
[5 9 3 0]
[5 9 3 0]
[5 9 2 0]
[5 8 2 0]
```

Cart Pole

Przy powyższej dyskretyzacji polityka π przypisująca każdej akcji (w lewo, w prawo) prawdopodobieństwo 0.25 może być zdefiniowana jako tensor (macierz) o wymiarach n_s[0]×n_s[1]×n_s[2]×n_s[3]×2 zawierająca tylko wartości 0.5:

```
import numpy as np
from cartpole import CartPoleEnv

env = CartPoleEnv()

n_s = np.array([10,10,10,10])

stochastic_policy =
        np.ones([n_s[0],n_s[1],n_s[2],n_s[3],2])/2
...
```

liczba akcji (2)

stan

Koniec części 4