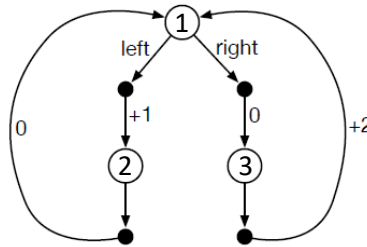


## Wprowadzenie do RL 4

### Zadanie 1

Rozważmy ciągły MDP przedstawiony na poniższym rysunku:



Rozważmy dwie deterministyczne polityki  $\pi_{\text{RIGHT}}$  i  $\pi_{\text{LEFT}}$  określające wybór w stanie 1. Jaka polityka jest optymalna dla:  $\gamma = 0$ ,  $\gamma = 0.9$ ,  $\gamma = 0.5$  ?

### Zadanie 2

Napisz program w którym agent w środowisku *Frozen Lake* przejdzie od stanu 0 do stanu 15. Wybierz odpowiednią ścieżkę i dostosuj akcje agenta. Wykorzystaj skrypt `FrozenLake_2.py`.

### Zadanie 3

Napisz program implementujący poniższy **algorytm iteracyjnego obliczenia polityki** dla środowiska *Frozen Lake* w celu znalezienia wartości  $V(s)$ .

#### Iterative Policy Evaluation, for estimating $V \approx v_\pi$

Input  $\pi$ , the policy to be evaluated

Algorithm parameter: a small threshold  $\theta > 0$  determining accuracy of estimation

Initialize  $V(s)$ , for all  $s \in \mathcal{S}^+$ , arbitrarily except that  $V(\text{terminal}) = 0$

Loop:

$\Delta \leftarrow 0$

Loop for each  $s \in \mathcal{S}$ :

$v \leftarrow V(s)$

$V(s) \leftarrow \sum_a \pi(a|s) \sum_{s', r} p(s', r | s, a) [r + \gamma V(s')]$

$\Delta \leftarrow \max(\Delta, |v - V(s)|)$

until  $\Delta < \theta$

Przyjmij, że **polityka stochastyczna**  $\pi$  określa jednakowe prawdopodobieństwo dla każdej akcji:  $\forall a \pi(a|s) = 1/4$ . Przetestuj program dla różnych wartości parametru  $\gamma$  np. 1, 0.9, 0.5. Wykorzystaj skrypt `FrozenLake_3.py`.