# Markov Decision Process

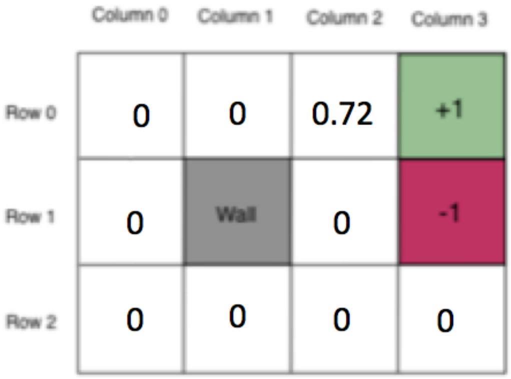
- Given the policy the utility of any state can be calculated based on:

  $$V(s) = R(s) + \max_{a\in A} * \gamma \sum_{s^{'}} P(s^{'}|s,a) * V(s^{'})$$

  Given $\gamma = 0.9, R(s) = 0, P(up) = 0.8, P(right) = P(left) = 0.1$
  After iteration 1, the policy of the agent is as given in the question.:

  - Value of cell (0,2) = 0+0.1*0.9*0+0.8*0.9*1+0.1*0.9*0=0.72
  - Value of cell (1,2) = 0 as P(down) = 0 and the neighboring utilities are 0 as well.
  - Value of all other cells would be 0 as the neighboring and local utilities are 0.

  The grid with values is as shown below:



| | Column 0 | Column 1 | Column 2 | Column 3 |
|---|---|---|---|---|
| Row 0 | 0 | 0 | 0.72 | +1 |
| Row 1 | 0 | Wall | 0 | -1 |
| Row 2 | 0 | 0 | 0 | 0 |

Figure 1: Values after 1st iteration

- Given these values a policy can be found based on:

  $$\arg\max_{s\in A} \sum_{s^{'}} P(s^{'}|s,a) * V(s^{'})$$

  The policies for each grid are calculated as follows:

  - 0,2 :
    - $\uparrow$ = 0.9*(0.8*0.72 + 0.1*1 + 0.1*0) = 0.6084
    - $\rightarrow$ = 0.9*(0.8*1 + 0.1*0.72 + 0.1*0) = 0.7848
    - $\leftarrow$ = 0.9*(0.8*0 + 0.1*0.72 + 0.1*0) = 0.0648
    - $\downarrow$ = 0.9*(0.8*0 + 0.1*1 + 0.1*0) = 0.09

    Hence the policy for this case would be $\rightarrow$ with a value of 0.7848.
  - 1,2 :
    - $\uparrow$ = 0.9*(0.8*0.72 + 0.1*-1 + 0.1*0) = 0.4284

    - $\rightarrow = 0.9*(0.8*-1 + 0.1*0.72 + 0.1*0) = -0.6552$
    - $\leftarrow = 0.9*(0.8*0 + 0.1*0.72 + 0.1*0) = 0.0648$
    - $\downarrow = 0.9*(0.8*0 + 0.1*-1 + 0.1*0) = -0.09$

  Hence the policy for this case would be ↑ with a value of 0.4284.

- 0,1 :
    - $\uparrow = 0.9*(0.8*0 + 0.1*0.72 + 0.1*0) = 0.0648$
    - $\rightarrow = 0.9*(0.8*0.72 + 0.1*0 + 0.1*0) = 0.5184$
    - $\leftarrow = 0.9*(0.8*0 + 0.1*0 + 0.1*0) = 0$
    - $\downarrow = 0.9*(0.8*0 + 0.1*0.72 + 0.1*0) = 0.0648$

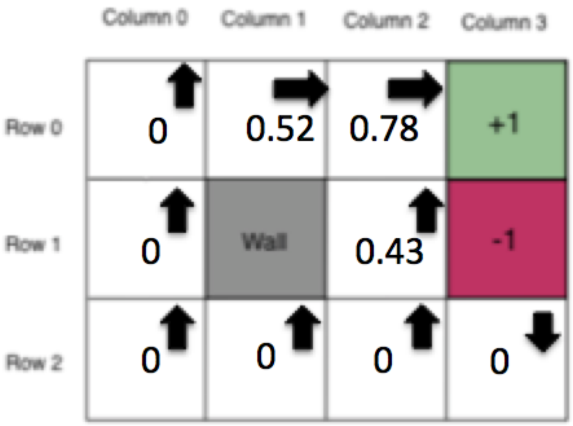
  Hence the policy for this case would be → with a value of 0.5184.

- 2,3 :
    - $\uparrow = 0.9*(0.8*-1 + 0.1*0 + 0.1*0) = -0.72$
    - $\rightarrow = 0.9*(0.8*0 + 0.1*-1 + 0.1*0) = -0.09$
    - $\leftarrow = 0.9*(0.8*0 + 0.1*-1 + 0.1*0) = -0.09$
    - $\downarrow = 0.9*(0.8*0 + 0.1*0 + 0.1*0) = 0$

  Hence the policy for this case would be ↓ with a value of 0.

- The rest of the cells have a policy ↑ as all the states would have values to be 0 but ↑ has a higher preference.

The grid with the policy and the values is as shown below:



| | Column 0 | Column 1 | Column 2 | Column 3 |
|---|---|---|---|---|
| Row 0 | ↑ 0 | → 0.52 | → 0.78 | +1 |
| Row 1 | ↑ 0 | Wall | ↑ 0.43 | -1 |
| Row 2 | ↑ 0 | ↑ 0 | ↑ 0 | ↓ 0 |

Figure 2: Policy and Values after 2nd iteration