Reinforcement Learning

MACHINE LEARNING

Pakarat Musikawan

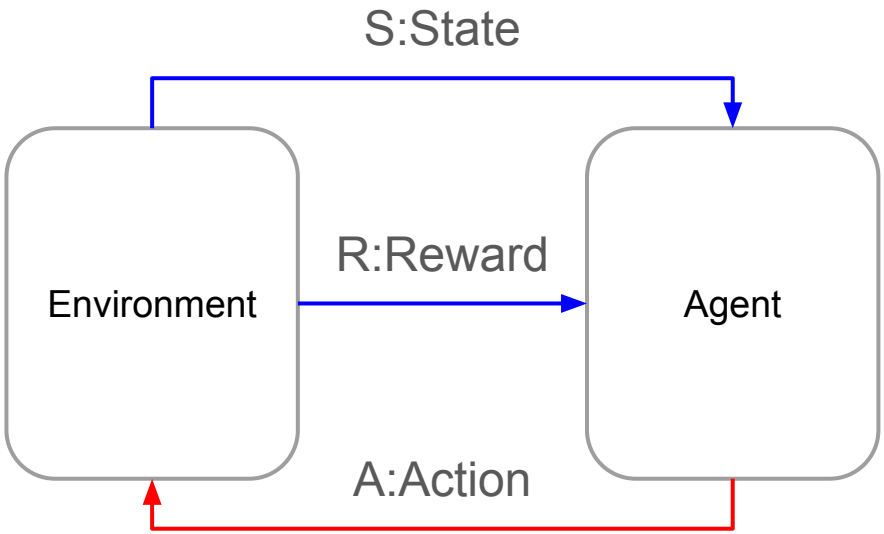
Introduction to Reinforcement Learning

Reinforcement Learning (RL) is a type of machine learning where an agent learns by interacting with an environment to maximize cumulative rewards.

Key Concepts:

- **Agent**: Learner/decision maker.
- **Environment**: What the agent interacts with.
- **State**: The current situation returned by the environment.
- **Action**: What the agent can do.
- **Reward**: Feedback from the environment.
- **Policy**: Strategy used by the agent.

Introduction to Reinforcement Learning



$$S_0 \xrightarrow[R_0]{A_0} S_1 \xrightarrow[R_1]{A_1} S_2 \xrightarrow[R_2]{A_2} \dots$$

$$G = R_0 + \gamma R_1 + \gamma^2 R_2 + \dots \gamma^T R_T$$

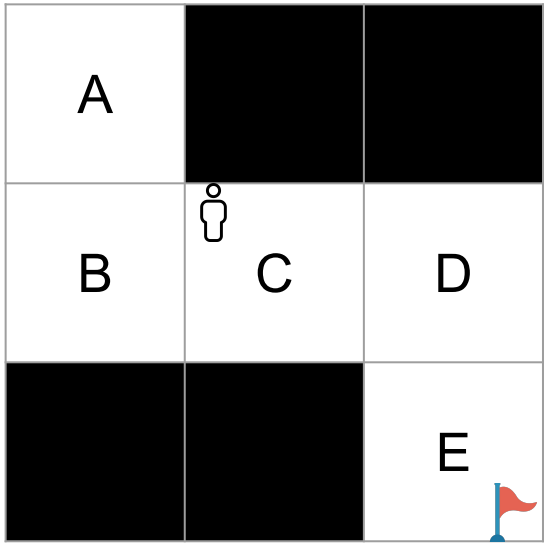
$$= \sum_{i=0}^{T} \gamma^i R_i$$

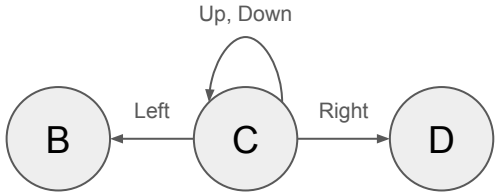
$$0 \leq \gamma \leq 1$$

Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

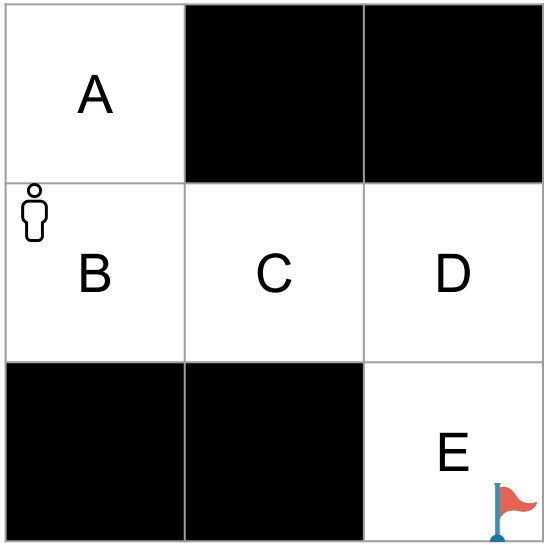


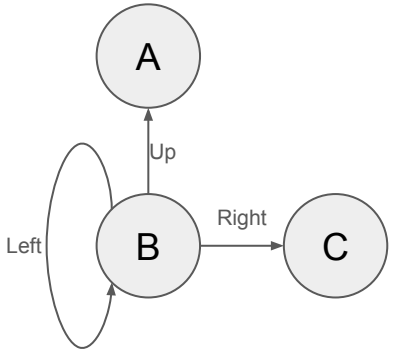


Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

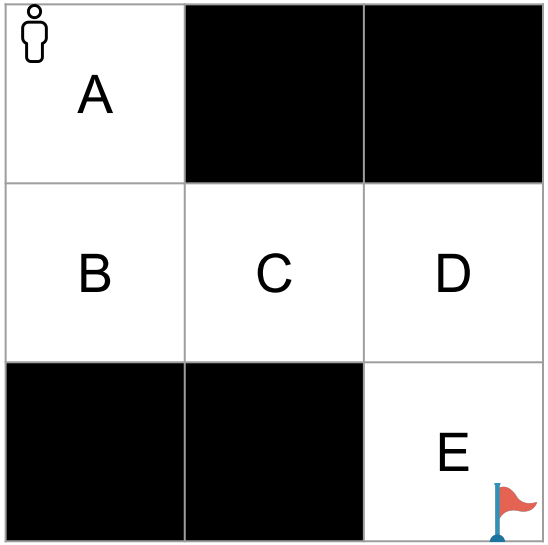


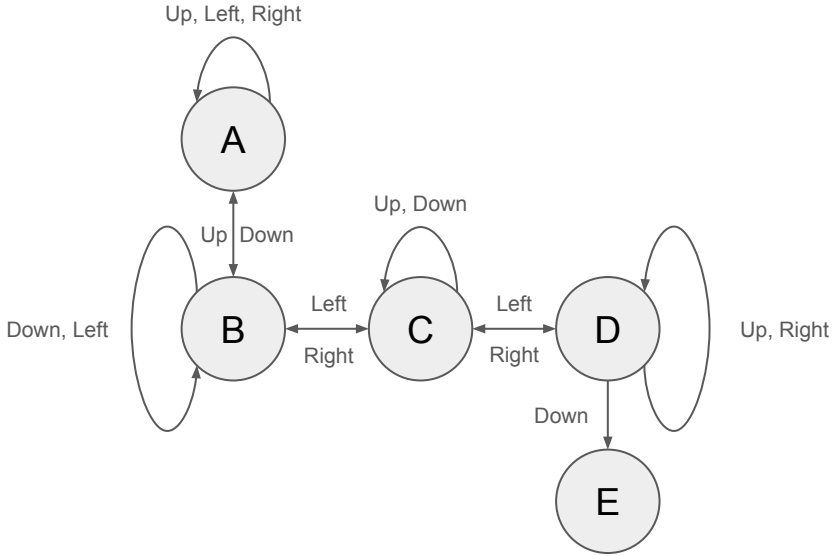


Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

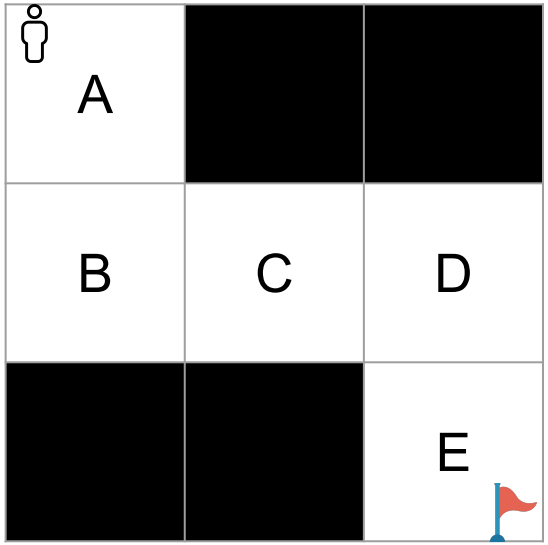


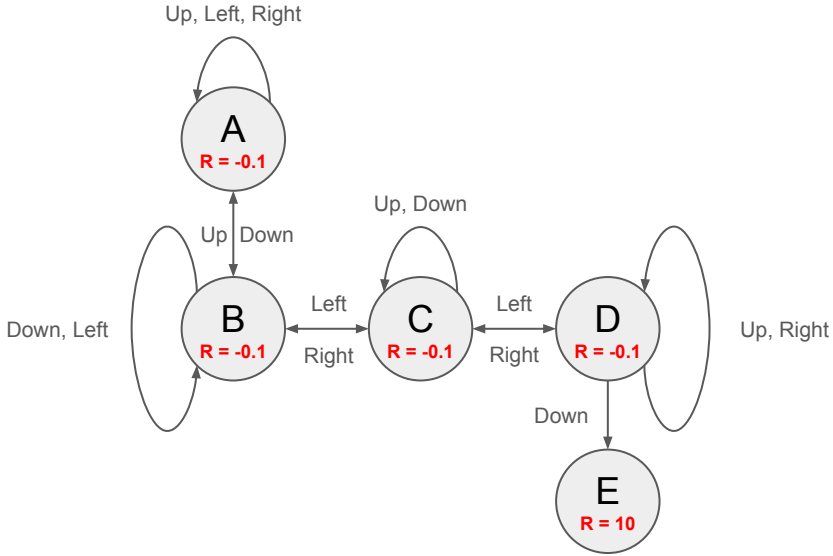


Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

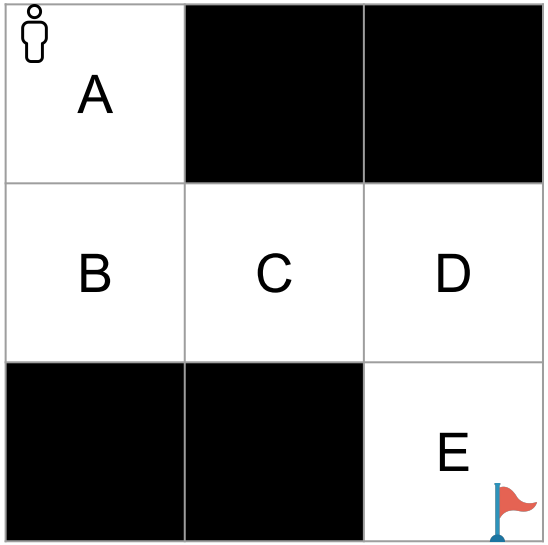


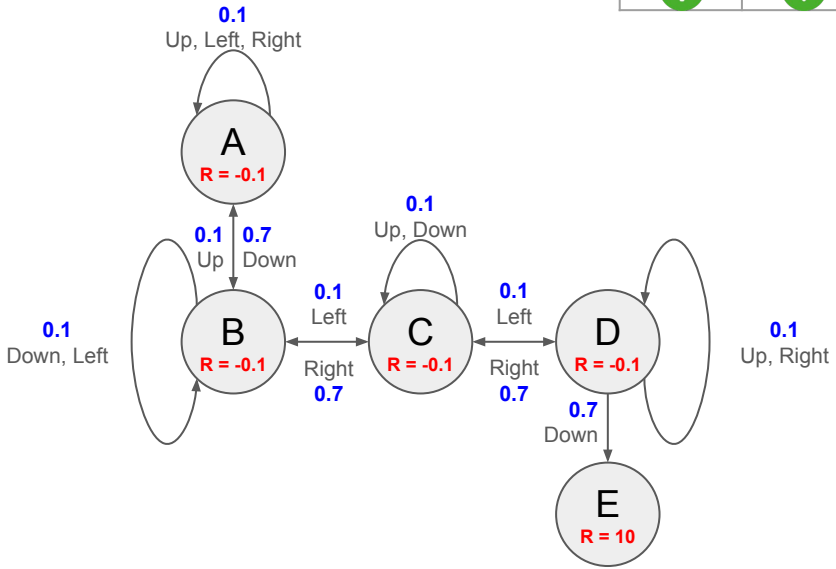


Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

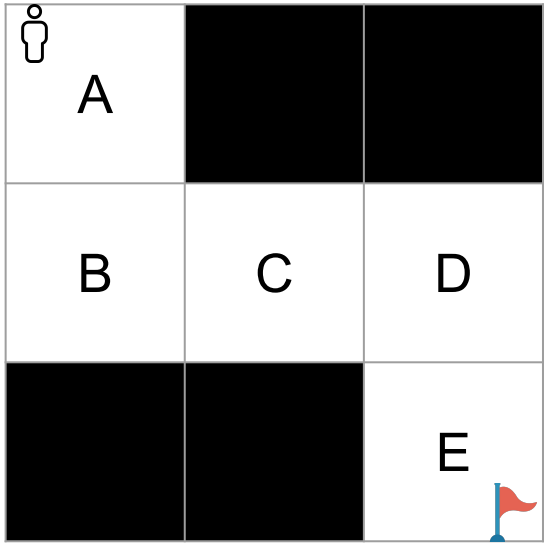




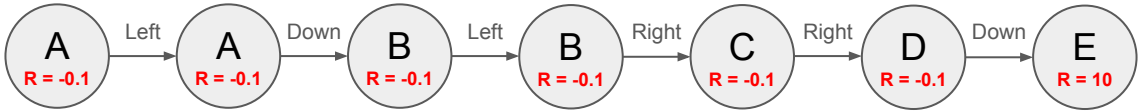
Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

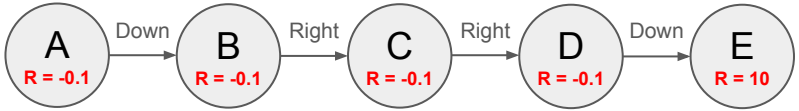
Markov decision process



$\gamma = 0.9$



$$= (-0.1 \times 0.9^0) + (-0.1 \times 0.9^1) + (-0.1 \times 0.9^2) + (-0.1 \times 0.9^3) + (-0.1 \times 0.9^4) + (-0.1 \times 0.9^5) + (10 \times 0.9^6)$$
$$\approx 4.845$$

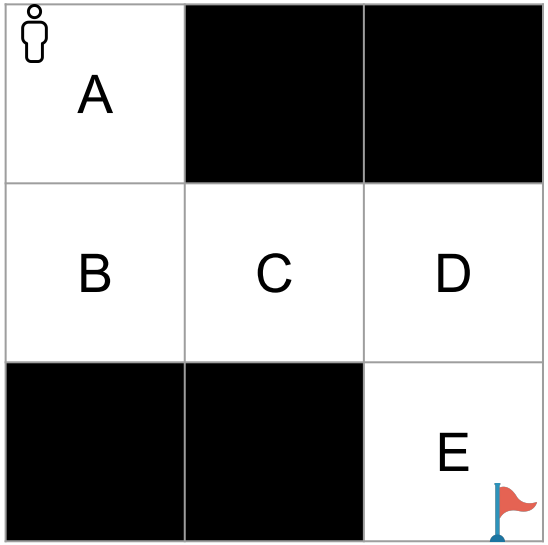


$$= (-0.1 \times 0.9^0) + (-0.1 \times 0.9^1) + (-0.1 \times 0.9^2) + (-0.1 \times 0.9^3) + (10 \times 0.9^4)$$
$$\approx 6.217$$

Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process

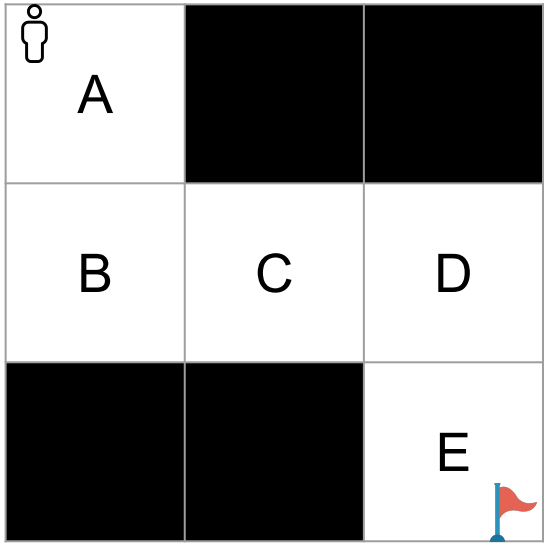


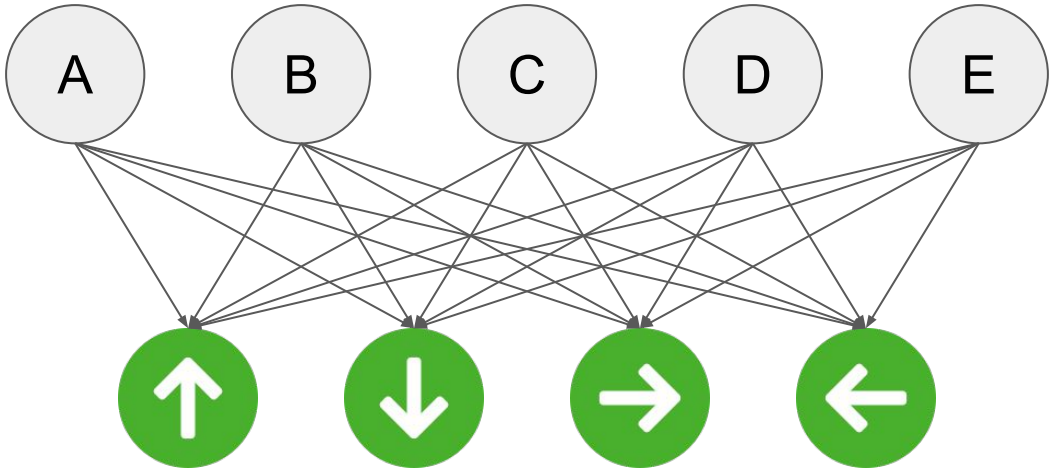
Q-Table	Action			
State	Up	Down	Right	Left
A	0.1	0.7	0.1	0.1
B	0.1	0.1	0.7	0.1
C	0.1	0.1	0.7	0.1
D	0.1	0.7	0.1	0.1
E				

Example of an RL Problem: Grid-World

Action			
Up	Down	Left	Right

Markov decision process



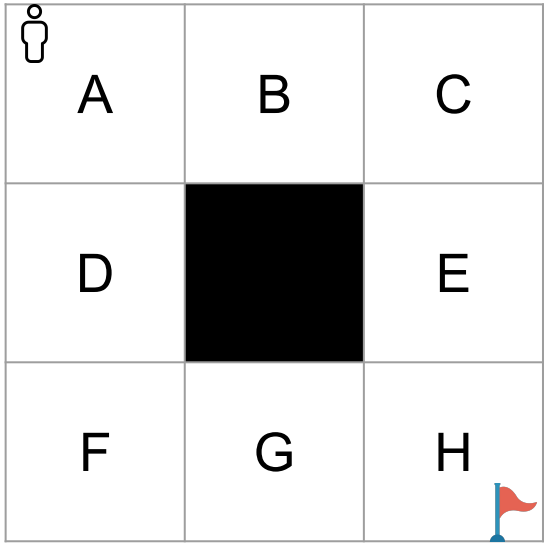


Q-Learning

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_t + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t) \right]$$

$Q(S_t, A_t)$	The current Q-value for the agent being in state **S_t** and and taking action **A_t**
α	Learning rate
γ	Discount factor
R_t	Immediate reward
$\max_a Q(S_{t+1}, a)$	The maximum Q-value over all possible actions ***a*** a the agent can take in the next state **S_{t+1}**

Q-Learning



Q-Table	Action			
State	Up	Down	Left	Right
A	-0.1	0.7	-0.1	0.9
B	-0.1	-0.1	0.1	0.9
C	-0.1	0.9	0.1	-0.1
D	0.1	0.8	-0.1	-0.1
E	0.1	0.9	-0.1	-0.1
F	0.1	-0.1	-0.1	0.9
G	-0.1	-0.1	0.1	0.9
H	1.0	1.0	1.0	1.0

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	-0.4	0.5
D	-0.6	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State C and chooses Action 2
- The agent receives a reward equal to 2
- The agent transitions to State D

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	-0.4	0.5
D	-0.6	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State C and chooses Action 2
- The agent receives a reward equal to 2
- The agent transitions to State D

$$Q(C,2) = -0.4$$
$$R = 2$$
$$\gamma = 0.9$$
$$\max Q(D,a) = \max(-0.6, 0.8, 0.0) = 0.8$$
$$\alpha = 0.1$$

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	**-0.088**	0.5
D	-0.6	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State C and chooses Action 2
- The agent receives a reward equal to 2
- The agent transitions to State D

$$Q(C,2) = -0.4$$
$$R = 2$$
$$\gamma = 0.9$$
$$\max Q(D,a) = \max(-0.6, 0.8, 0.0) = 0.8$$
$$\alpha = 0.1$$

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_t + \gamma \max Q(S_{t+1}, a) - Q(S_t, A_t)\right]$$

$$\begin{aligned} Q(C,2) &\leftarrow Q(C,2) + 0.1\left[2 + 0.9 \max(-0.6, 0.8, 0.0) - Q(C,2)\right] \\ &\leftarrow (-0.4) + 0.1\left[2 + 0.9 \times 0.8 - (-0.4)\right] \\ &\leftarrow (-0.4) + 0.1 \times 3.12 \\ &\leftarrow (-0.4) + 0.312 = -0.088 \end{aligned}$$

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	-0.088	0.5
D	-0.6	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State D and chooses Action 1
- The agent receives a reward equal to -1
- The agent transitions to State E

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	-0.088	0.5
D	-0.6	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State D and chooses Action 1
- The agent receives a reward equal to -1
- The agent transitions to State E

$$Q(D,1) = -0.6$$
$$R = -1$$
$$\gamma = 0.9$$
$$\max Q(E,a) = \max(0.2, -0.1, 0.4) = 0.4$$
$$\alpha = 0.1$$

Q-Learning – Example

Q-Table	Action		
State	1	2	3
A	0.5	-0.2	0.1
B	0.0	1.0	-0.3
C	0.7	-0.088	0.5
D	**-0.604**	0.8	0.0
E	0.2	-0.1	0.4

- The agent starts in State D and chooses Action 1
- The agent receives a reward equal to -1
- The agent transitions to State E

$$Q(D,1) = -0.6$$
$$R = -1$$
$$\gamma = 0.9$$
$$\max Q(E,a) = \max(0.2, -0.1, 0.4) = 0.4$$
$$\alpha = 0.1$$

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_t + \gamma \max Q(S_{t+1}, a) - Q(S_t, A_t)\right]$$

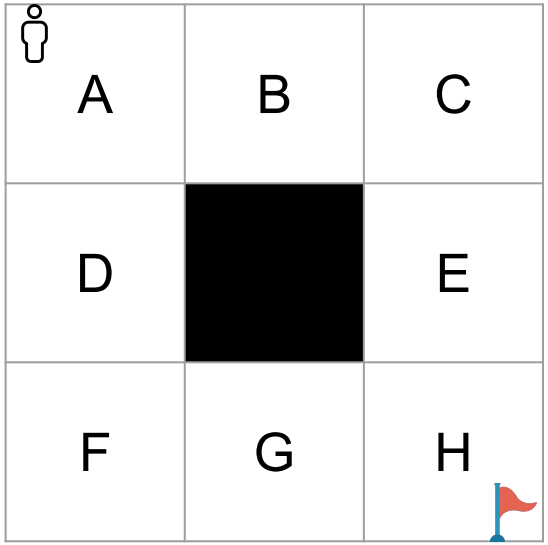
$$\begin{aligned} Q(D,1) &\leftarrow Q(D,1) + 0.1\left[(-1) + 0.9\max(0.2, -0.1, 0.4) - Q(D,1)\right] \\ &\leftarrow (-0.6) + 0.1\left[(-1) + 0.9 \times 0.4 - (-0.6)\right] \\ &\leftarrow (-0.6) + 0.1 \times (-0.04) \\ &\leftarrow (-0.6) + (-0.004) = -0.604 \end{aligned}$$

Hand On

$$\alpha = 0.1$$

$$\gamma = 0.9$$

$$R = \begin{cases} \text{Positive} & 1 \\ \text{Negative} & -1 \end{cases}$$



$$Q(A,1) \rightarrow Q(A,4) \rightarrow Q(B,2) \rightarrow Q(B,4)$$

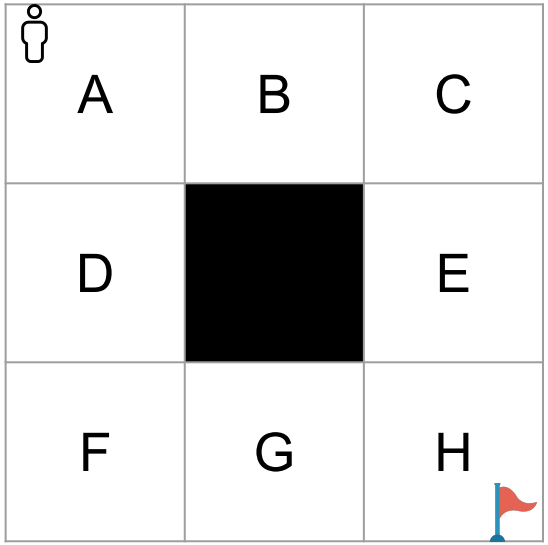
Q-Table	Action			
State	Up	Down	Left	Right
A	-0.1	0.7	-0.1	0.9
B	-0.1	-0.1	0.1	0.9
C	-0.1	0.9	0.1	-0.1
D	0.1	0.8	-0.1	-0.1
E	0.1	0.9	-0.1	-0.1
F	0.1	-0.1	-0.1	0.9
G	-0.1	-0.1	0.1	0.9
H	1.0	1.0	1.0	1.0

Hand On

$$\alpha = 0.1$$

$$\gamma = 0.9$$

$$R = \begin{cases} \text{Positive} & 1 \\ \text{Negative} & -1 \end{cases}$$



$$Q(A,4) \rightarrow Q(B,4) \rightarrow Q(C,4) \rightarrow Q(C,2) \rightarrow Q(E,2)$$

Q-Table	Action			
State	Up	Down	Left	Right
A	-0.1	0.7	-0.1	0.9
B	-0.1	-0.1	0.1	0.9
C	-0.1	0.9	0.1	-0.1
D	0.1	0.8	-0.1	-0.1
E	0.1	0.9	-0.1	-0.1
F	0.1	-0.1	-0.1	0.9
G	-0.1	-0.1	0.1	0.9
H	1.0	1.0	1.0	1.0