Aprendizagem por Reforço com Deep Learning, PyTorch e Python

© SuperDataScience & IA Expert

Conteúdo do Curso

O que você aprenderá neste curso:

- Parte 1 – Fundamentos de Aprendizagem por Reforço (Q-Learning)
- Parte 2 – Deep Q-Learning
- Parte 3 – Implementação Deep Q-Learning

© SuperDataScience & IA Expert

Pré-requisitos

- Lógica de programação
- Programação básica em Python
- Instalação de softwares
- Redes Neurais Artificiais (desejável)
- PyTorch (desejável)

- Nível do curso: intermediário
- Dica: aumentar a velocidade do player
- Avaliação do curso!

© SuperDataScience & IA Expert

Conteúdo

© SuperDataScience & IA Expert

Conteúdo

O que você aprenderá nesta seção:

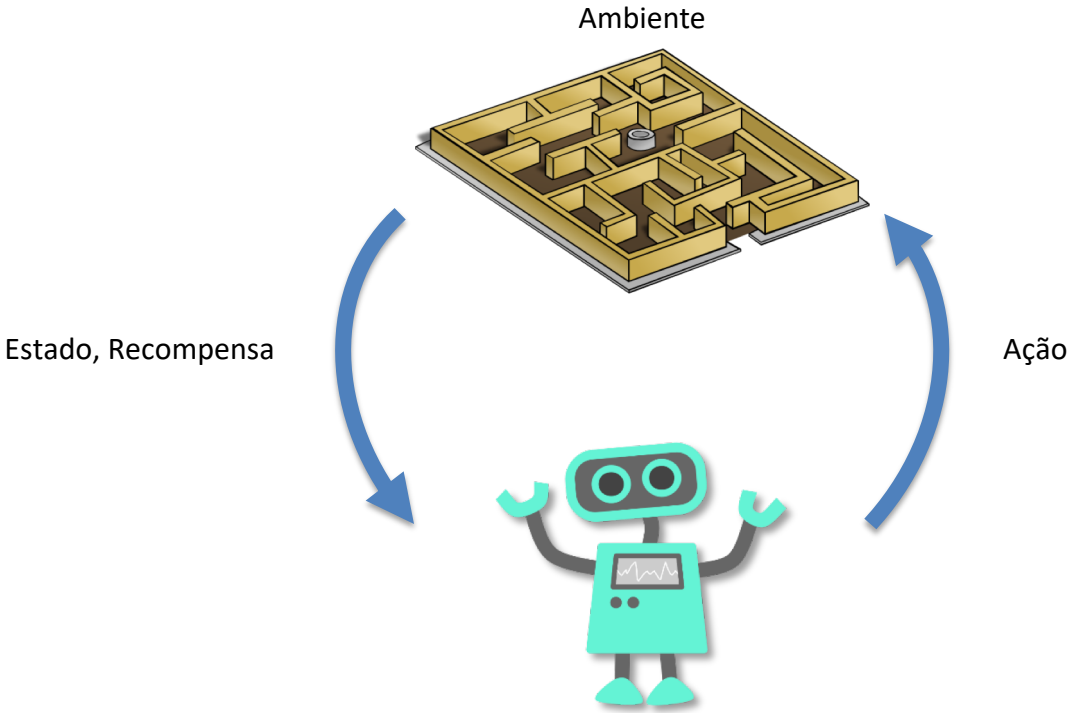
- O que é aprendizagem por reforço?
- A Equação de Bellman
- O "Plano"
- Markov Decision Process (MDP)
- Política x Plano
- Adição de penalidades ("Living Penalty”)
- Q-Learning – Intuição
- Diferença Temporal
- Q-Learning – Visualização

© SuperDataScience & IA Expert

O que é Aprendizagem por Reforço?

© SuperDataScience & IA Expert

O que é Aprendizagem por Reforço?

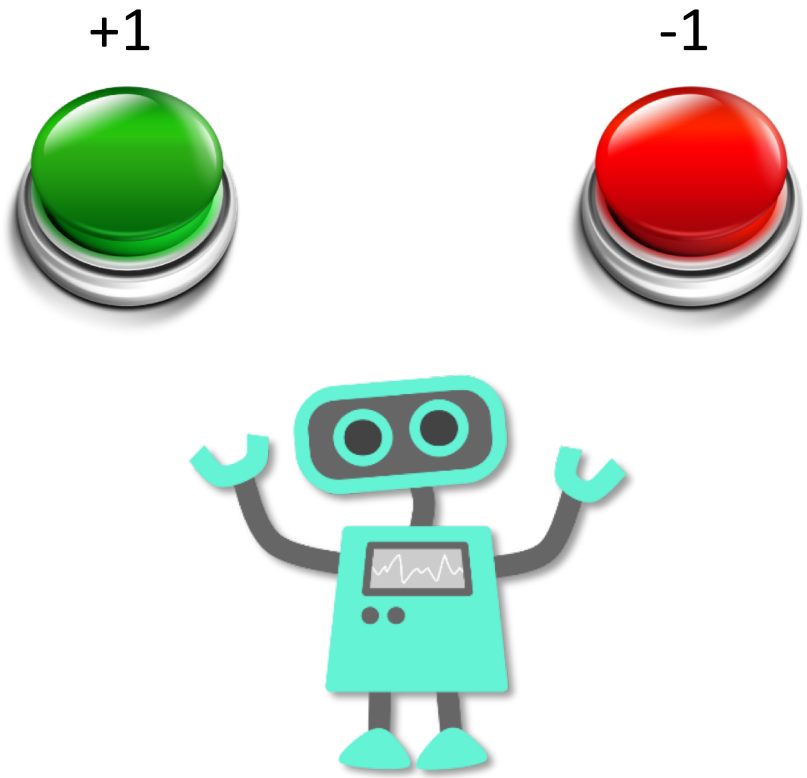


© SuperDataScience & IA Expert

O que é Aprendizagem por Reforço?

© SuperDataScience & IA Expert

O que é Aprendizagem por Reforço?



© SuperDataScience & IA Expert

O que é Aprendizagem por Reforço?

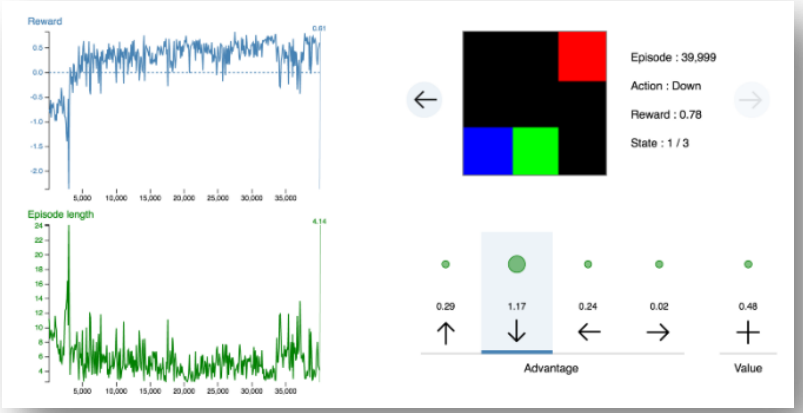
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Simple Reinforcement Learning with Tensorflow (10 Parts)

Arthur Juliani (2016)



Link:

https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-0-q-learning-with-tables-and-neural-networks-d195264329d0

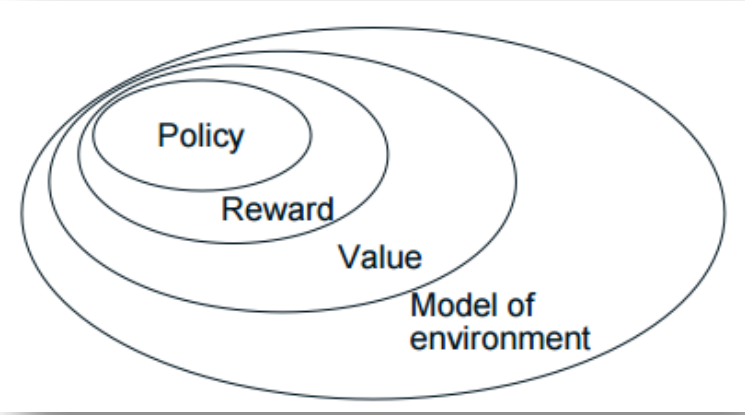
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Reinforcement Learning I: Introduction

Richard Sutton et al. (1998)



Link:

http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.7692

© SuperDataScience & IA Expert

A Equação de Bellman

© SuperDataScience & IA Expert

A Equação de Bellman

Conceitos:

- s – Estado
- a – Ação
- R – Recompensa
- γ – Desconto

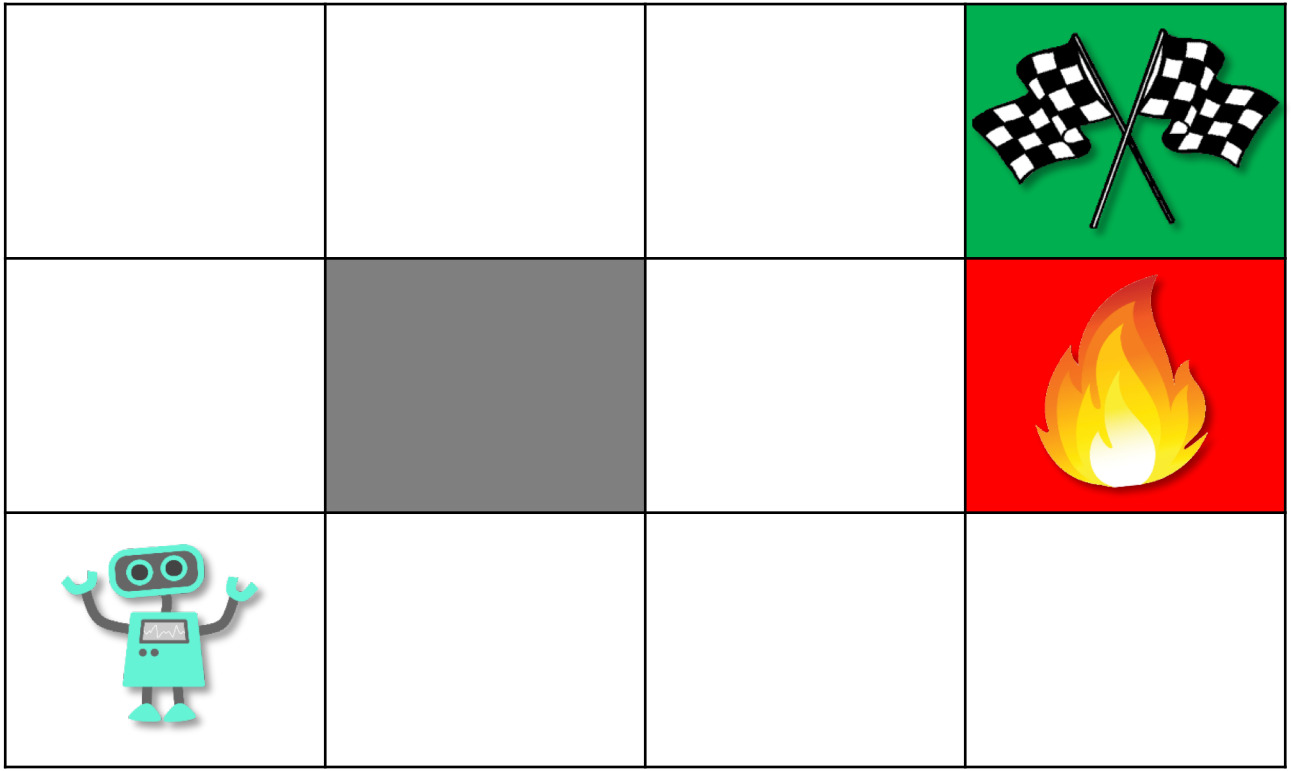
© SuperDataScience & IA Expert

A Equação de Bellman

Richard Ernest Bellman

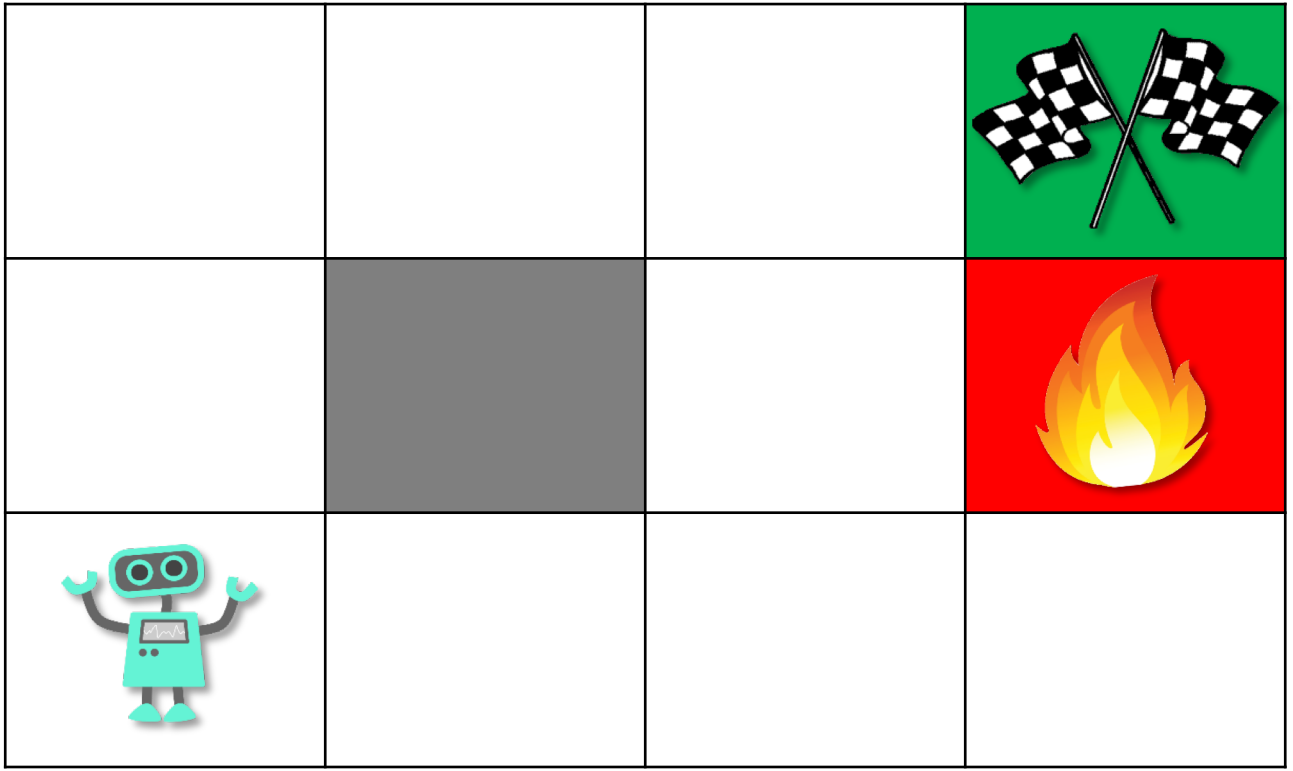
© SuperDataScience & IA Expert

A Equação de Bellman



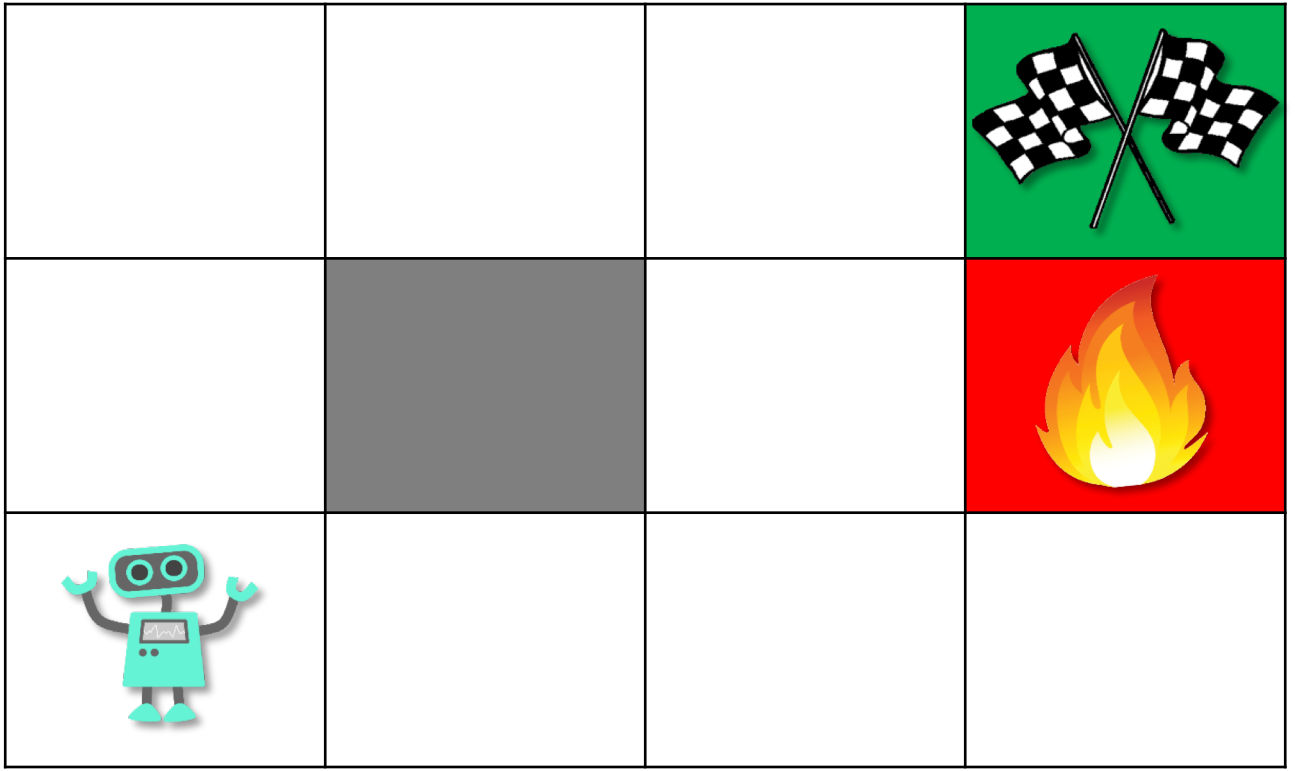
© SuperDataScience & IA Expert

A Equação de Bellman



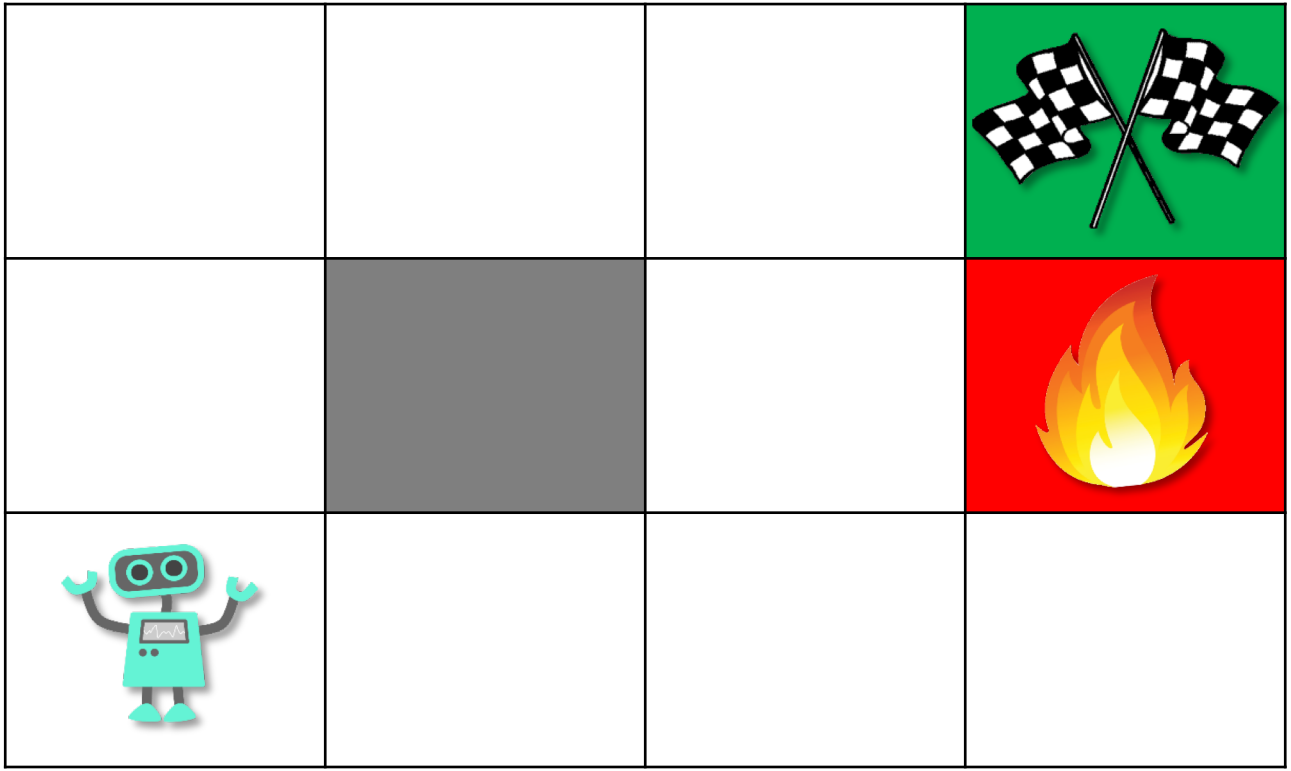
© SuperDataScience & IA Expert

A Equação de Bellman



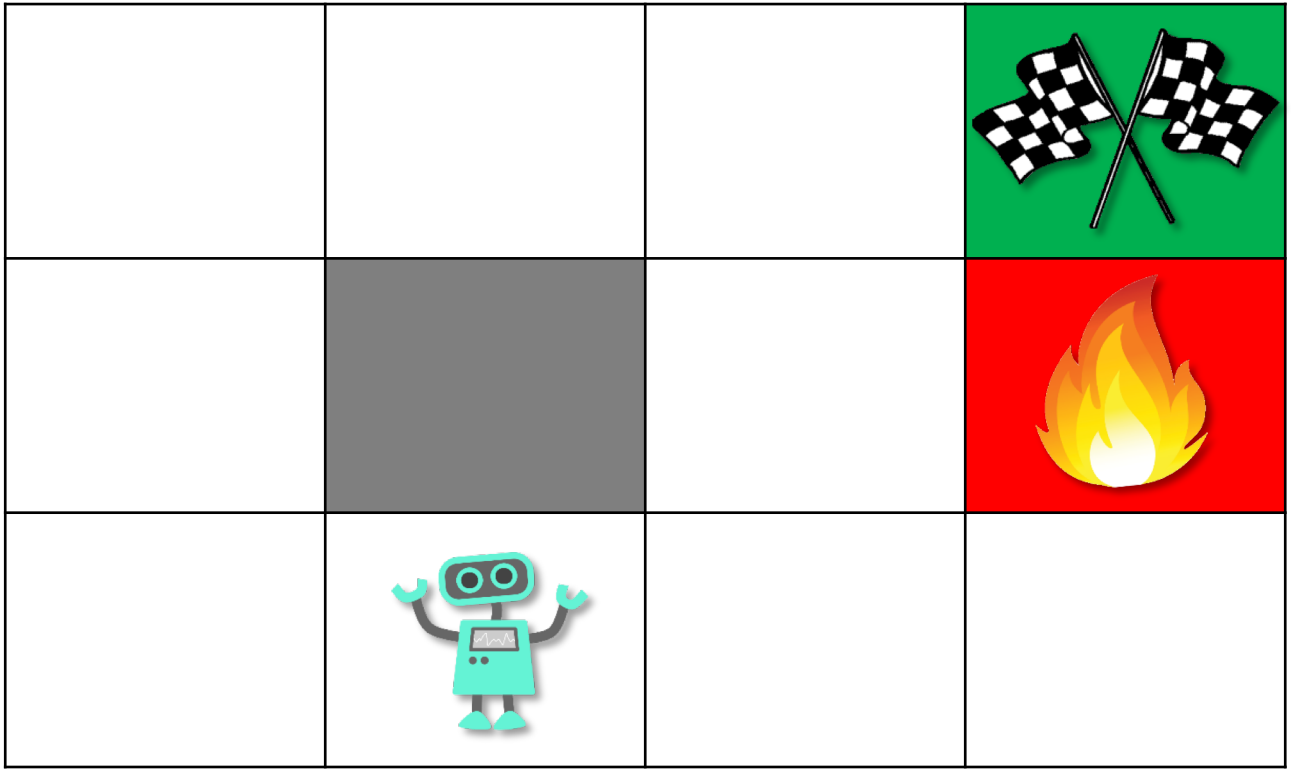
© SuperDataScience & IA Expert

A Equação de Bellman



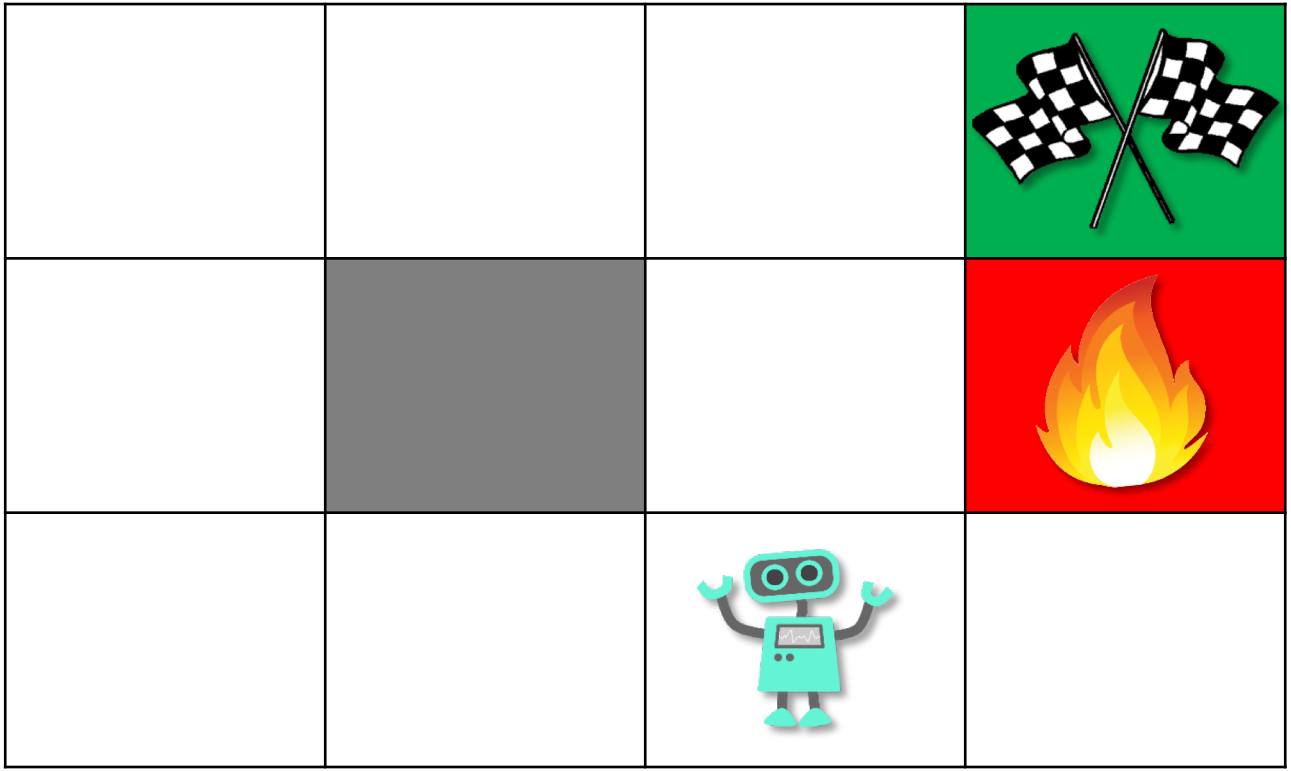
© SuperDataScience & IA Expert

A Equação de Bellman



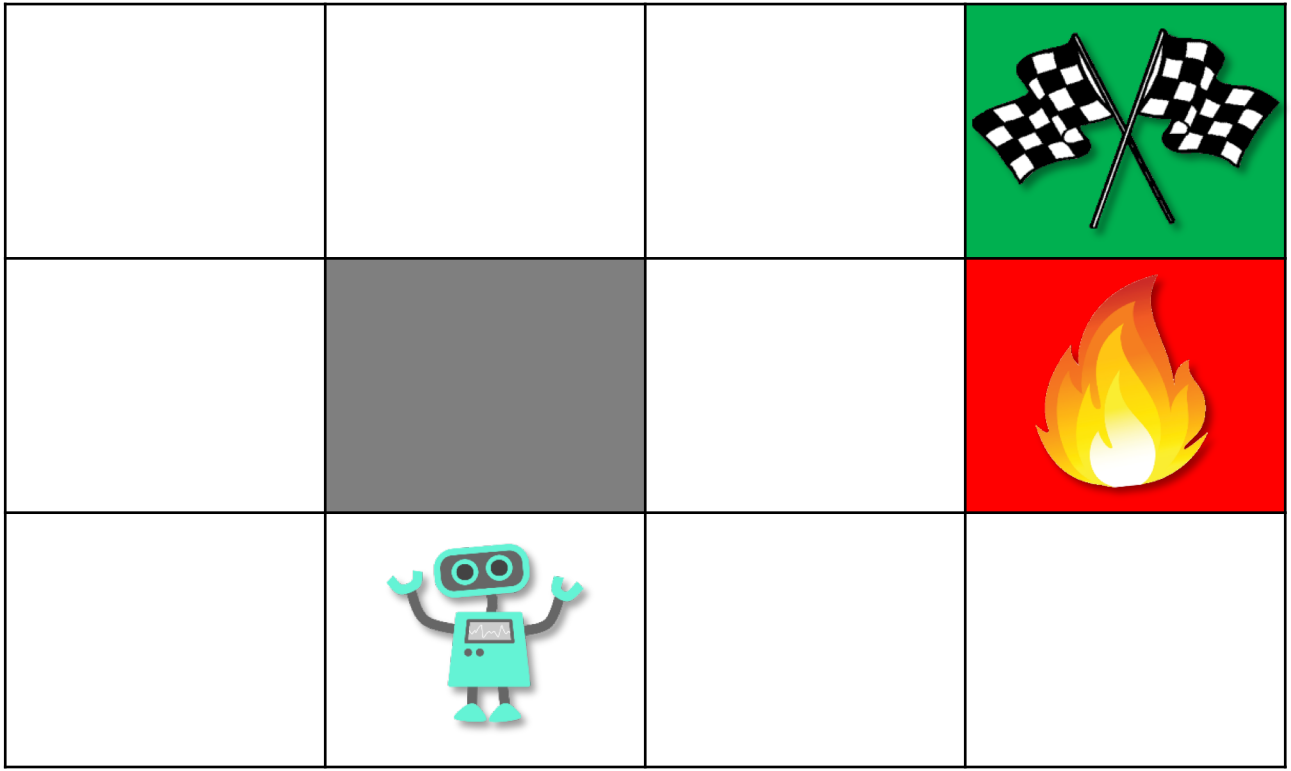
© SuperDataScience & IA Expert

A Equação de Bellman



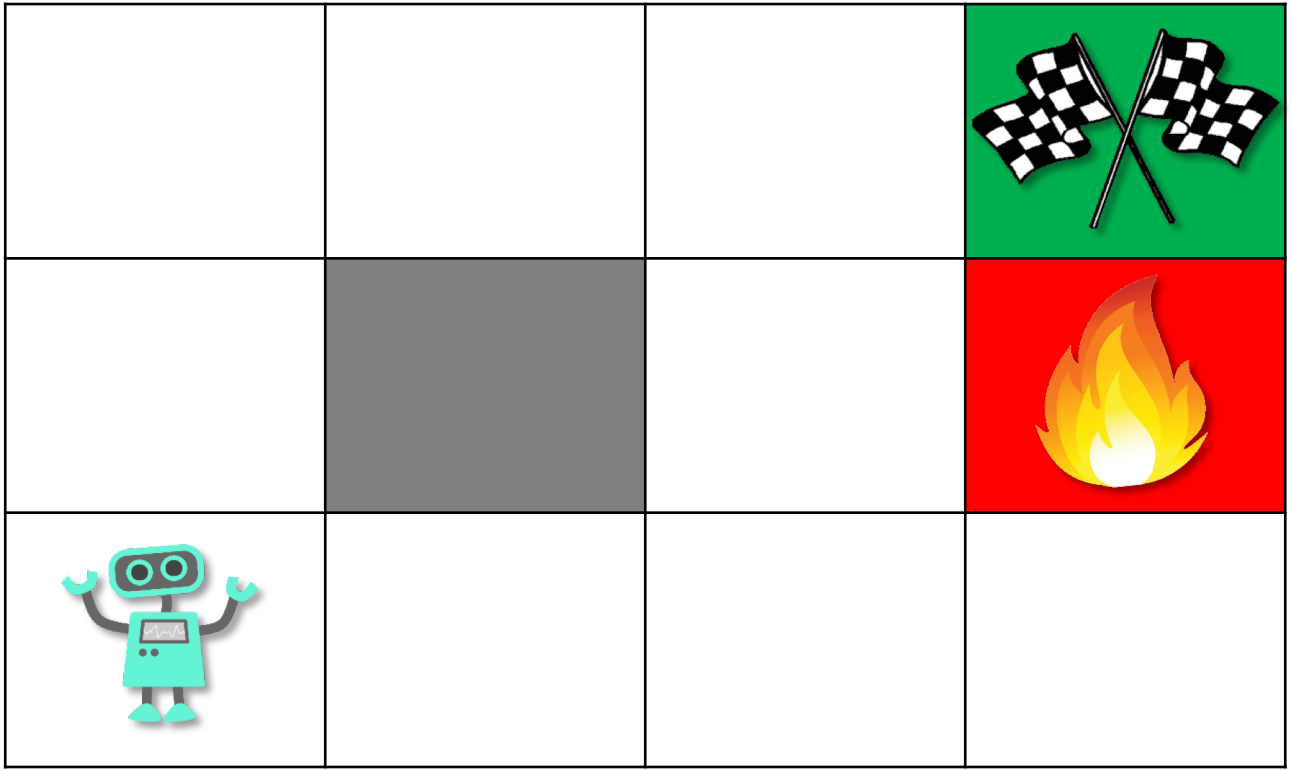
© SuperDataScience & IA Expert

A Equação de Bellman



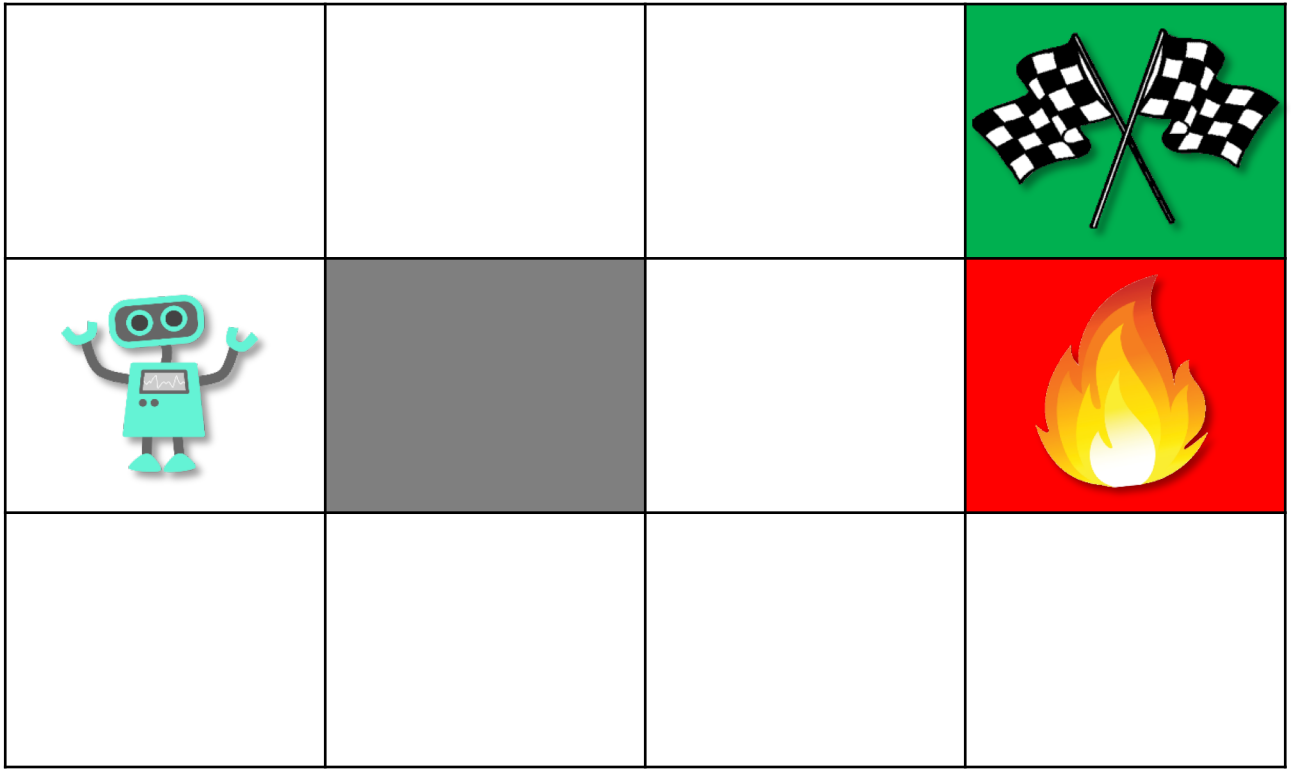
© SuperDataScience & IA Expert

A Equação de Bellman



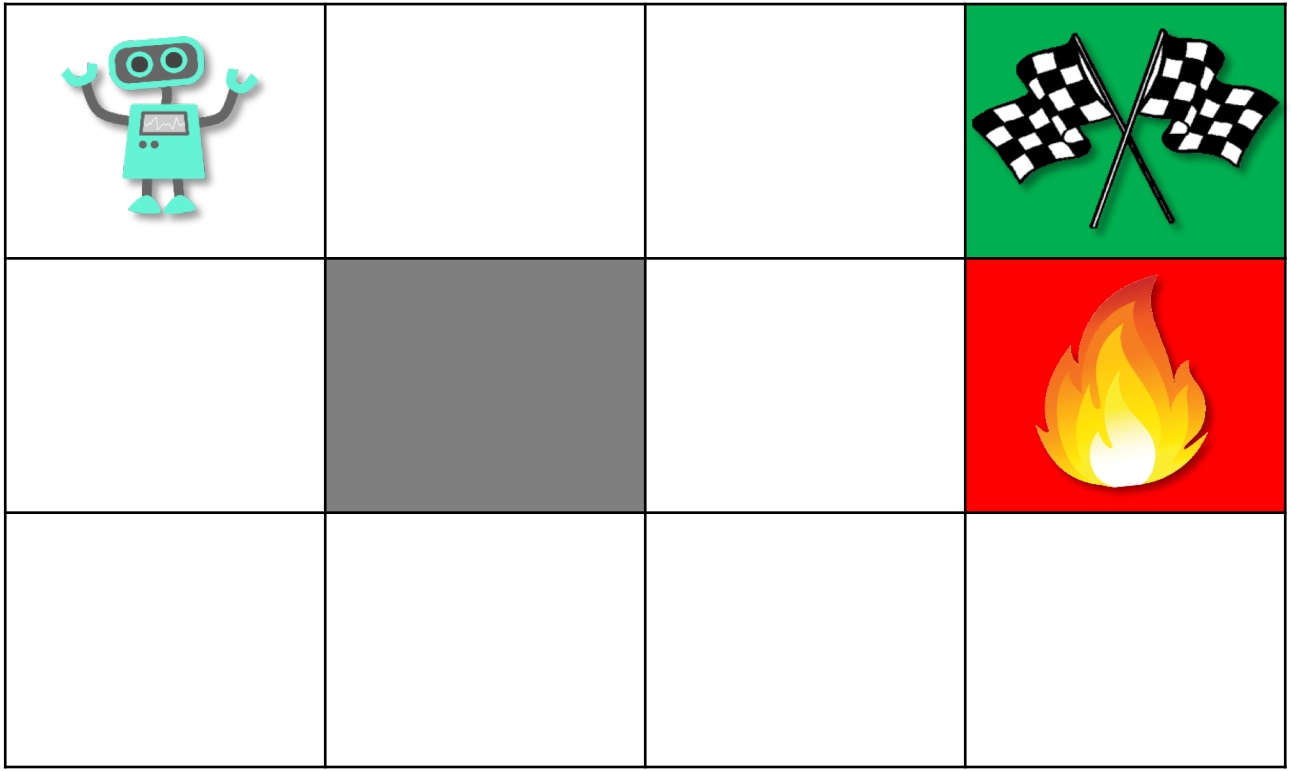
© SuperDataScience & IA Expert

A Equação de Bellman



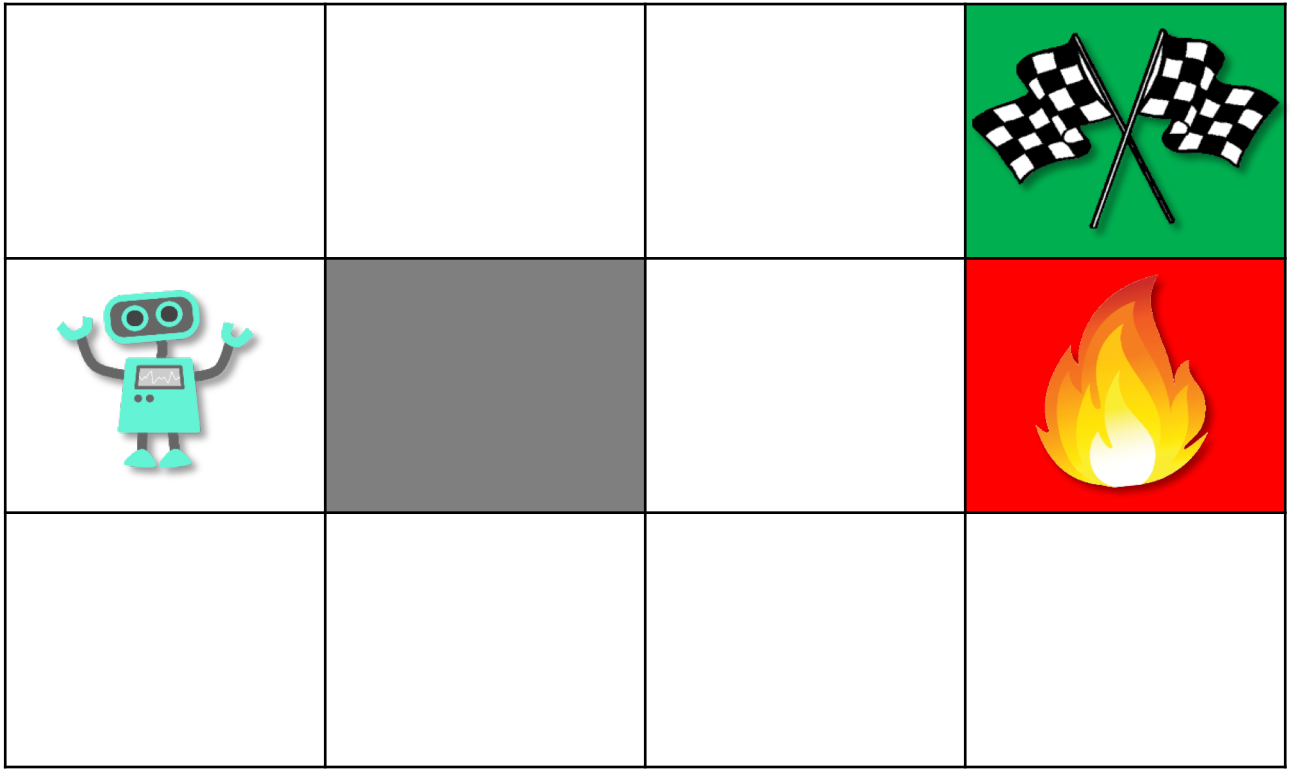
© SuperDataScience & IA Expert

A Equação de Bellman



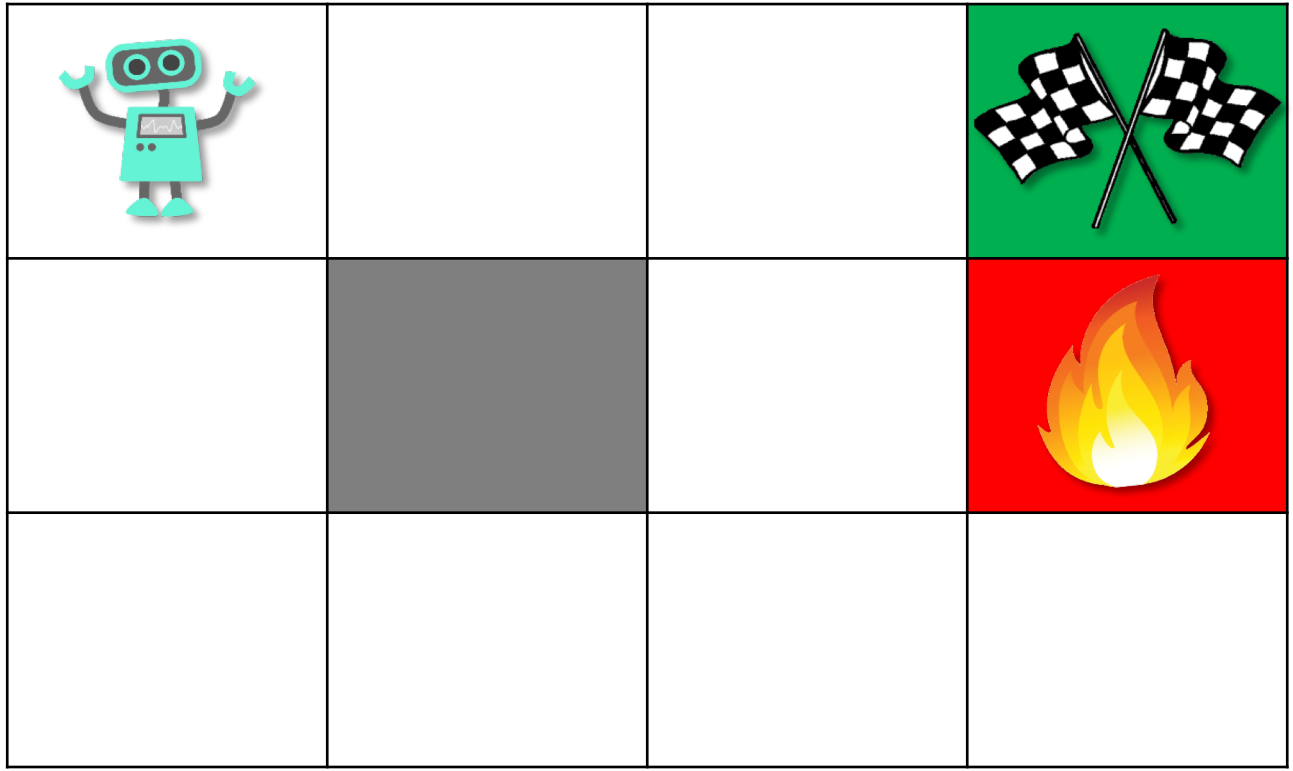
© SuperDataScience & IA Expert

A Equação de Bellman



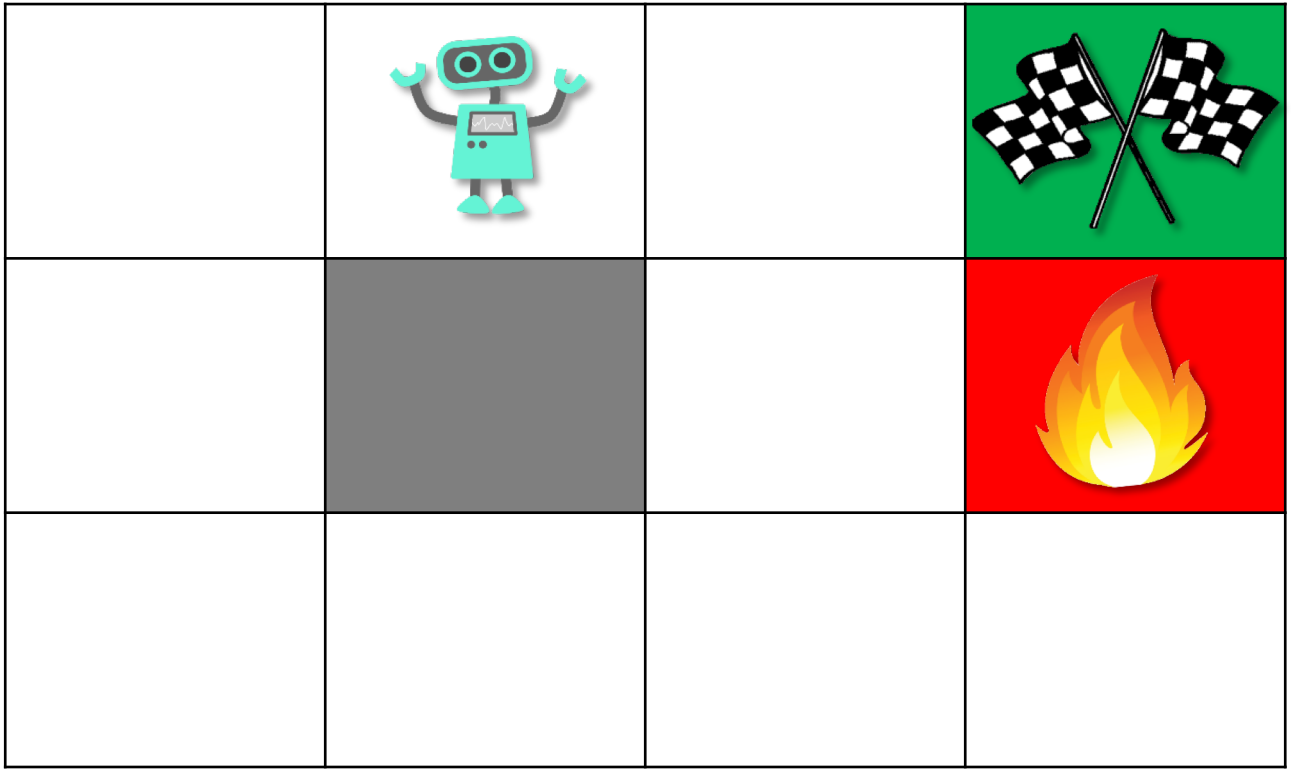
© SuperDataScience & IA Expert

A Equação de Bellman



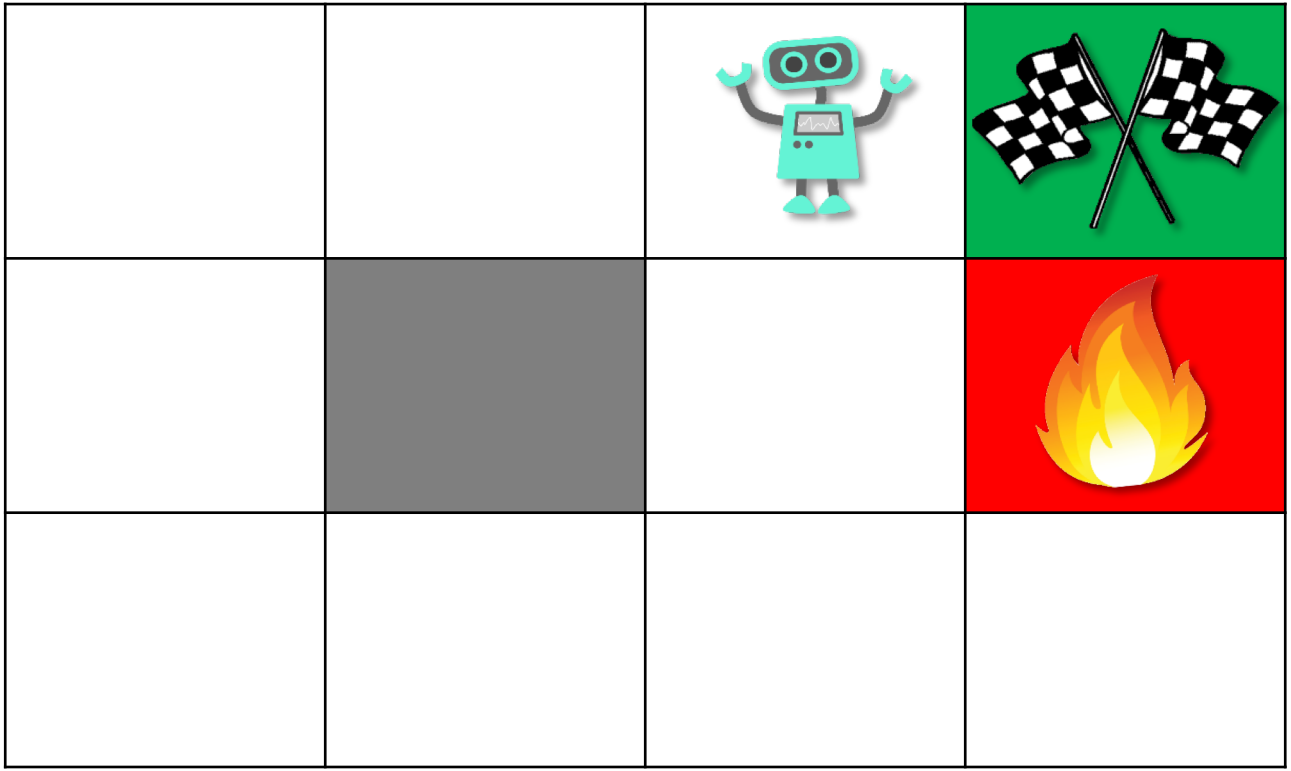
© SuperDataScience & IA Expert

A Equação de Bellman



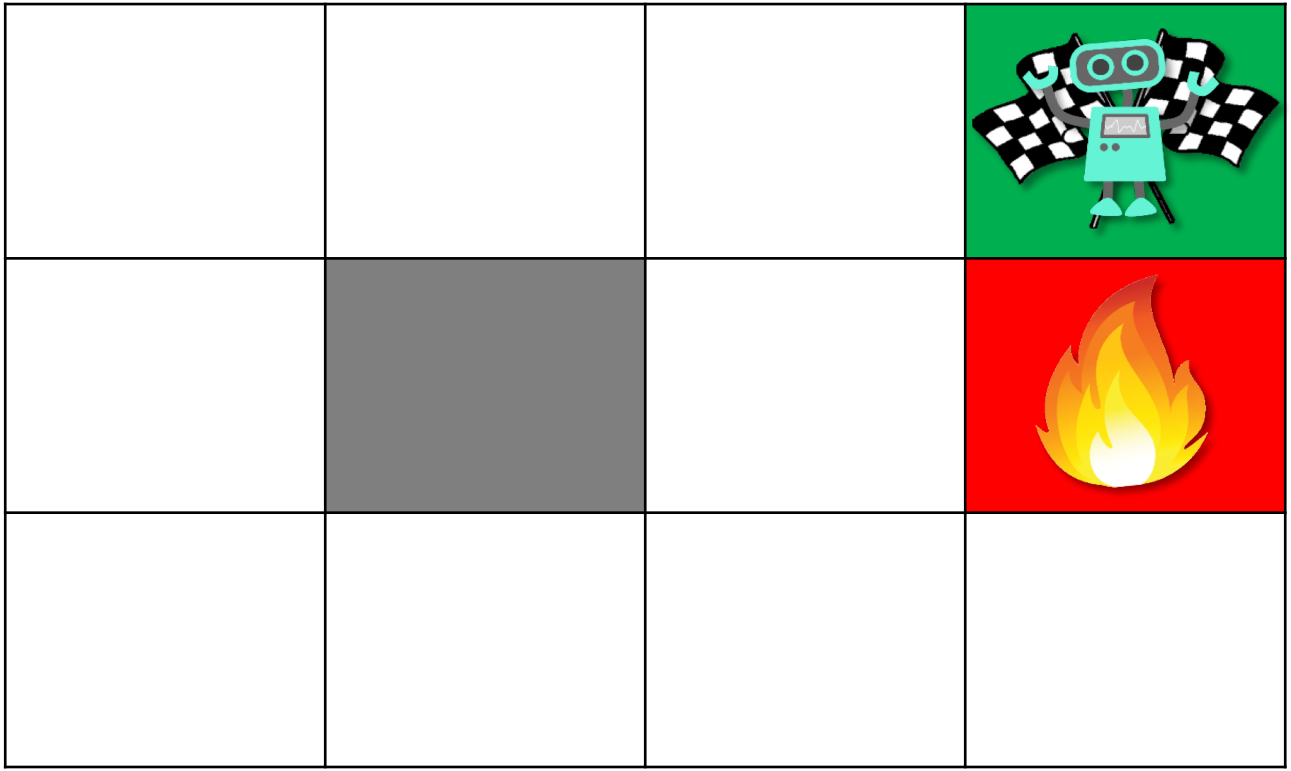
© SuperDataScience & IA Expert

A Equação de Bellman



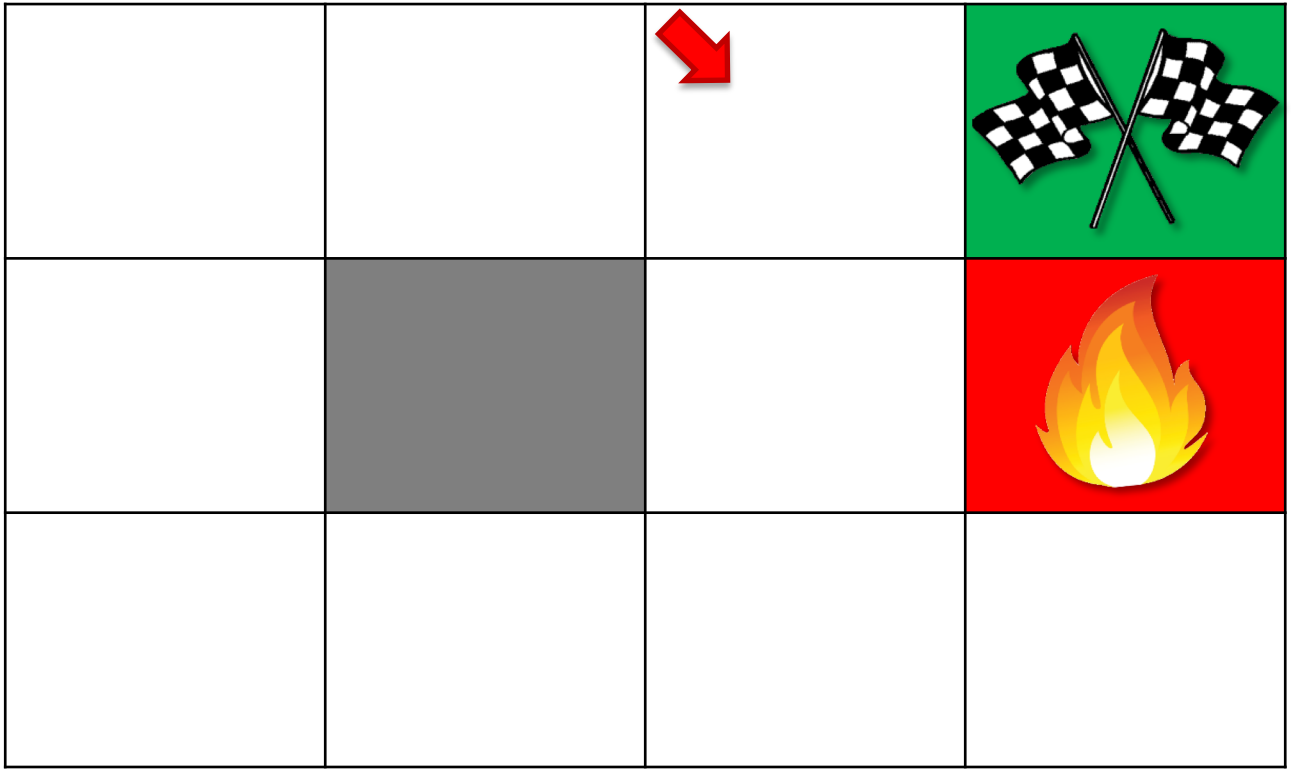
© SuperDataScience & IA Expert

A Equação de Bellman



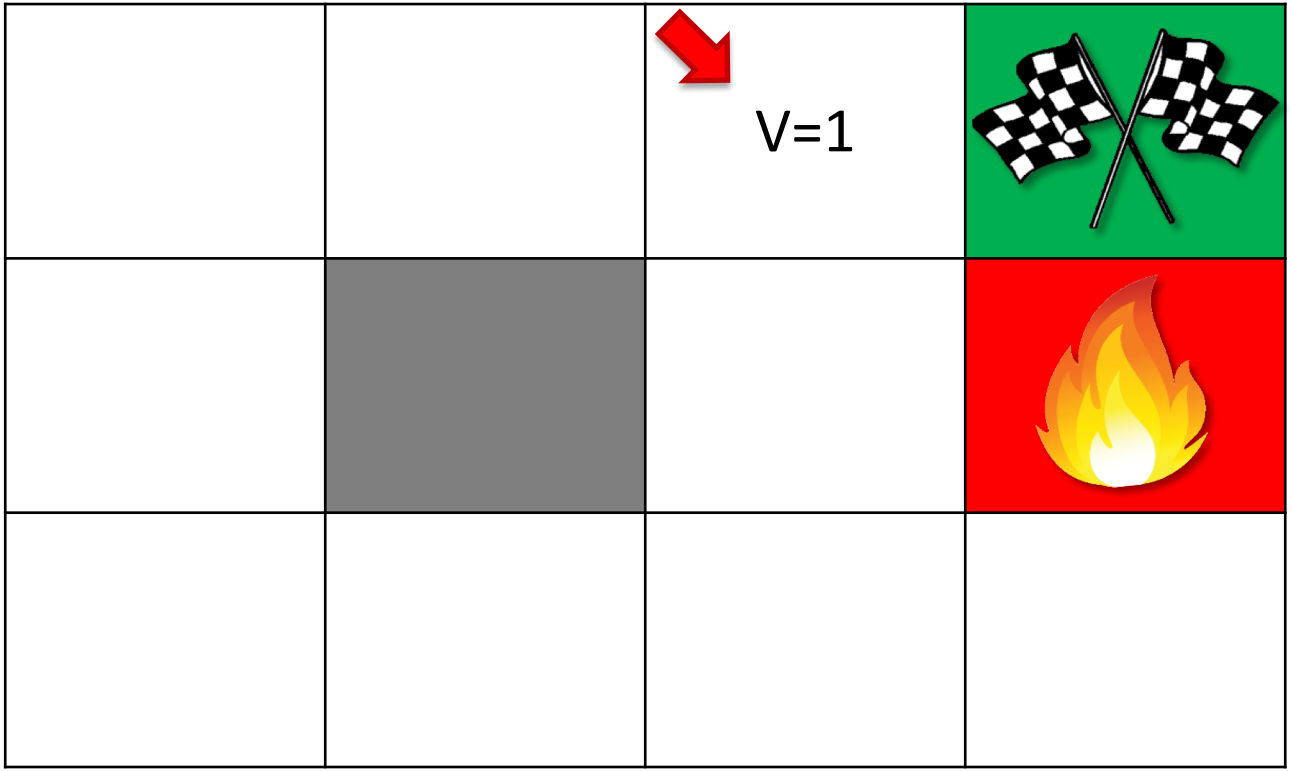
© SuperDataScience & IA Expert

A Equação de Bellman



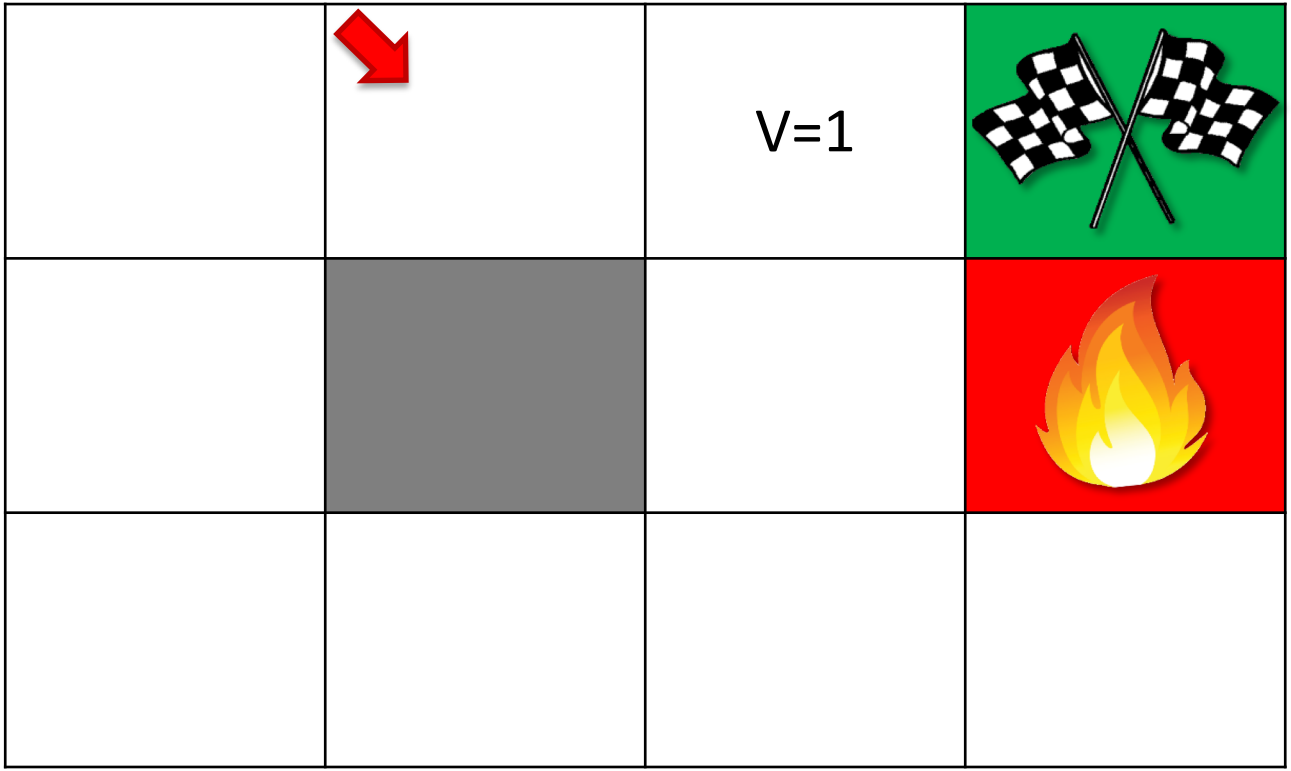
© SuperDataScience & IA Expert

A Equação de Bellman



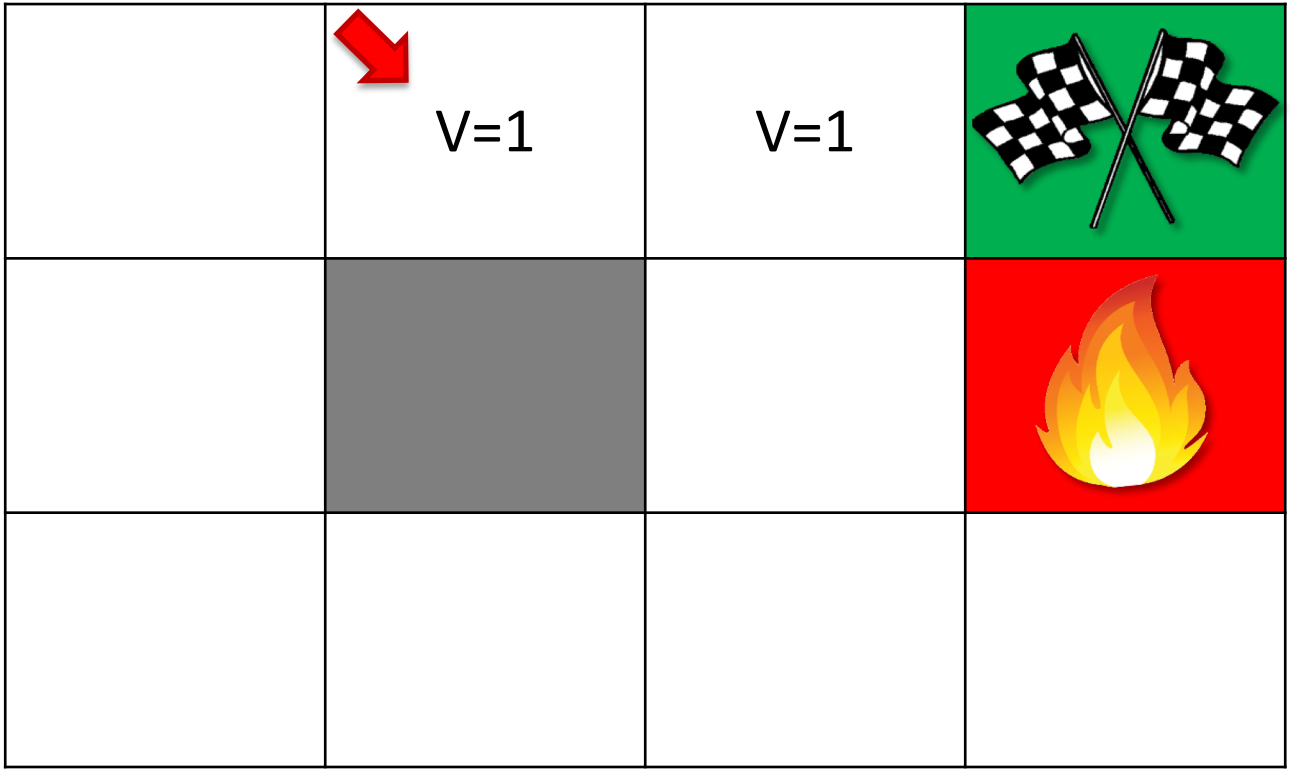
© SuperDataScience & IA Expert

A Equação de Bellman



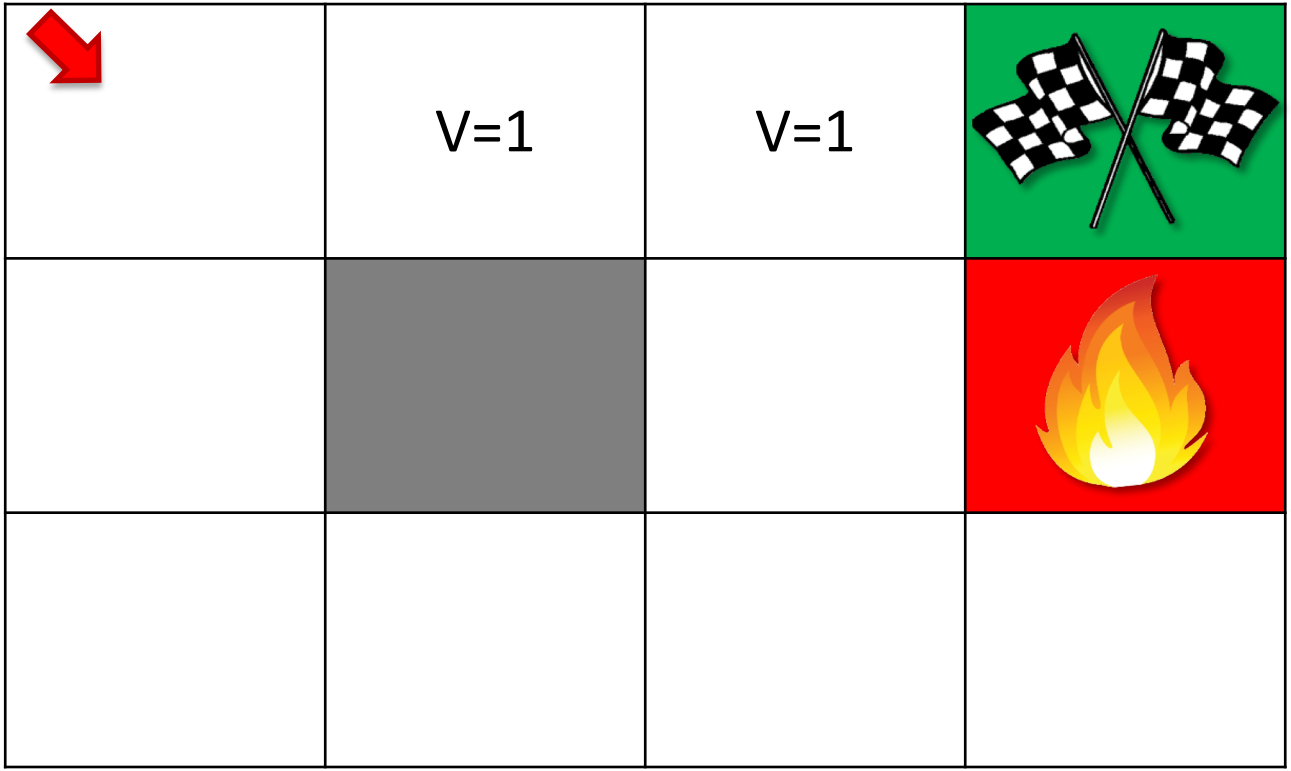
© SuperDataScience & IA Expert

A Equação de Bellman



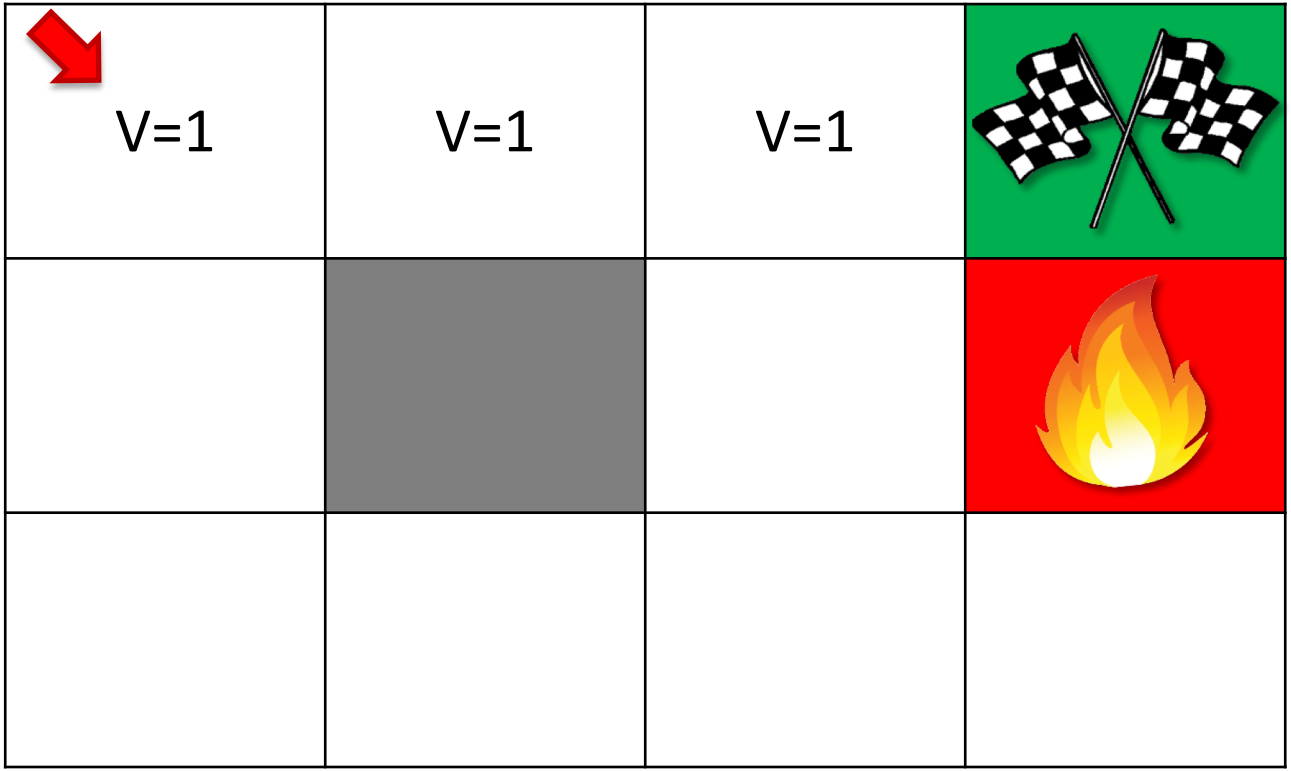
© SuperDataScience & IA Expert

A Equação de Bellman



© SuperDataScience & IA Expert

A Equação de Bellman



© SuperDataScience & IA Expert

A Equação de Bellman

V=1	V=1	V=1	
			

© SuperDataScience & IA Expert

A Equação de Bellman

V=1	V=1	V=1	
V=1			

© SuperDataScience & IA Expert

A Equação de Bellman

V=1	V=1	V=1	
V=1			
			

© SuperDataScience & IA Expert

A Equação de Bellman

V=1	V=1	V=1	
V=1			
V=1			

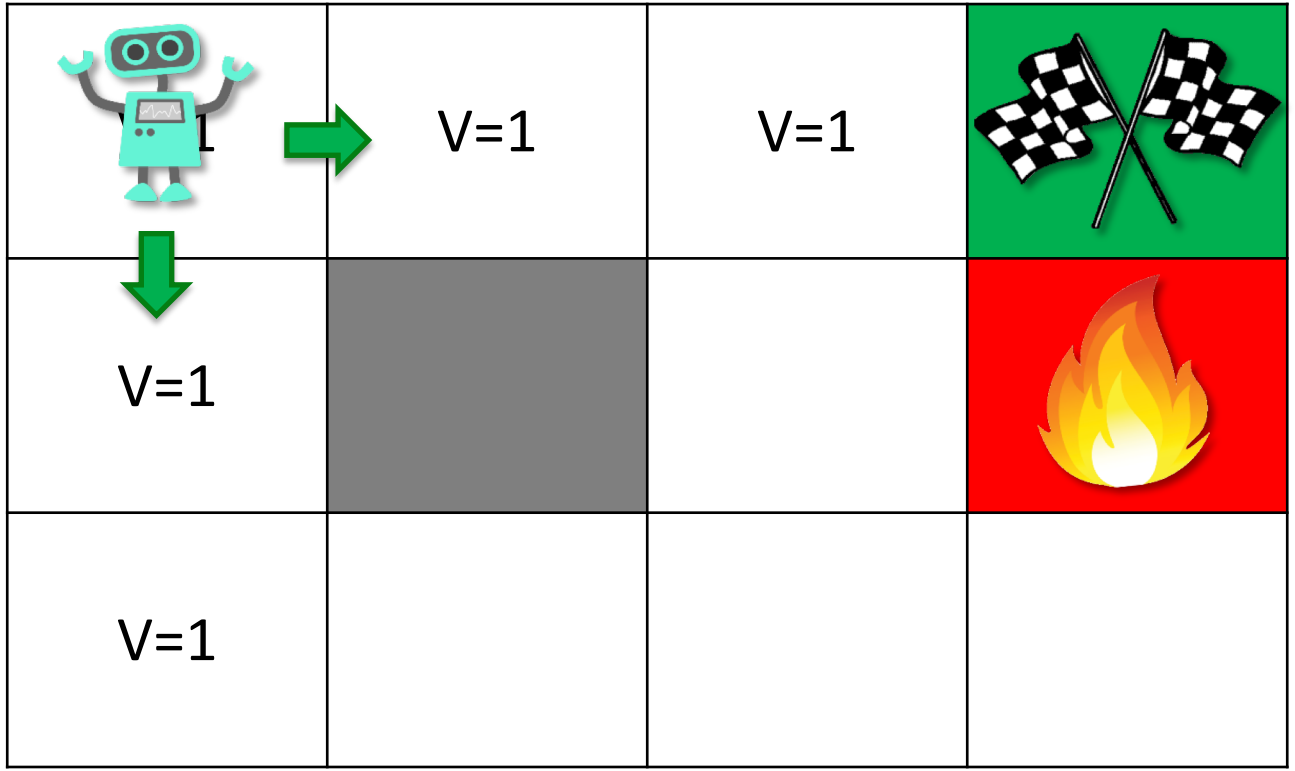
© SuperDataScience & IA Expert

A Equação de Bellman

V=1	V=1	V=1	
V=1			
V=1			

© SuperDataScience & IA Expert

A Equação de Bellman



© SuperDataScience & IA Expert

A Equação de Bellman

V = 1	V = 1	V = 1	
V = 1			
V = 1			

© SuperDataScience & IA Expert

A Equação de Bellman

© SuperDataScience & IA Expert

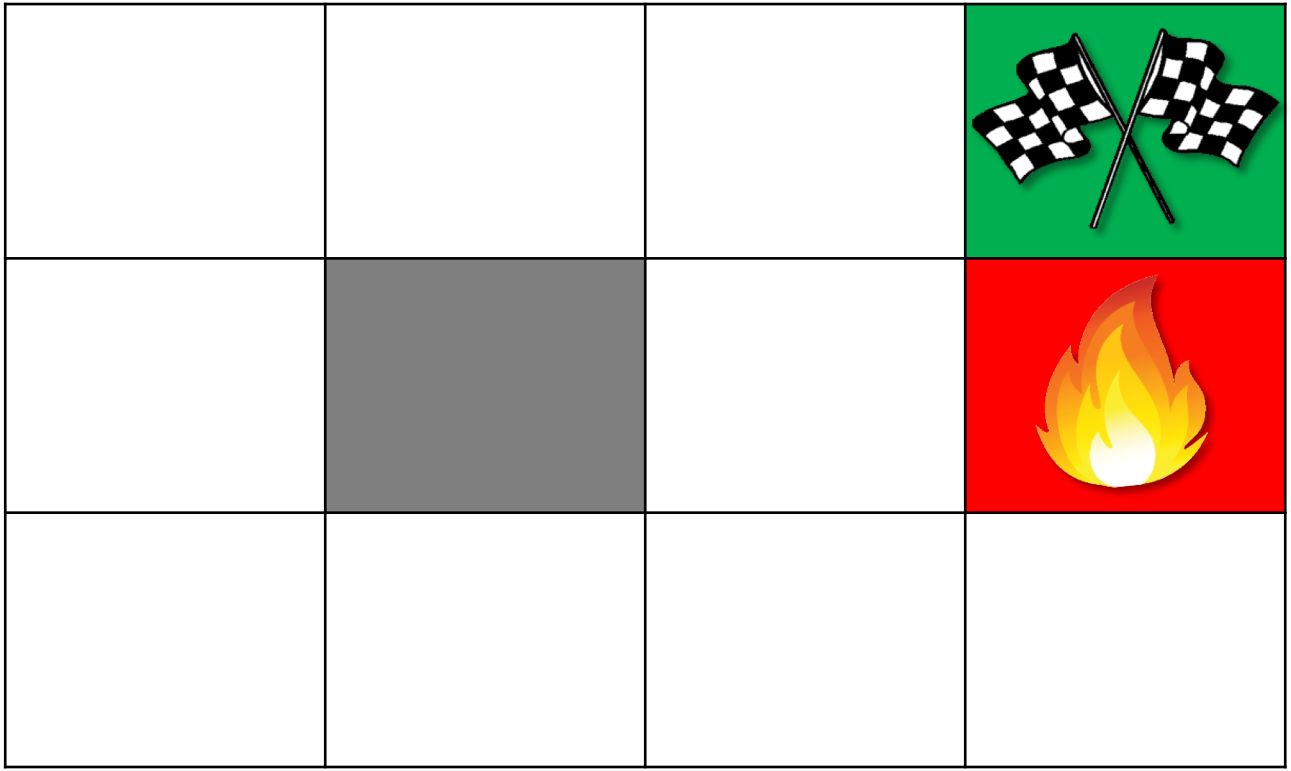
A Equação de Bellman

$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$

© SuperDataScience & IA Expert

A Equação de Bellman

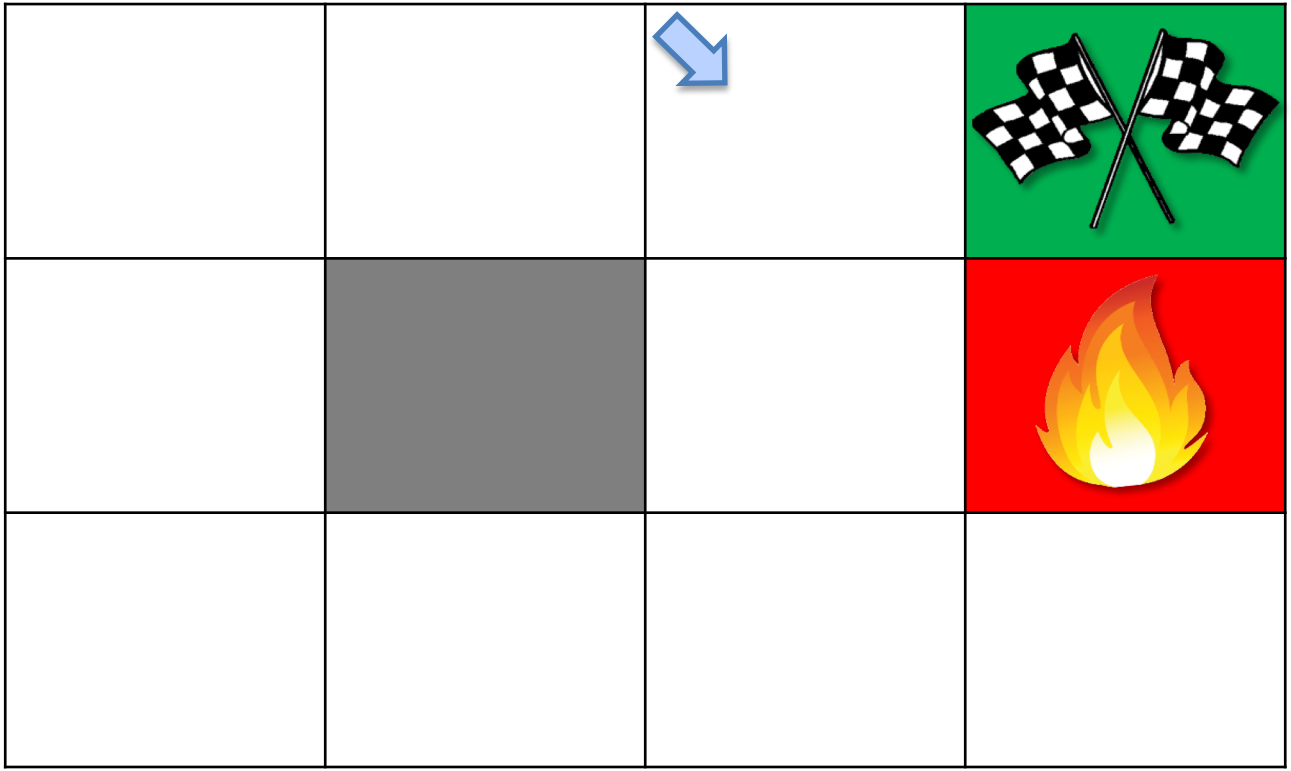
$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$



© SuperDataScience & IA Expert

A Equação de Bellman

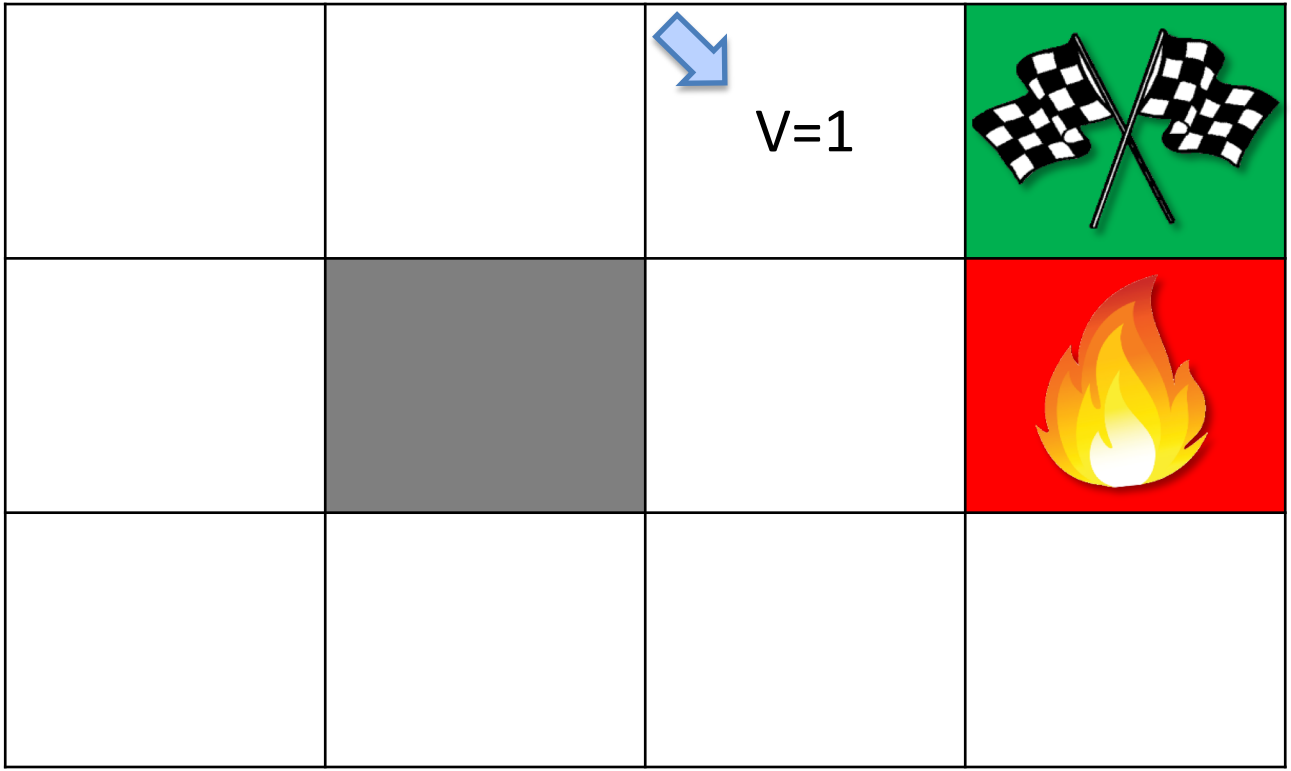
$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$



© SuperDataScience & IA Expert

A Equação de Bellman

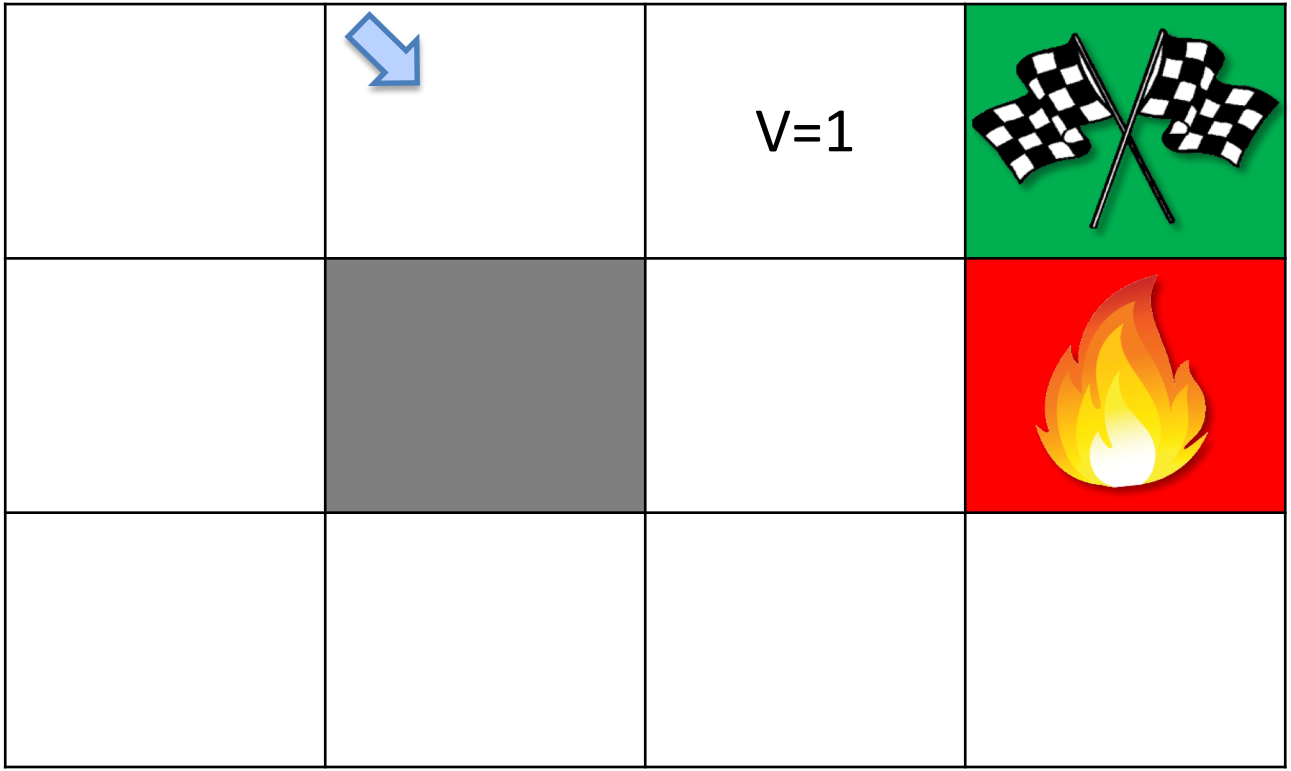
$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$



© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

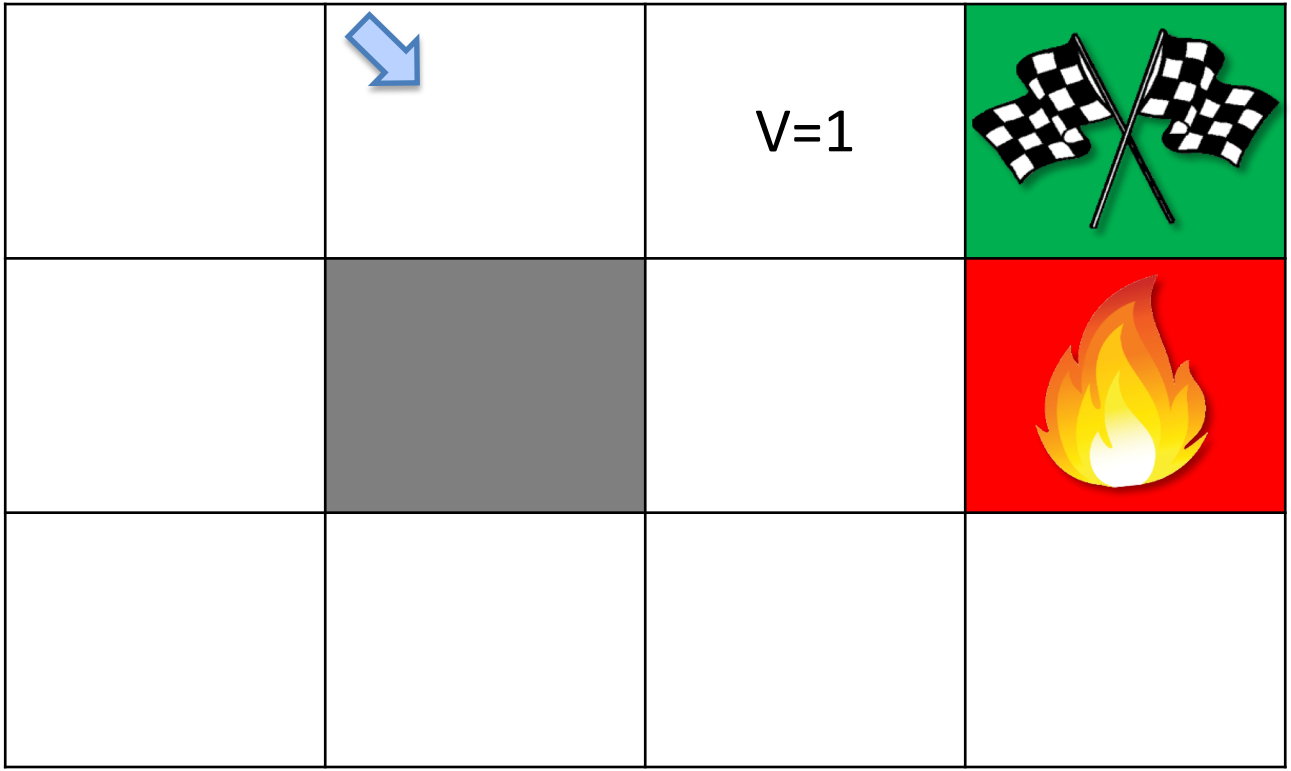


© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

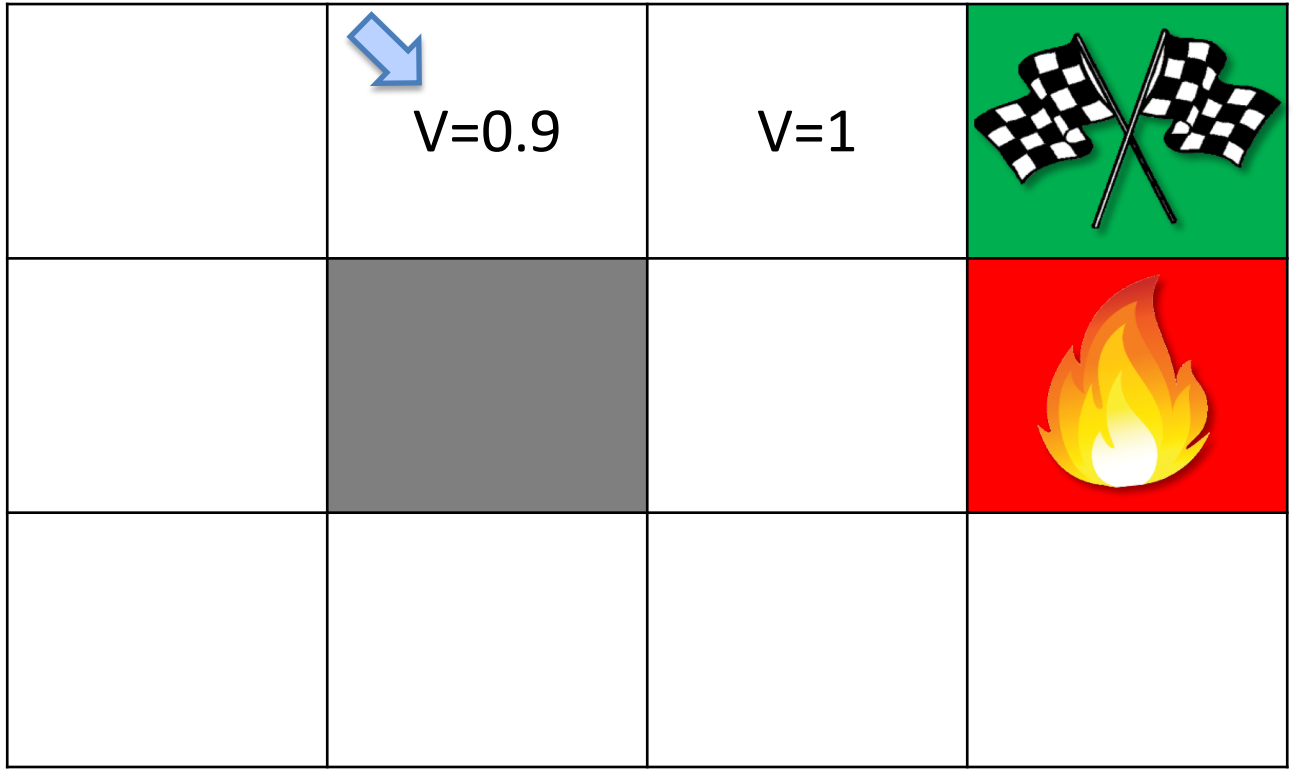


© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$



© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

↘	V=0.9	V=1	
			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73			
V=0.66			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73			
V=0.66			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73			
V=0.66			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

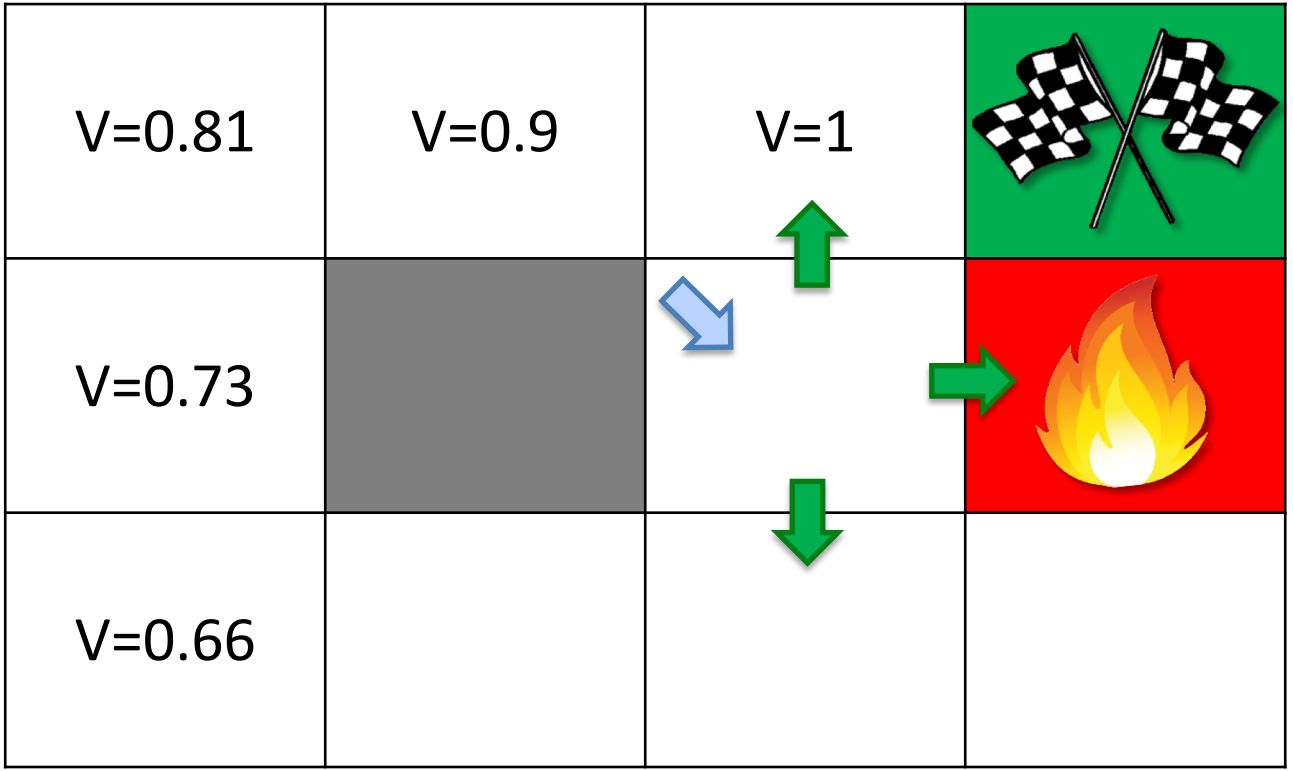
V=0.81	V=0.9	V=1	
V=0.73			
V=0.66			

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

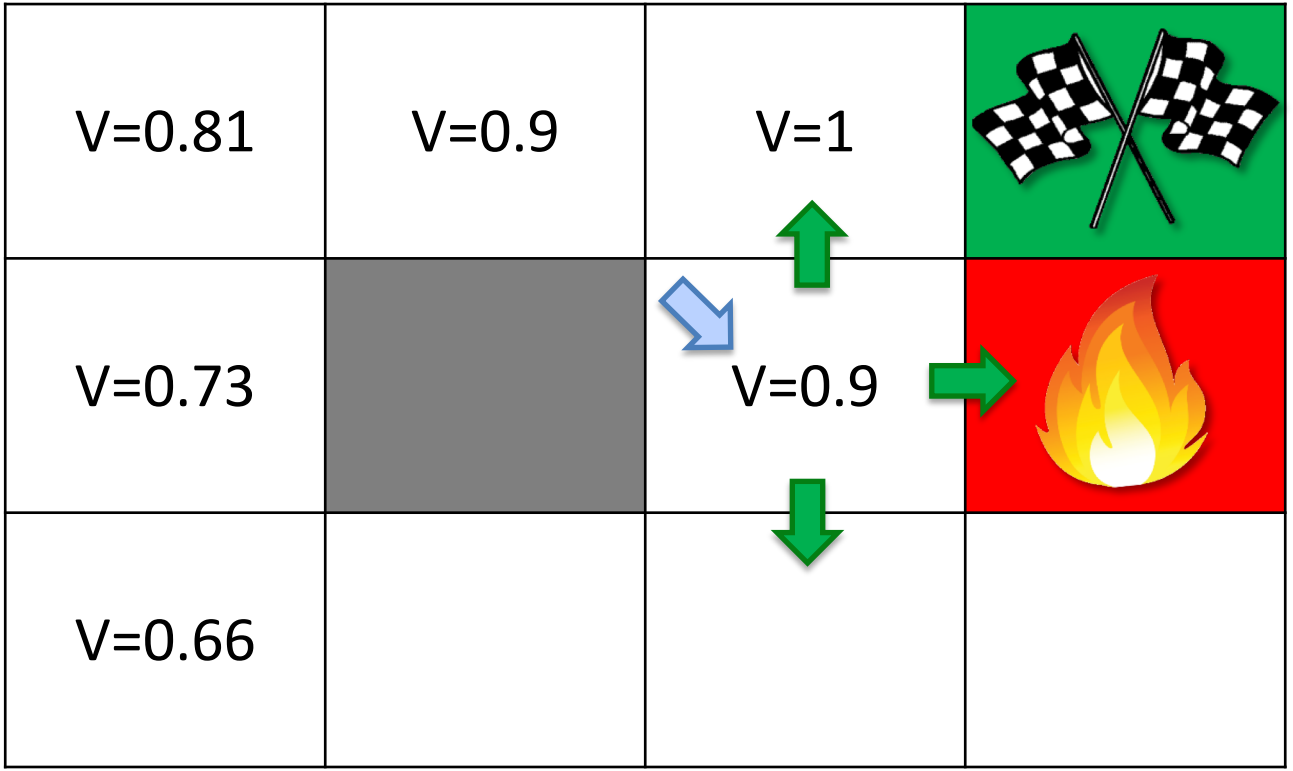


© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$



© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

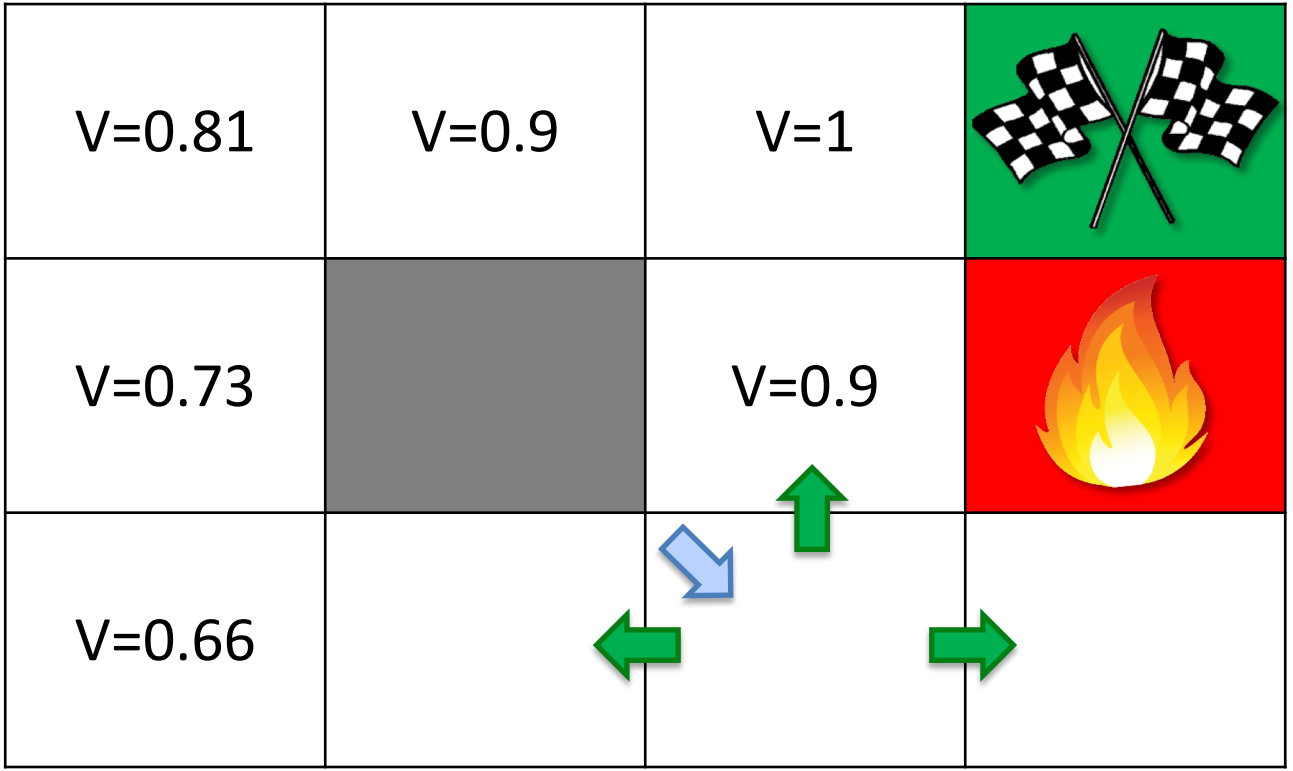
V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66		↘	

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

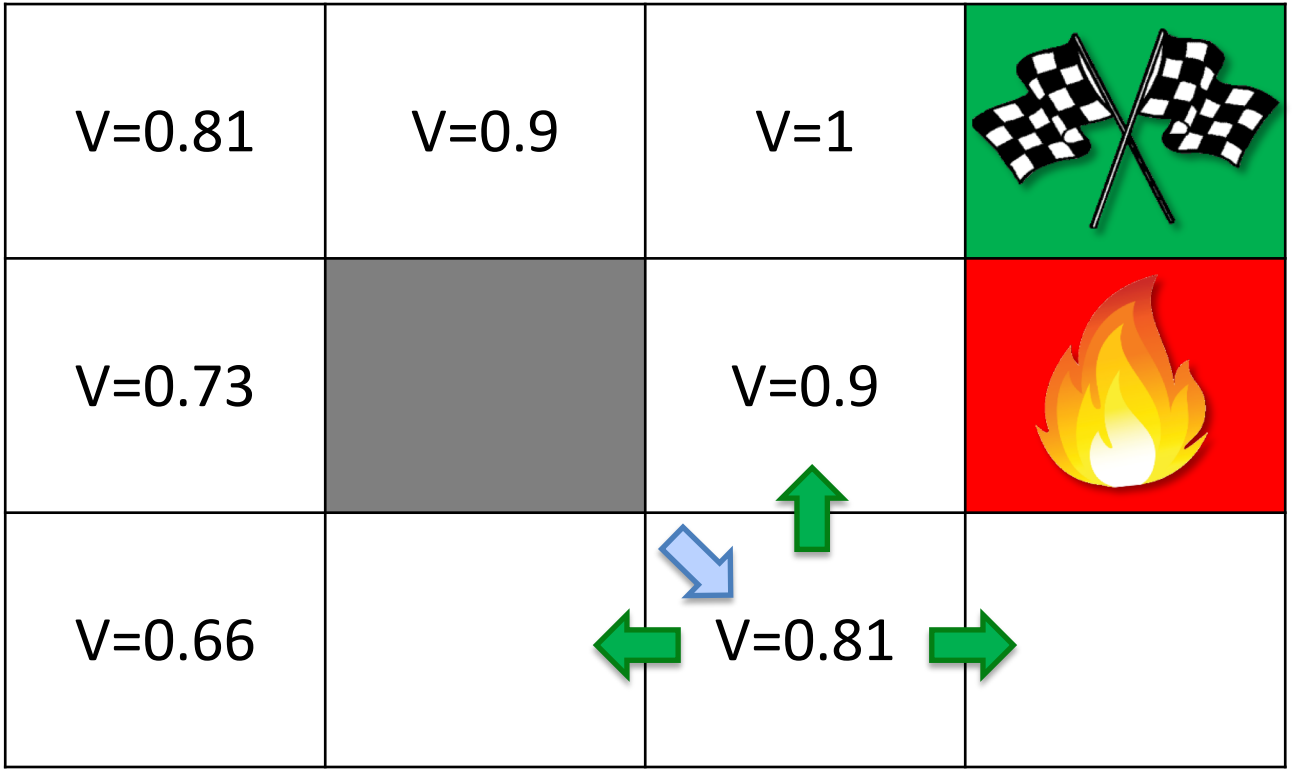


© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$



© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66		V=0.81	

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66		V=0.81	

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

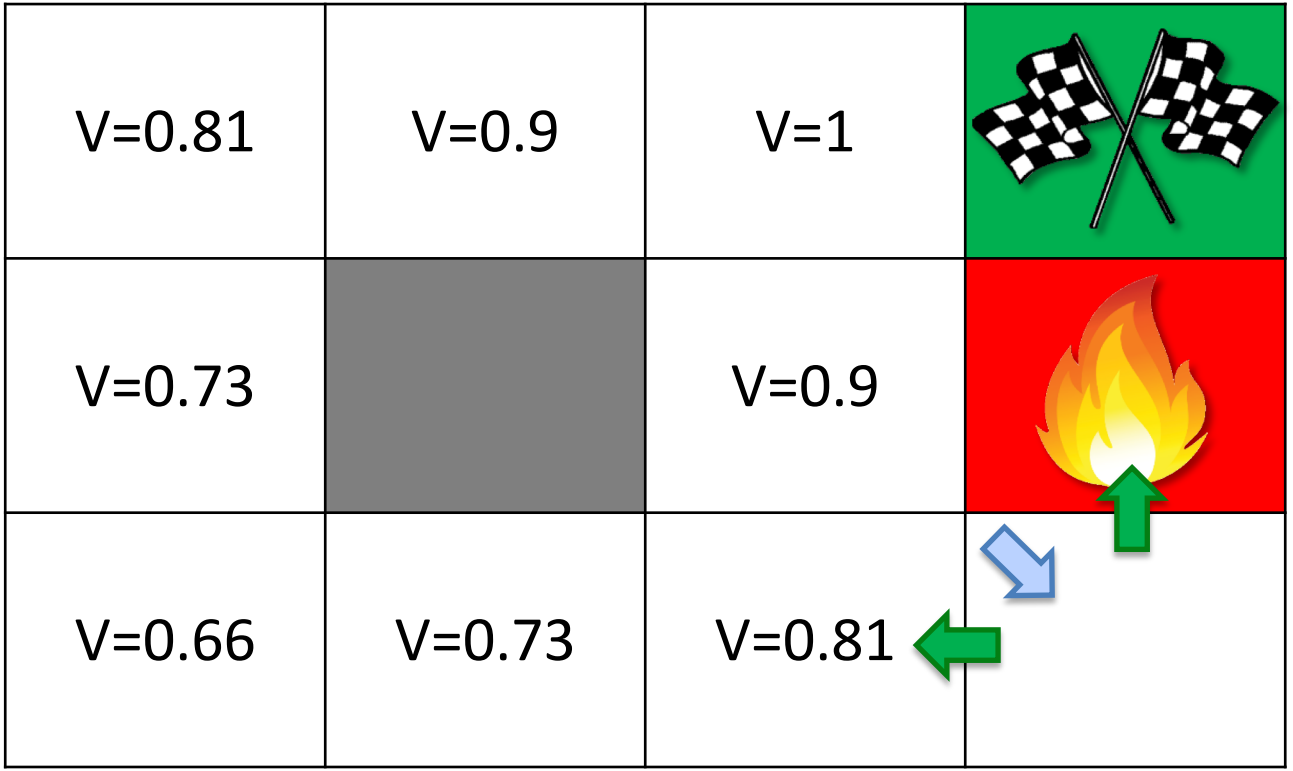
V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	

© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

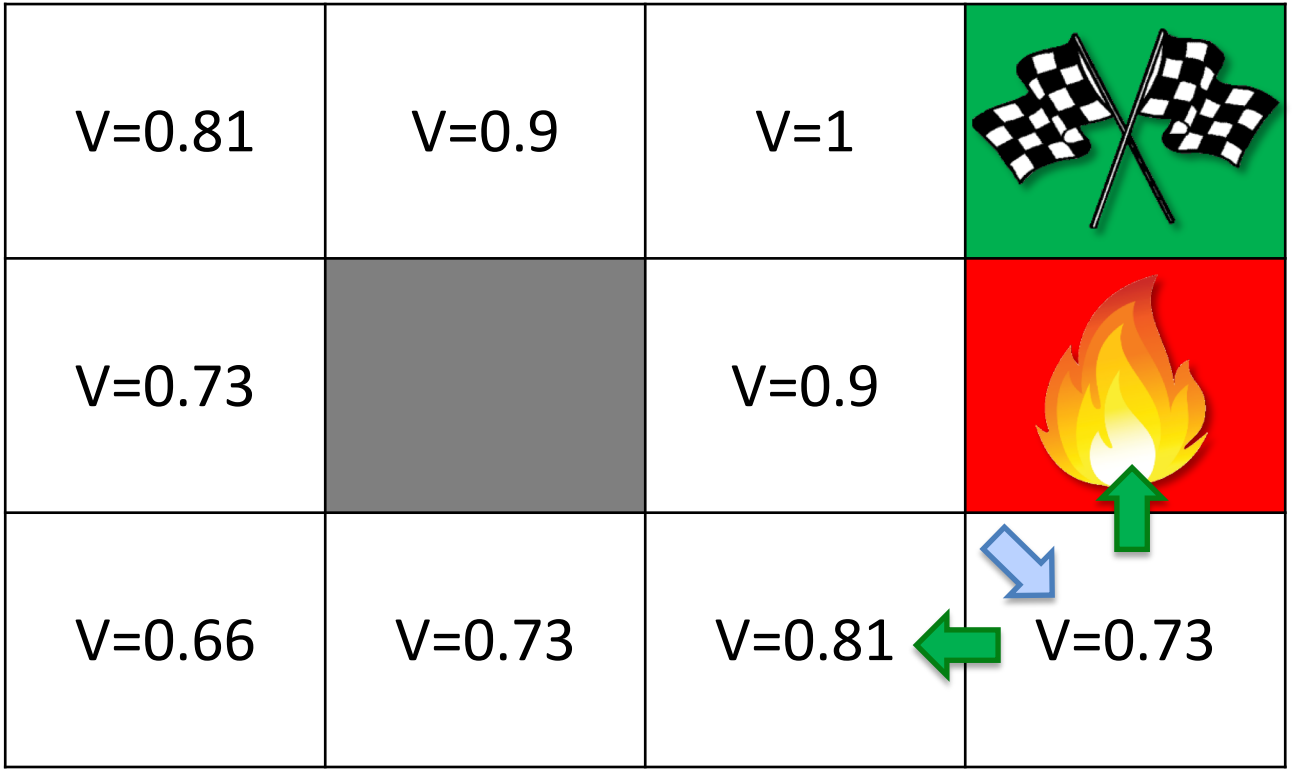


© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$



© SuperDataScience & IA Expert

A Equação de Bellman

$$V(s) = \max_a(R(s,a) + \gamma V(s'))$$

$\gamma = 0.9$

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

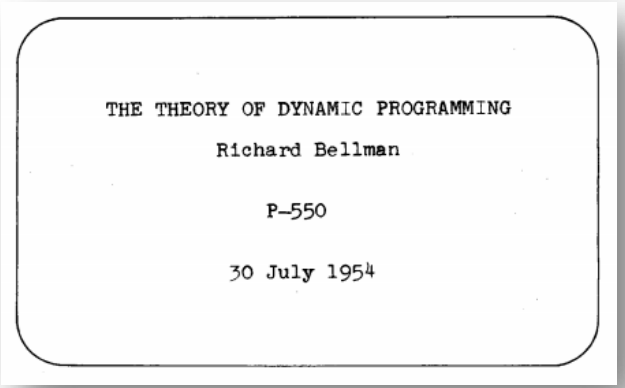
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

The Theory of Dynamic Programming

Richard Bellman (1954)


THE THEORY OF DYNAMIC PROGRAMMING

Richard Bellman

P–550

30 July 1954

Link:

https://www.rand.org/content/dam/rand/pubs/papers/2008/P550.pdf

© SuperDataScience & IA Expert

O “Plano”

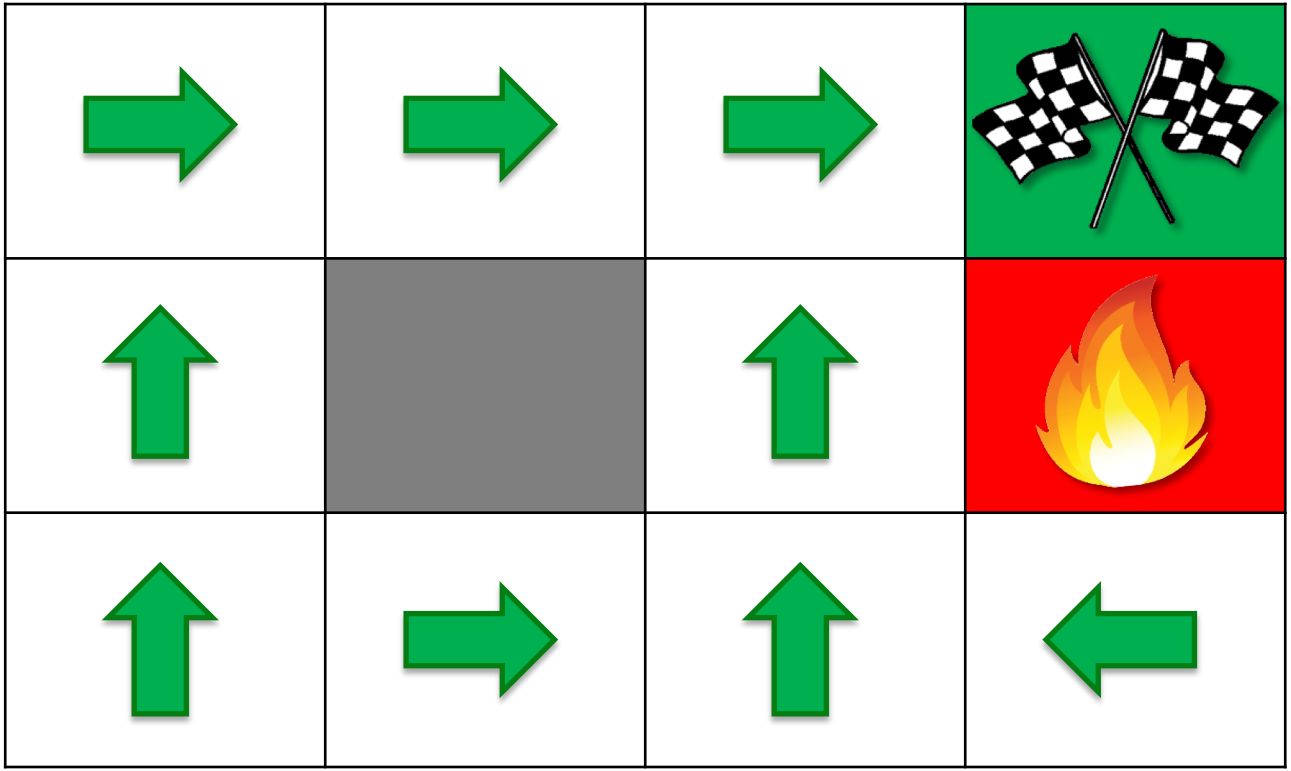
© SuperDataScience & IA Expert

A Equação de Bellman

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

© SuperDataScience & IA Expert

A Equação de Bellman

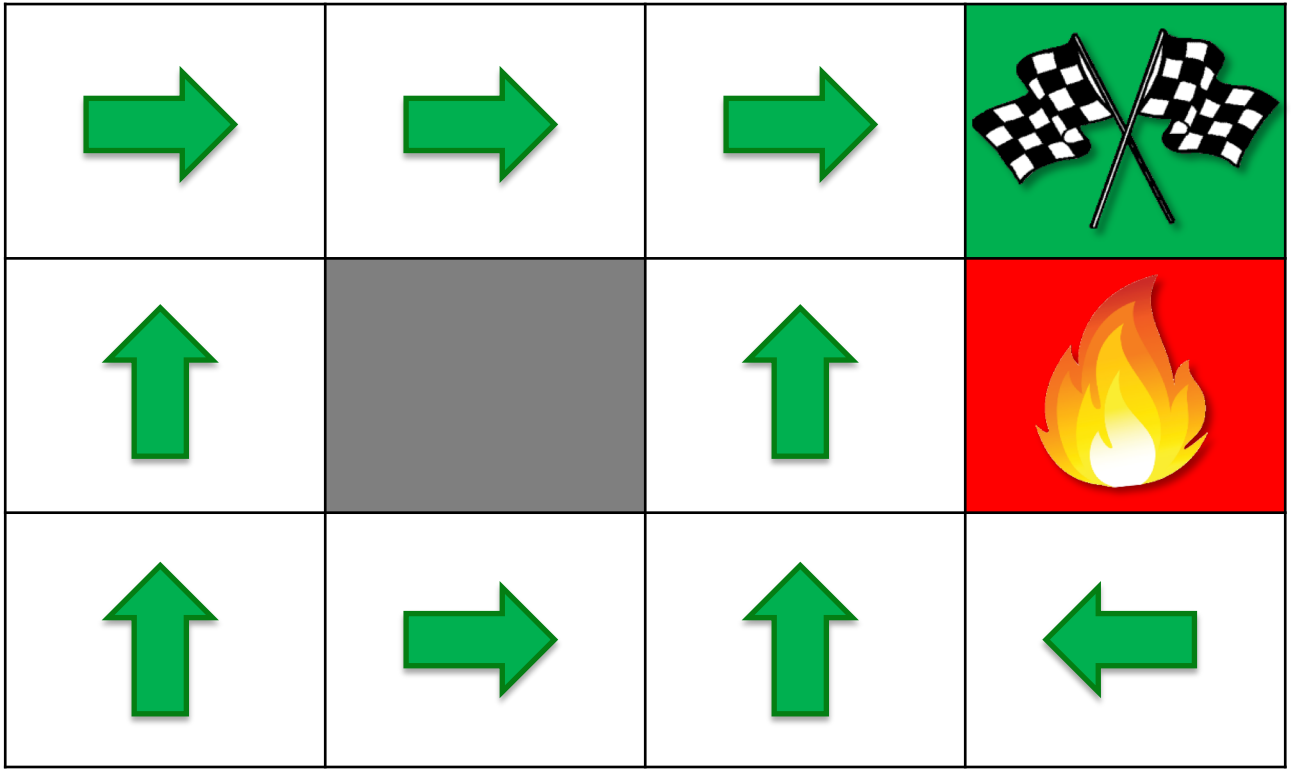


© SuperDataScience & IA Expert

Markov Decision Process (MDP)

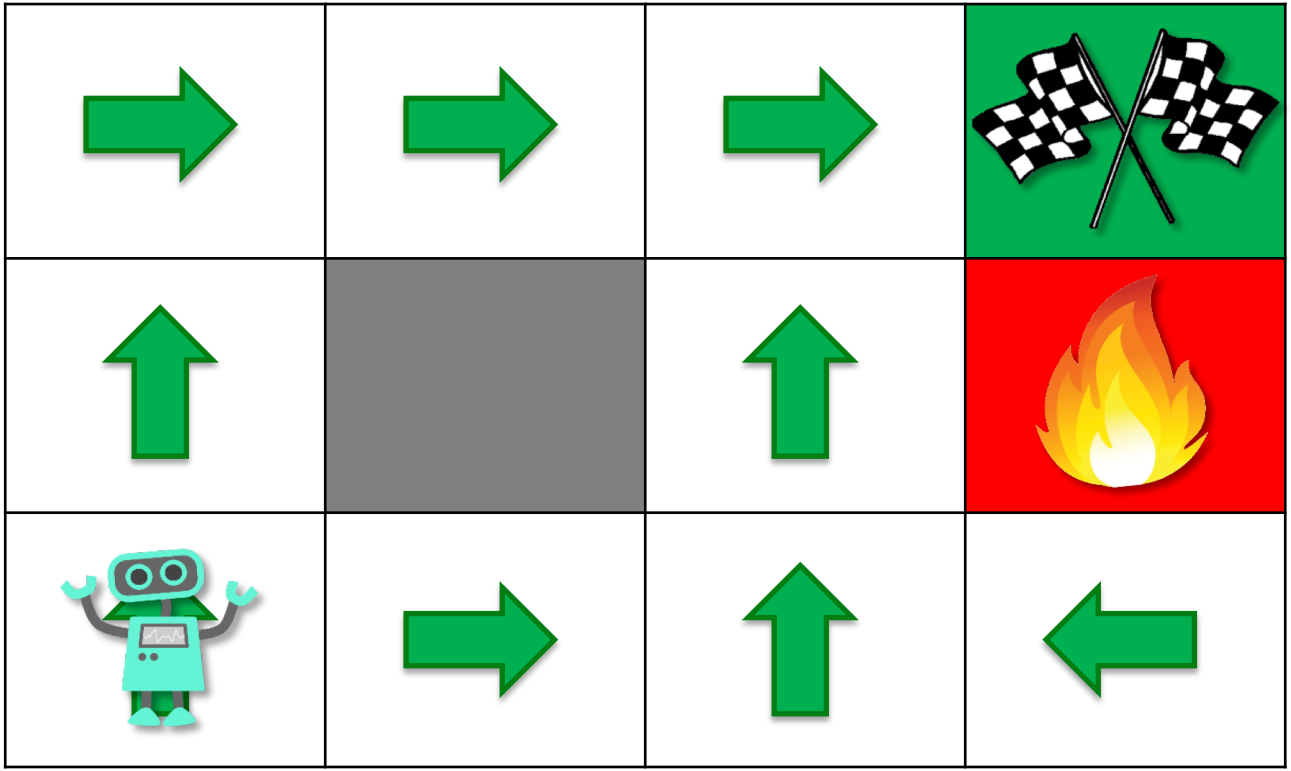
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



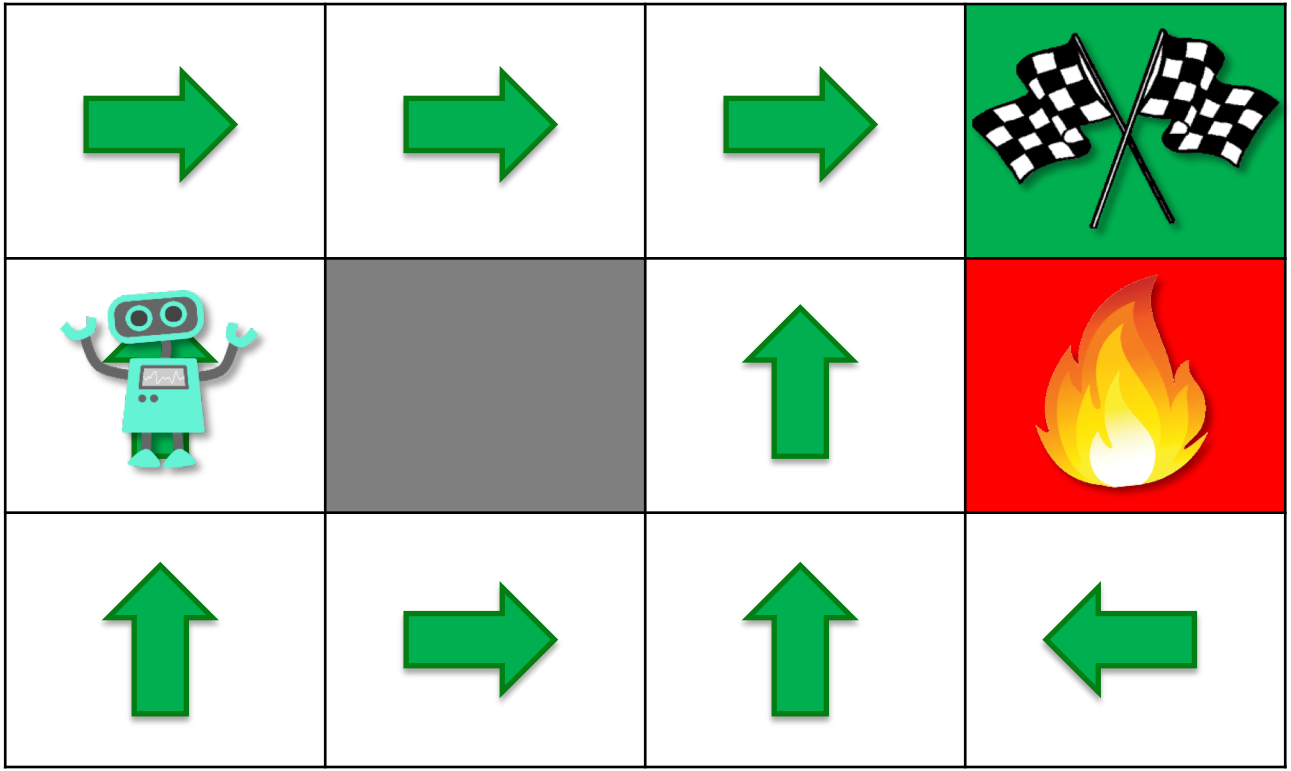
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



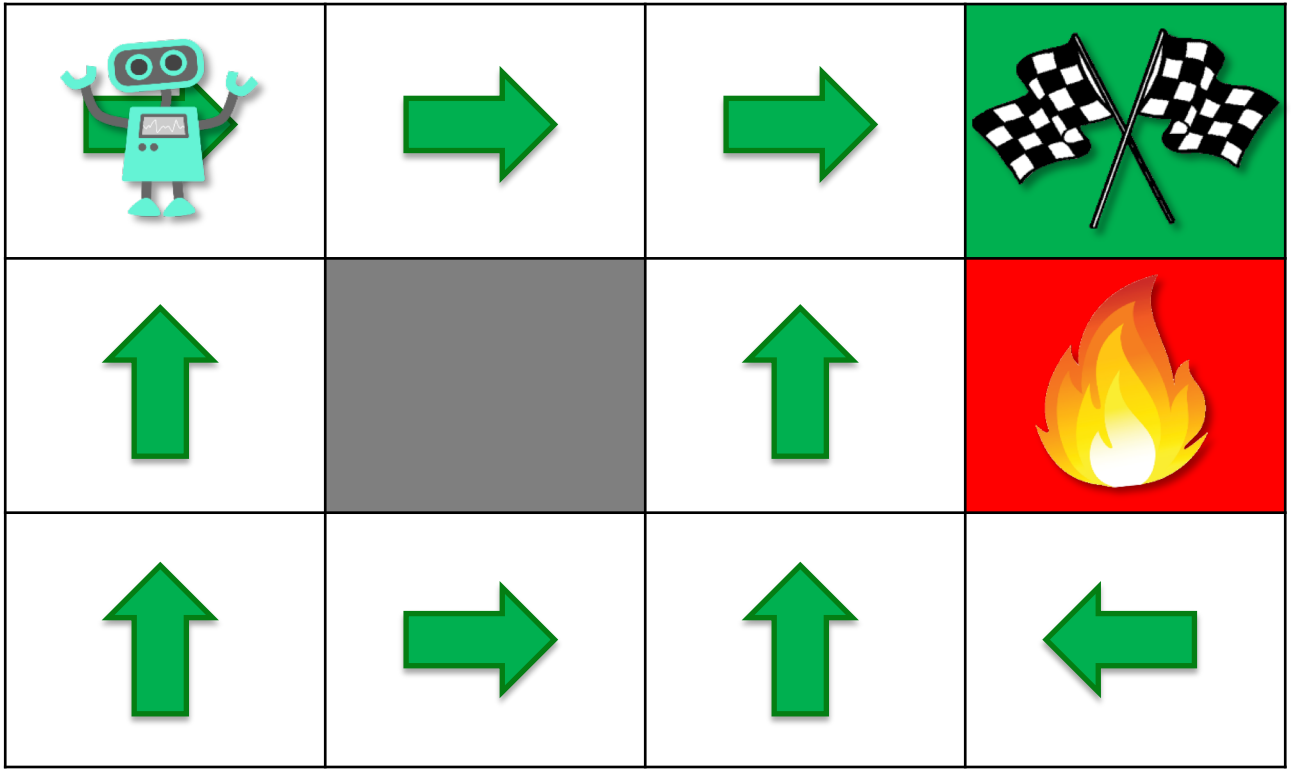
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



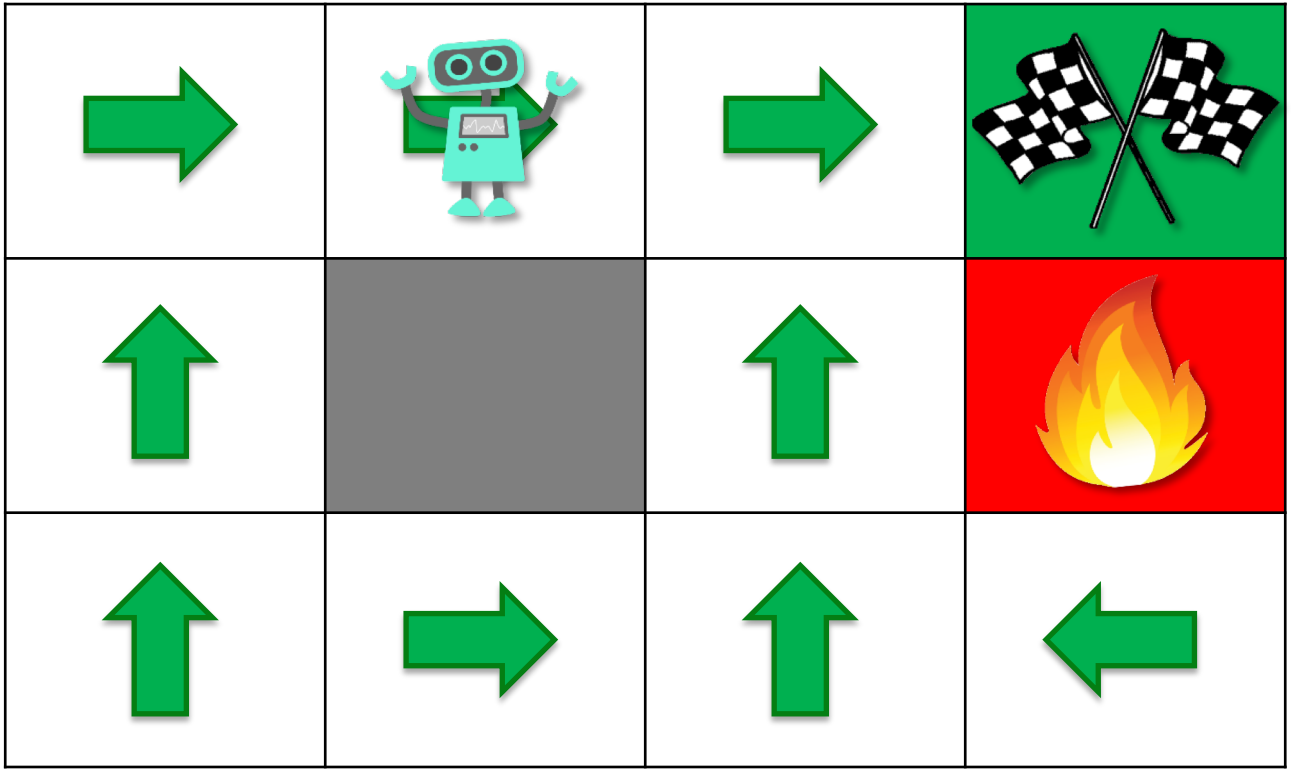
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



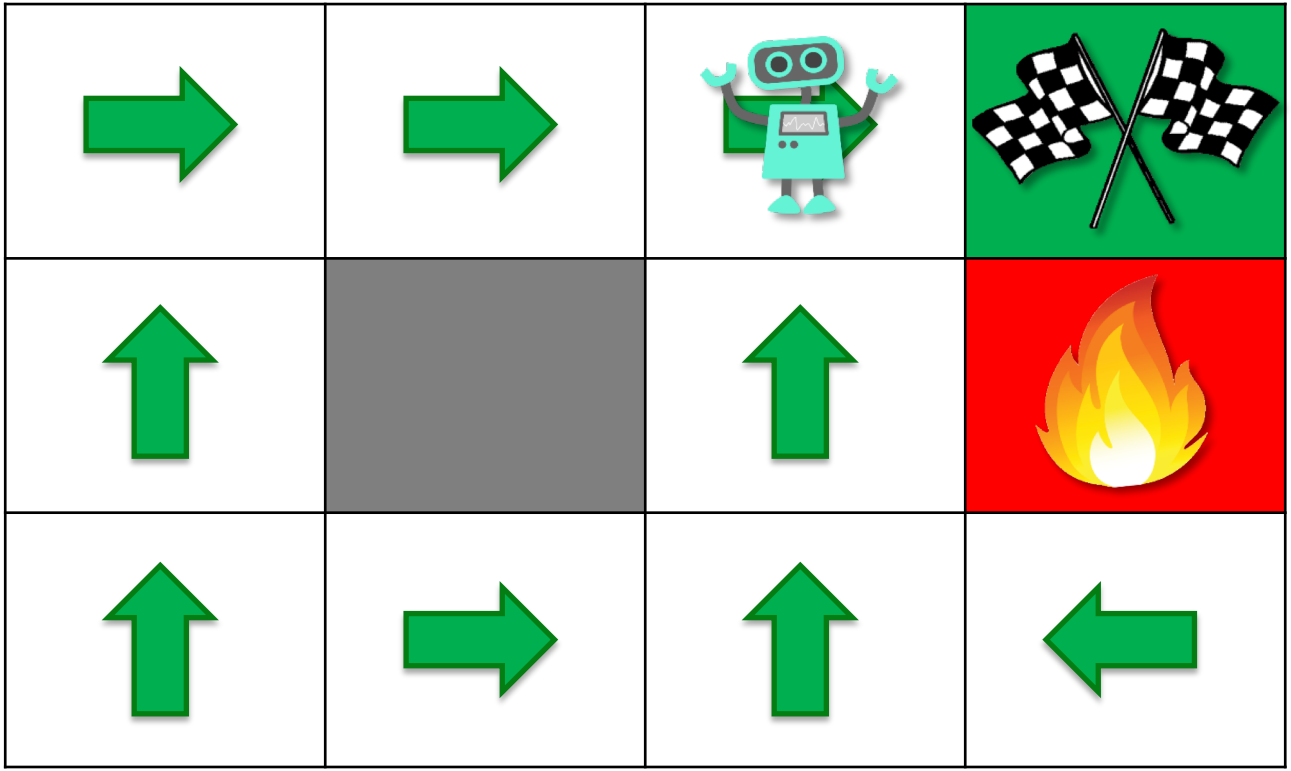
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



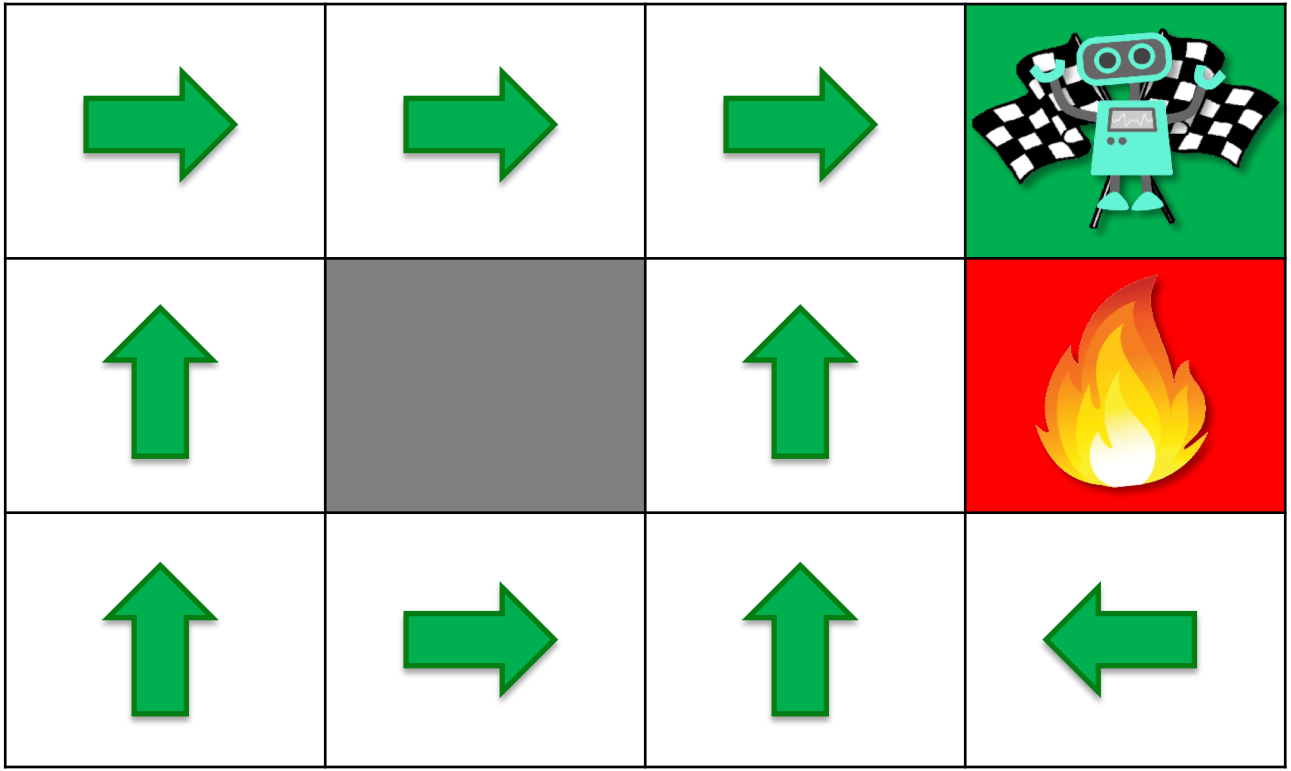
© SuperDataScience & IA Expert

Markov Decision Process (MDP)



© SuperDataScience & IA Expert

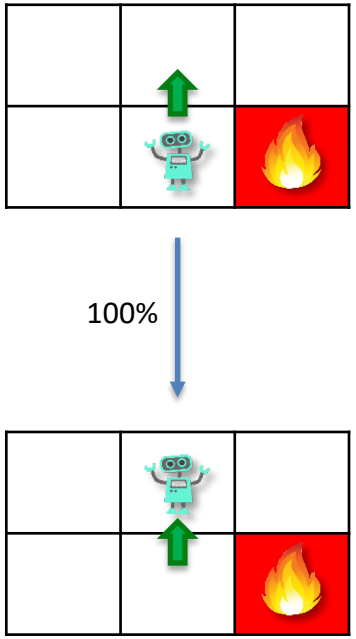
Markov Decision Process (MDP)



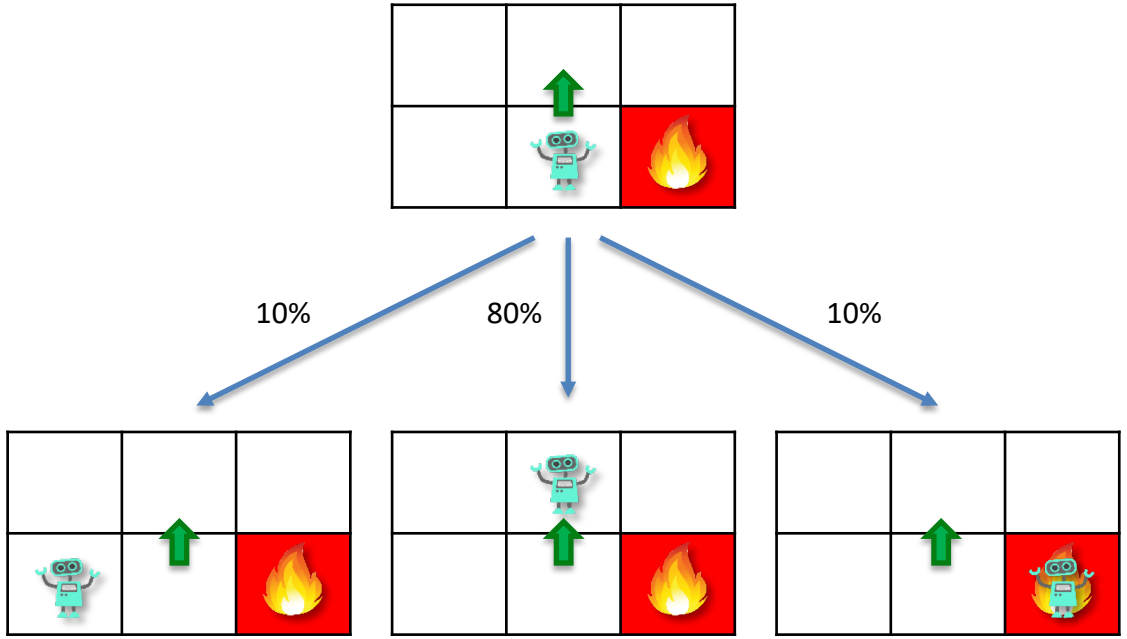
© SuperDataScience & IA Expert

Markov Decision Process (MDP)

Busca Determinística



Busca Não-Determinística



© SuperDataScience & IA Expert

Markov Decision Process (MDP)

Markov Process

Markov Decision Process (MDP)

© SuperDataScience & IA Expert

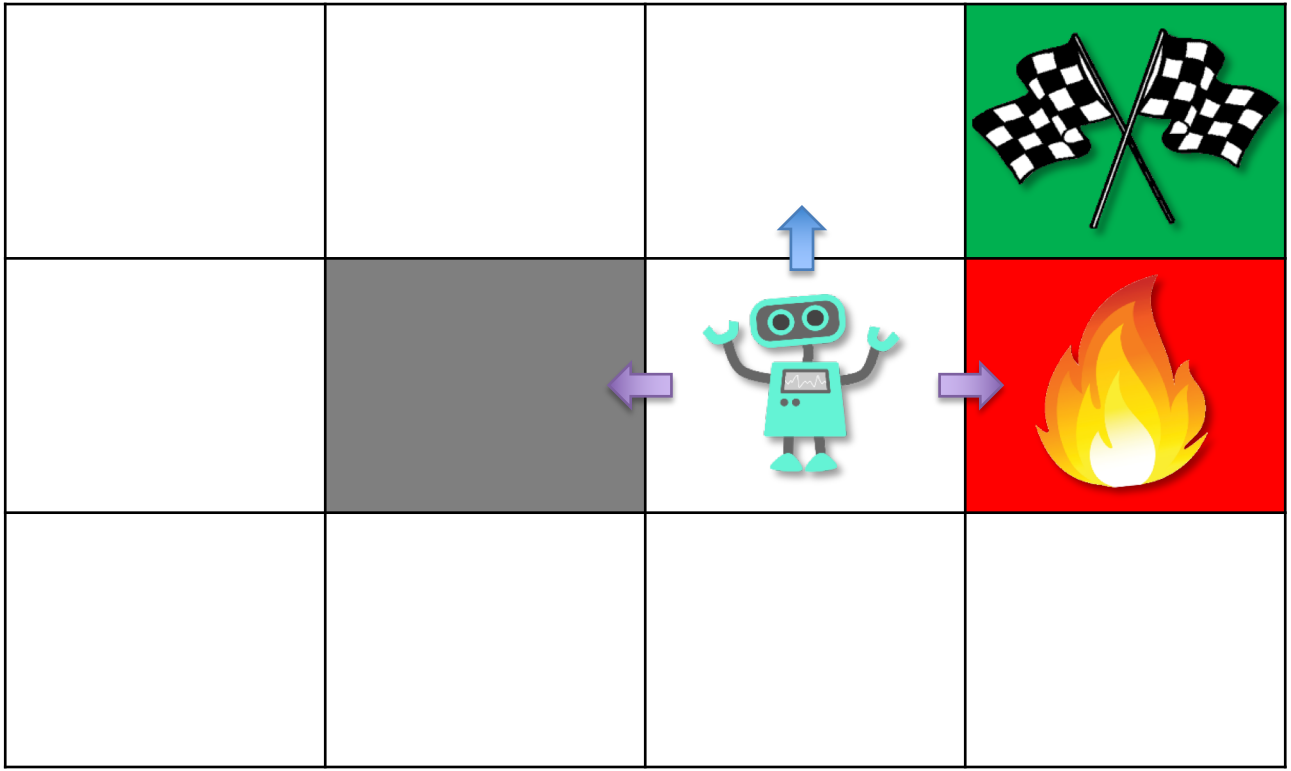
Markov Decision Process (MDP)

Um processo estocástico possui a propriedade Markov se a distribuição de probabilidade condicional de estados futuros do processo (condicional em estados passados e presentes) depende apenas do estado presente, não da sequência de eventos que o precedeu. Um processo com essa propriedade é chamado de processo de Markov.

Wikipedia (traduzido para português)

© SuperDataScience & IA Expert

Markov Decision Process (MDP)



© SuperDataScience & IA Expert

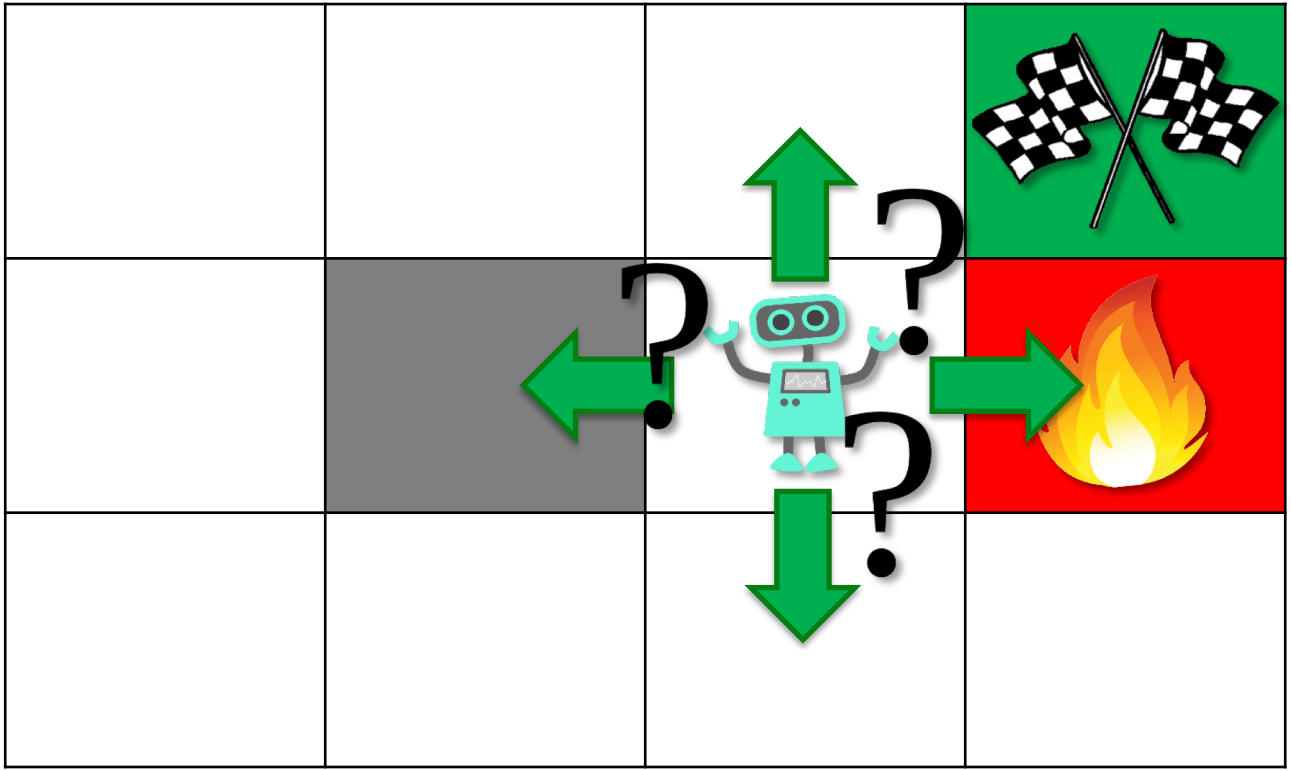
Markov Decision Process (MDP)

Markov Decision Processes (MDPs – Processos de Decisão de Makov) fornecem uma estrutura matemática para modelar a tomada de decisões em situações onde os resultados são parcialmente aleatórios e parcialmente sob o controle de um tomador de decisão.

Wikipedia (traduzido para o português)

© SuperDataScience & IA Expert

Markov Decision Process (MDP)



© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$

© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$

© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$

$s'_1 \quad s'_2 \quad s'_3$

© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$V(s) = \max_{a}(R(s,a) + \gamma V(s'))$$

$V(s_1')$ $V(s_2')$ $V(s_3')$

© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$0.8 * V(s'_1) + 0.1 * V(s'_2) + 0.1 * V(s'_3)$$

$$V(s) = \max_a (R(s,a) + \gamma V(s'))$$

© SuperDataScience & IA Expert

Markov Decision Process (MDP)

$$V(s) = \max_{a} \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

A Survey of Applications of Markov Decision Processes

D. J. White (1993)

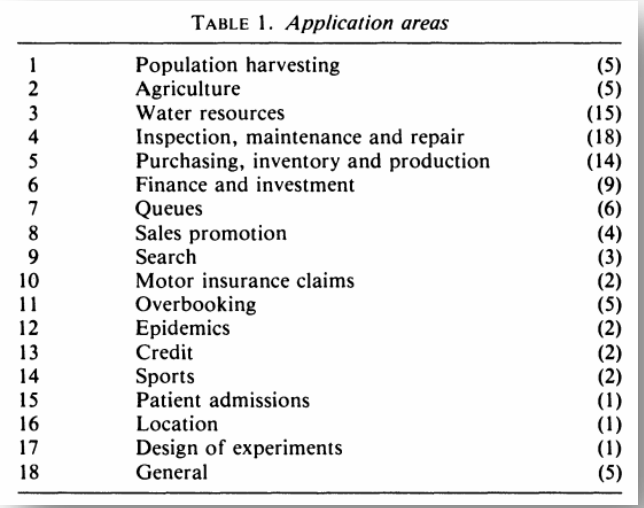


TABLE 1. *Application areas*

1	Population harvesting	(5)
2	Agriculture	(5)
3	Water resources	(15)
4	Inspection, maintenance and repair	(18)
5	Purchasing, inventory and production	(14)
6	Finance and investment	(9)
7	Queues	(6)
8	Sales promotion	(4)
9	Search	(3)
10	Motor insurance claims	(2)
11	Overbooking	(5)
12	Epidemics	(2)
13	Credit	(2)
14	Sports	(2)
15	Patient admissions	(1)
16	Location	(1)
17	Design of experiments	(1)
18	General	(5)

Link:

http://www.cs.uml.edu/ecg/uploads/AIfall14/MDPApplications3.pdf

© SuperDataScience & IA Expert

Política x Plano

© SuperDataScience & IA Expert

Política x Plano

$$V(s) = \max_{a}\left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

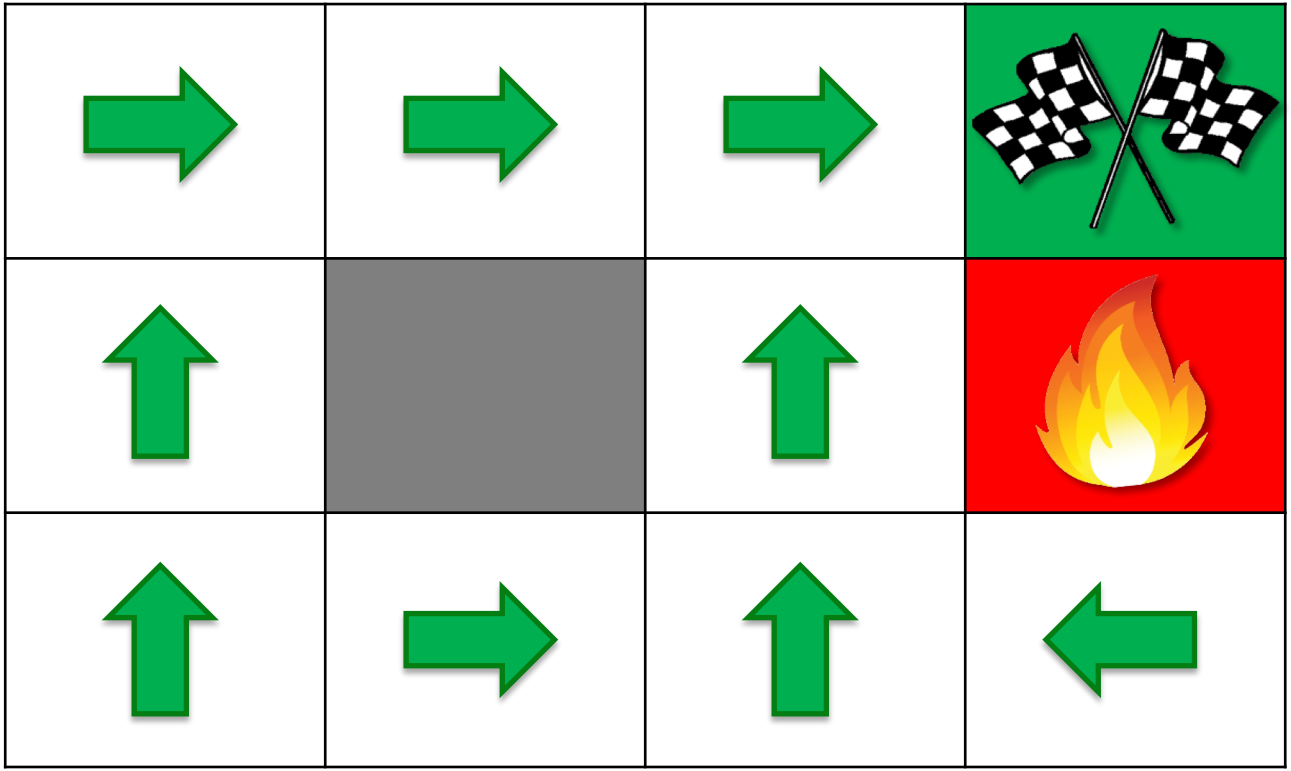
© SuperDataScience & IA Expert

Política x Plano

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

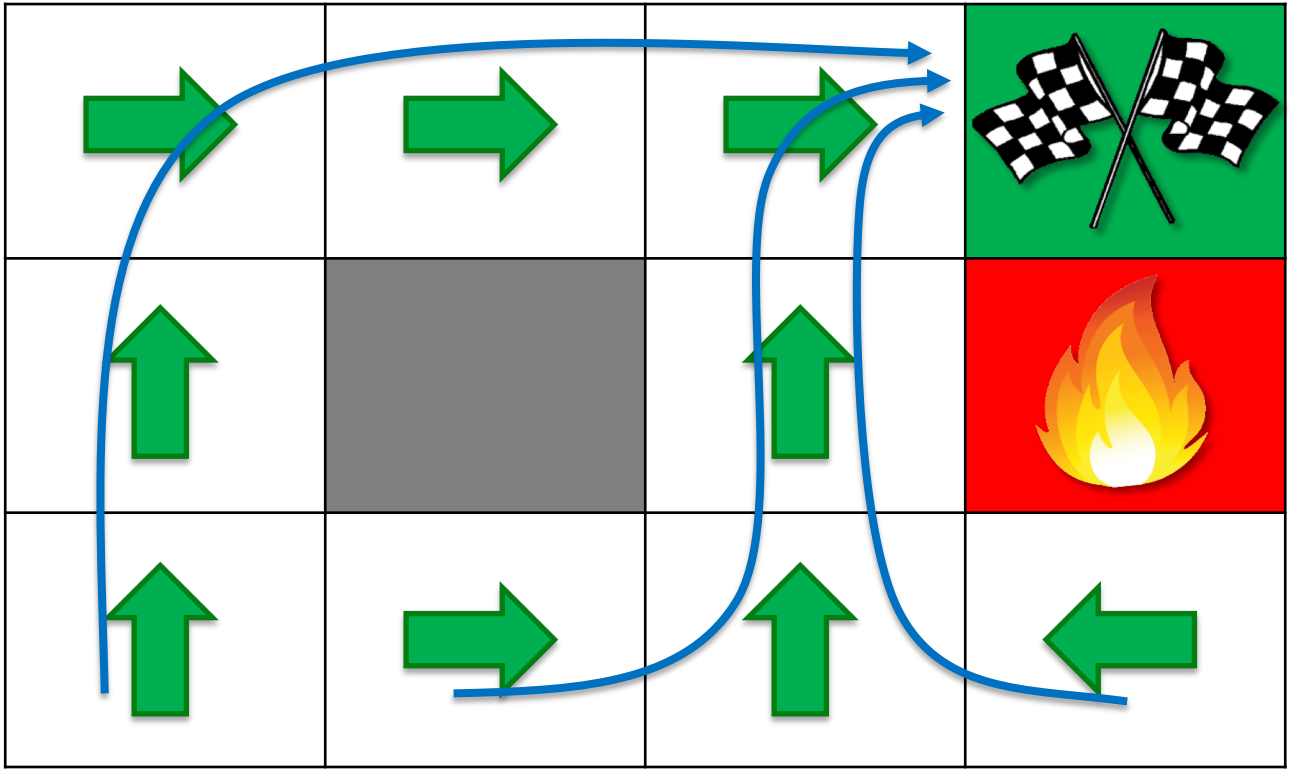
© SuperDataScience & IA Expert

Política x Plano



© SuperDataScience & IA Expert

Política x Plano



© SuperDataScience & IA Expert

Política x Plano

V=0.71	V=0.74	V=0.86	
V=0.63		V=0.39	
V=0.55	V=0.46	V=0.36	V=0.22

© SuperDataScience & IA Expert

Política x Plano

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

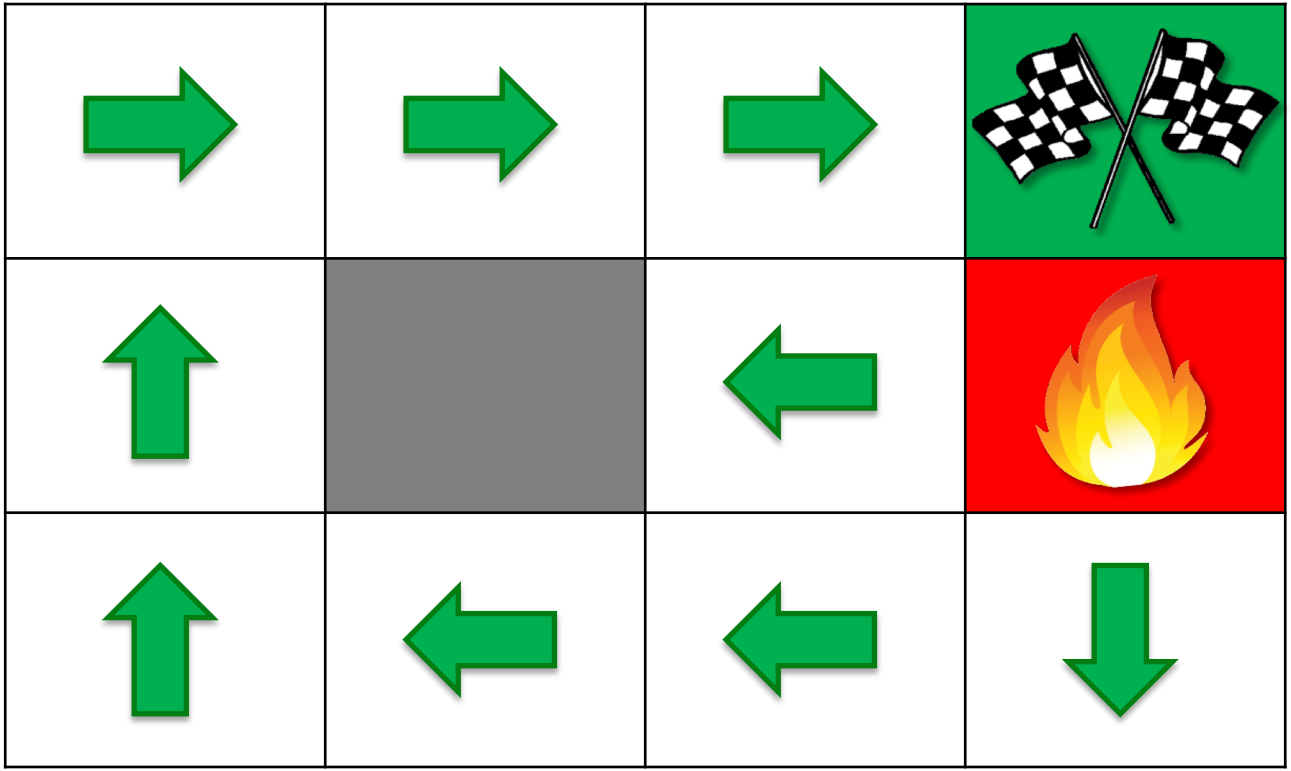
© SuperDataScience & IA Expert

Política x Plano

V=0.71	V=0.74	V=0.86	
V=0.63		V=0.39	
V=0.55	V=0.46	V=0.36	V=0.22

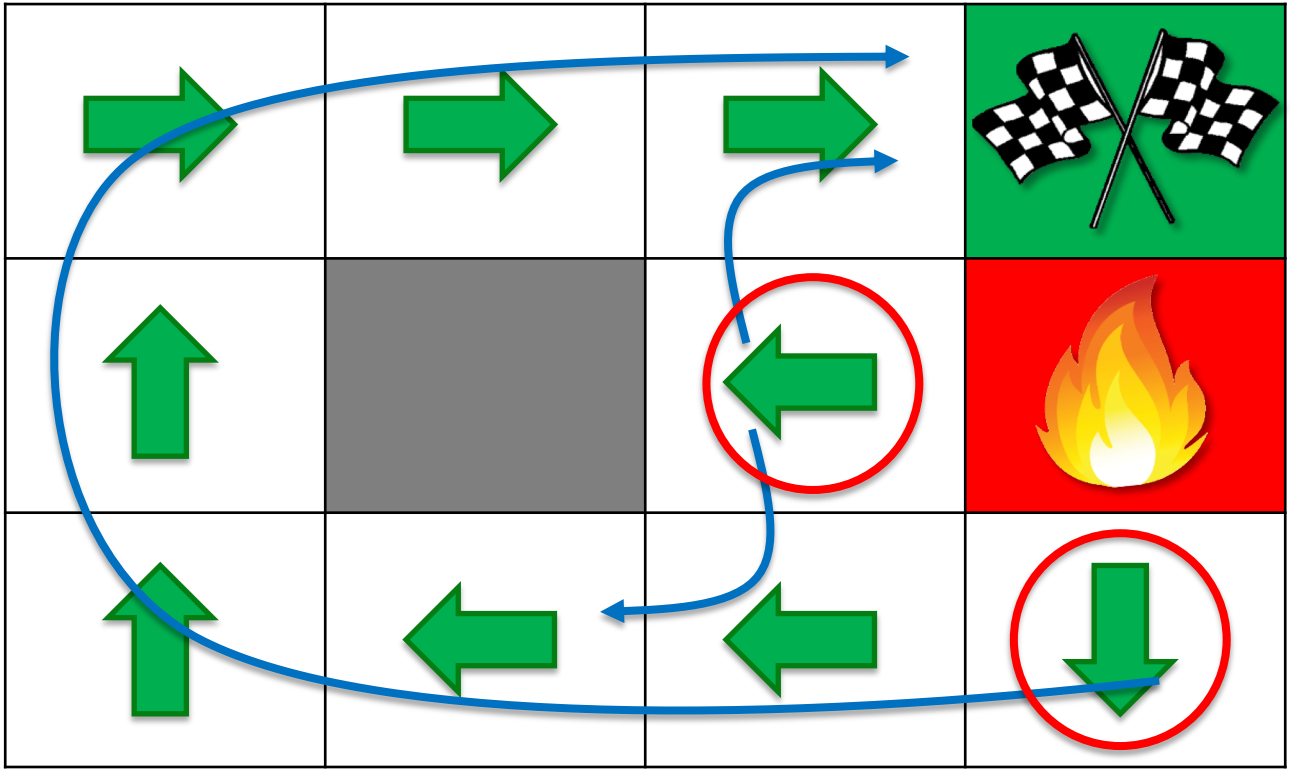
© SuperDataScience & IA Expert

Política x Plano



© SuperDataScience & IA Expert

Política x Plano



© SuperDataScience & IA Expert

Adição de Penalidades – "Living Penalty"

© SuperDataScience & IA Expert

“Living Penalty”

$$V(s) = \max_{a}\left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

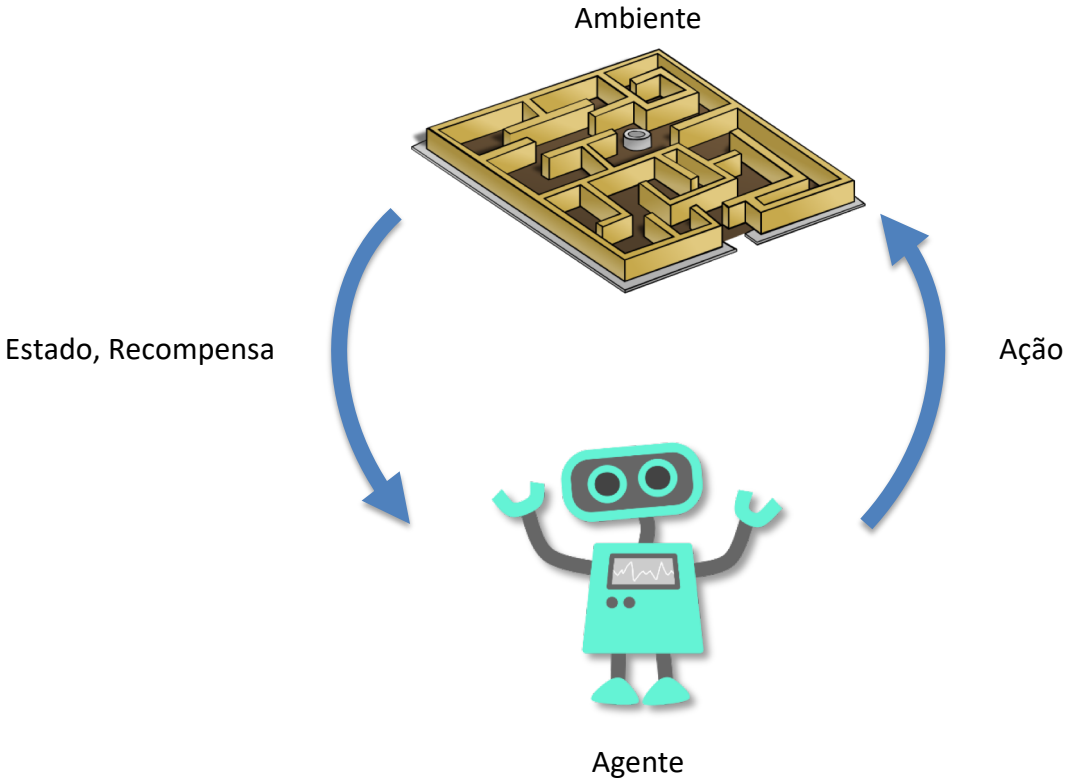
© SuperDataScience & IA Expert

"Living Penalty"

$$V(s) = \max_{a}\left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

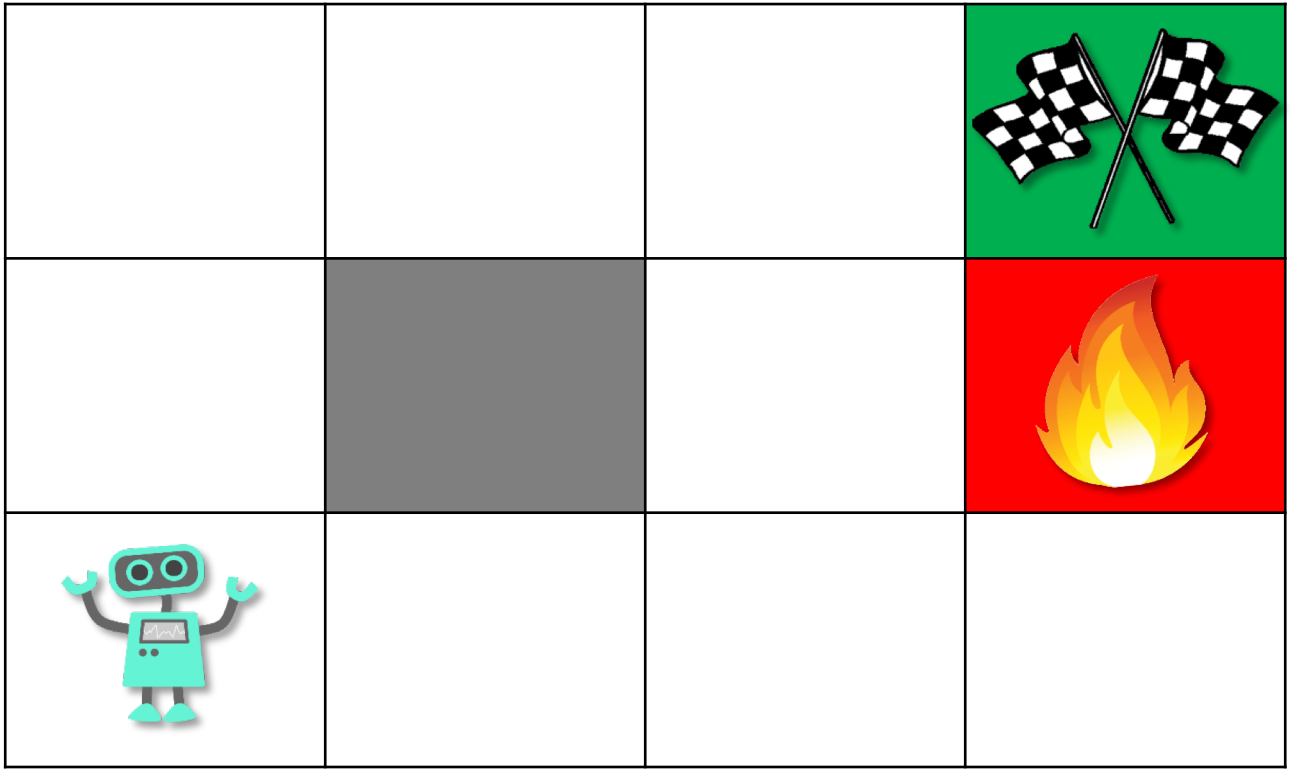
© SuperDataScience & IA Expert

“Living Penalty”



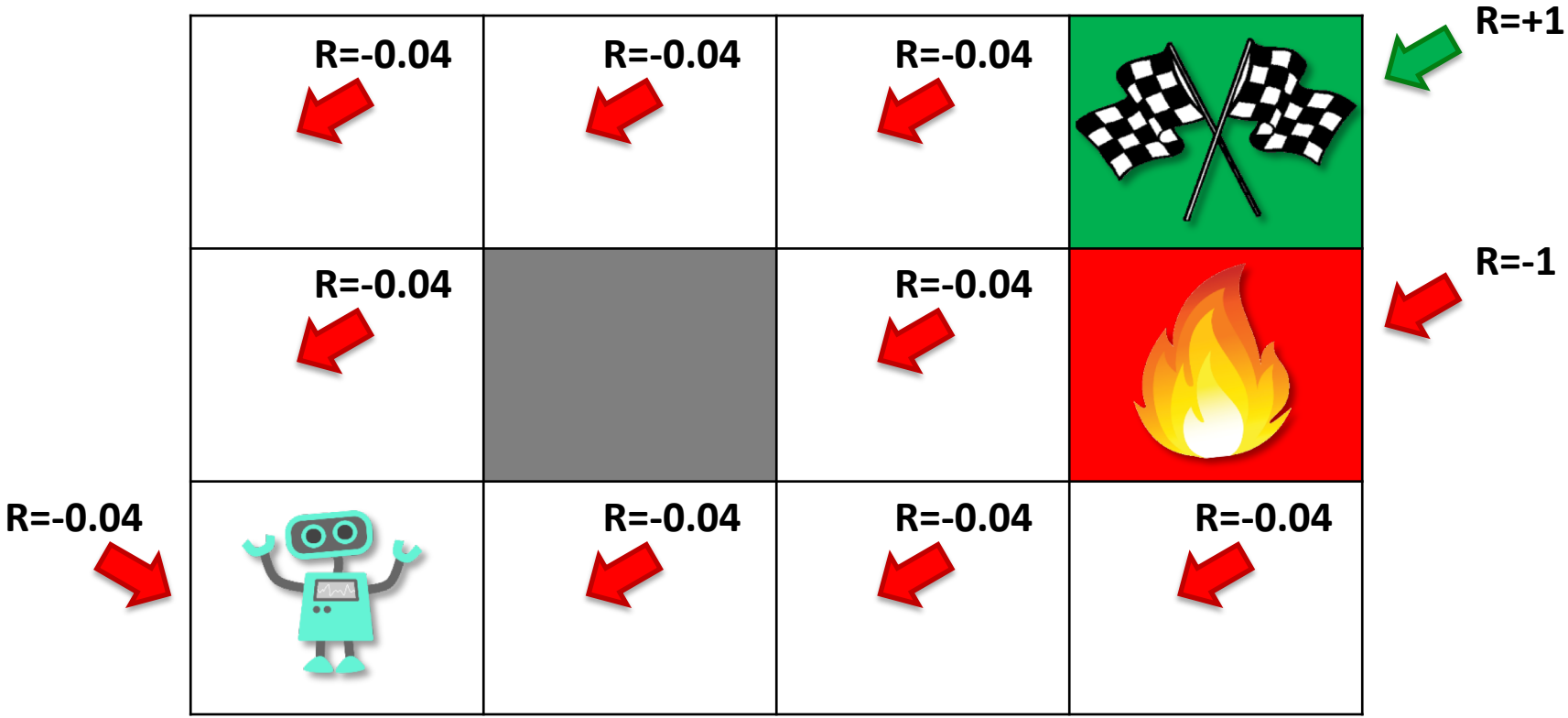
© SuperDataScience & IA Expert

“Living Penalty”



© SuperDataScience & IA Expert

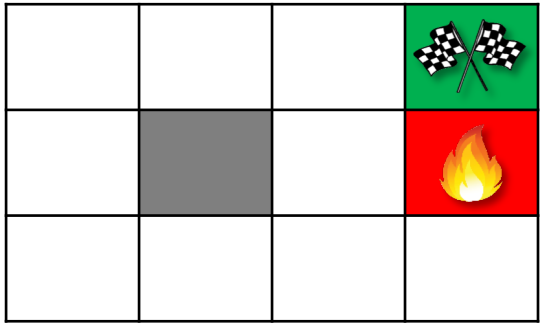
“Living Penalty”



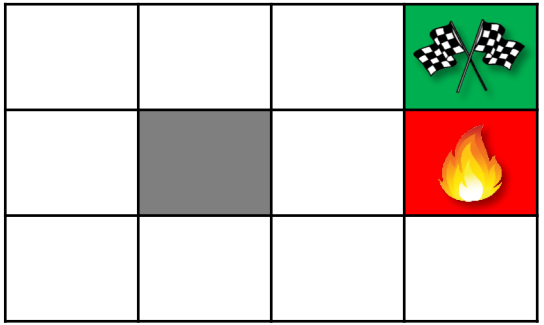
© SuperDataScience & IA Expert

“Living Penalty”

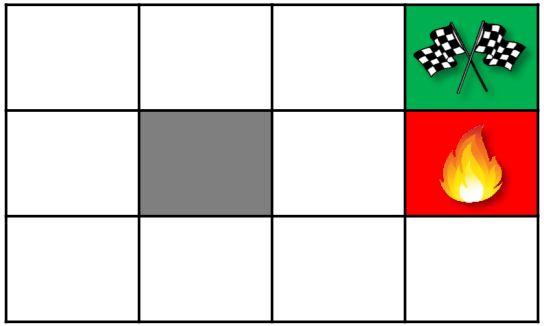
R(s)=0



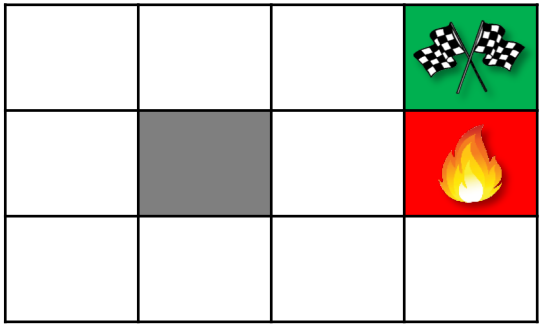
R(s)=-0.04



R(s)=-0.5



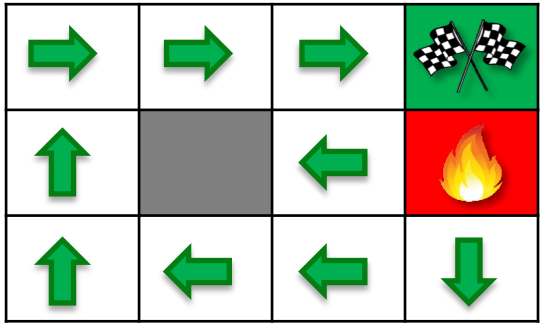
R(s)=-2.0



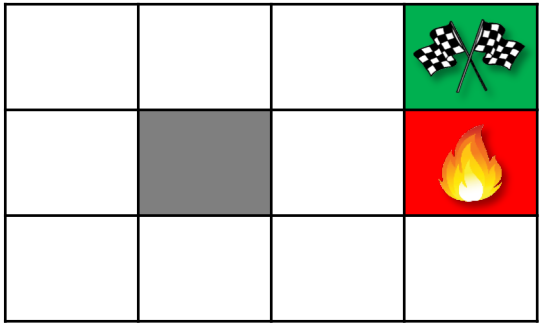
© SuperDataScience & IA Expert

“Living Penalty”

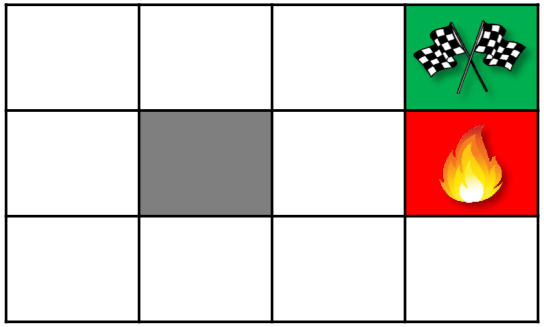
R(s)=0



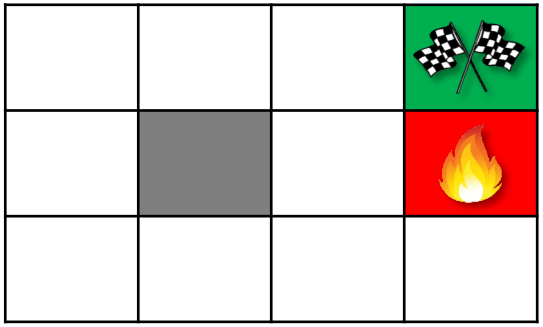
R(s)=-0.04



R(s)=-0.5



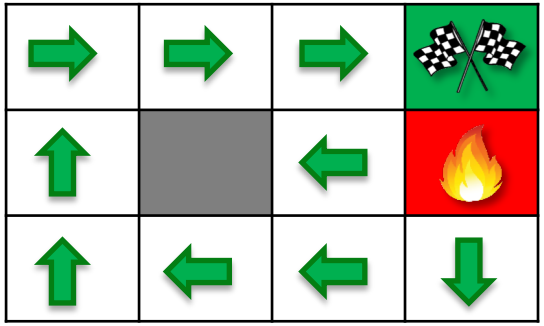
R(s)=-2.0



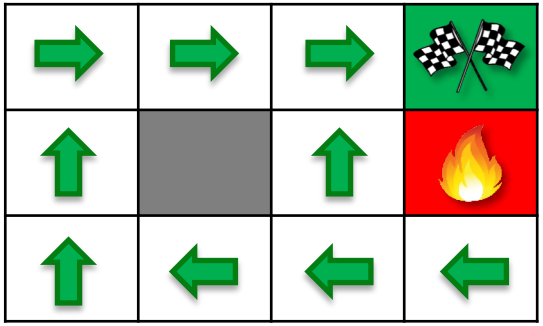
© SuperDataScience & IA Expert

“Living Penalty”

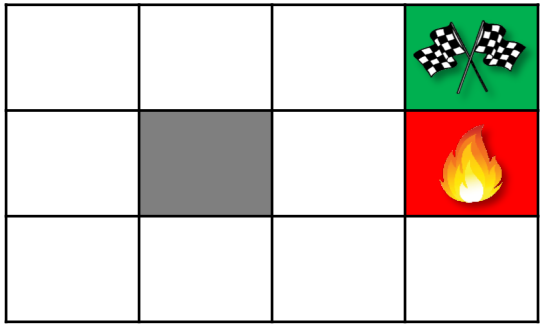
R(s)=0



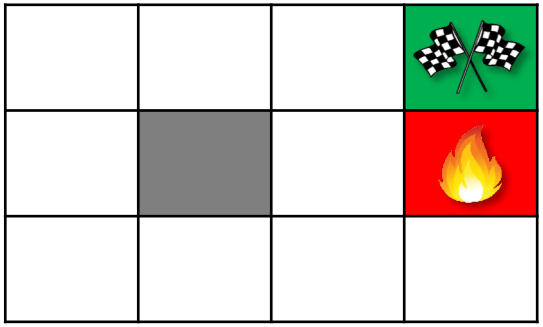
R(s)=-0.04



R(s)=-0.5



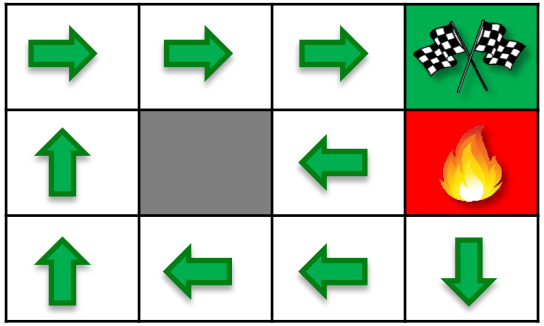
R(s)=-2.0



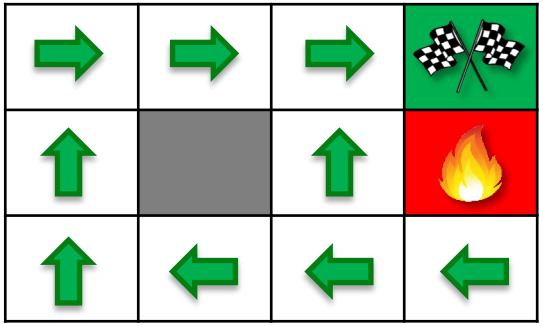
© SuperDataScience & IA Expert

“Living Penalty”

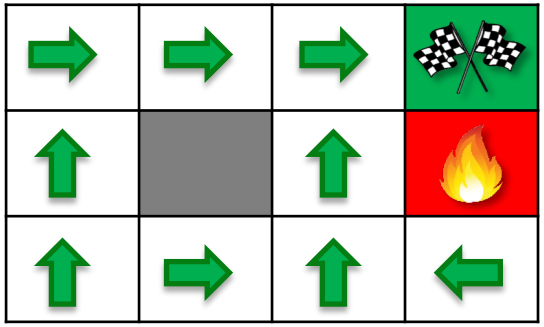
R(s)=0



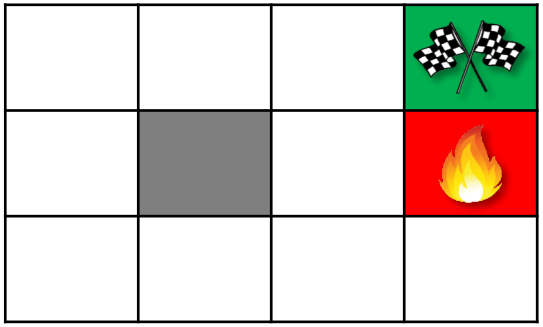
R(s)=-0.04



R(s)=-0.5



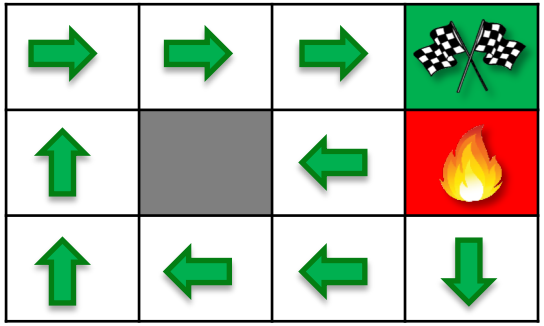
R(s)=-2.0



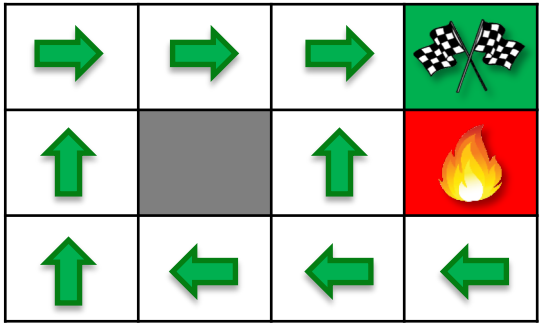
© SuperDataScience & IA Expert

“Living Penalty”

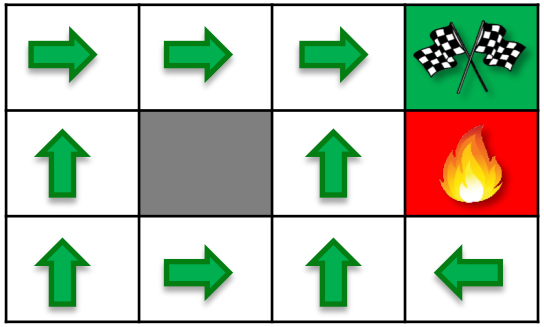
R(s)=0



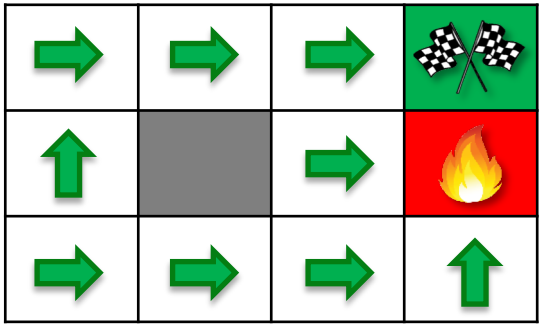
R(s)=-0.04



R(s)=-0.5



R(s)=-2.0



© SuperDataScience & IA Expert

Q-Learning Intuição

© SuperDataScience & IA Expert

Q-Learning Intuição

$$V(s) = \max_{a} \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

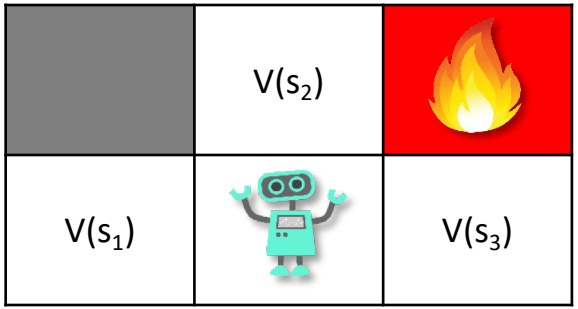
© SuperDataScience & IA Expert

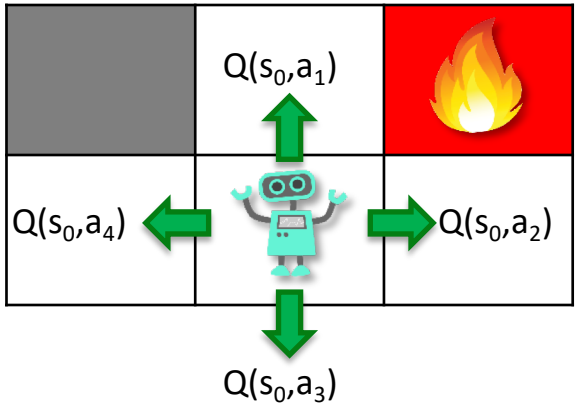
Q-Learning Intuição

Onde está o Q?

© SuperDataScience & IA Expert

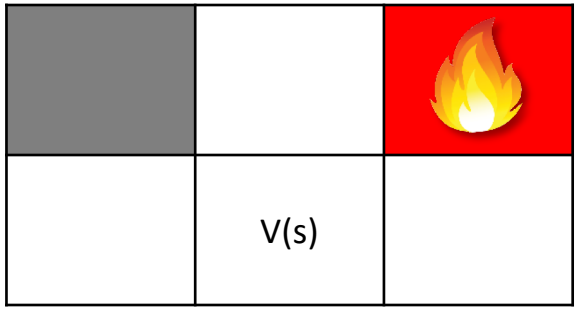
Q-Learning Intuição

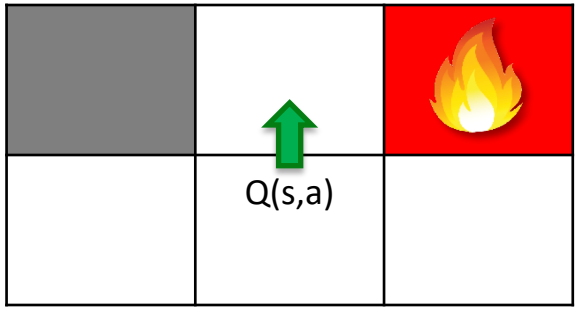




© SuperDataScience & IA Expert

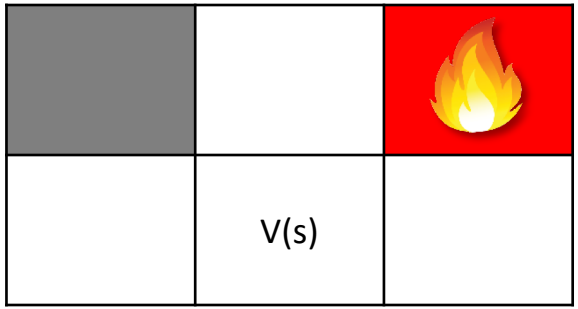
Q-Learning Intuição

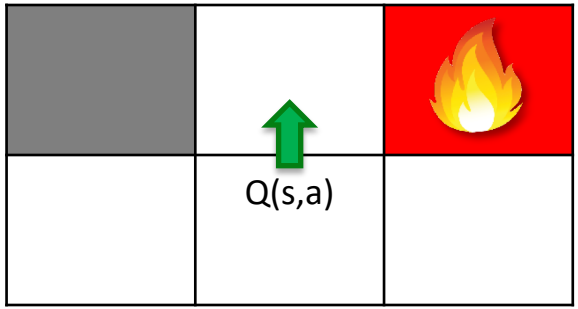




© SuperDataScience & IA Expert

Q-Learning Intuição

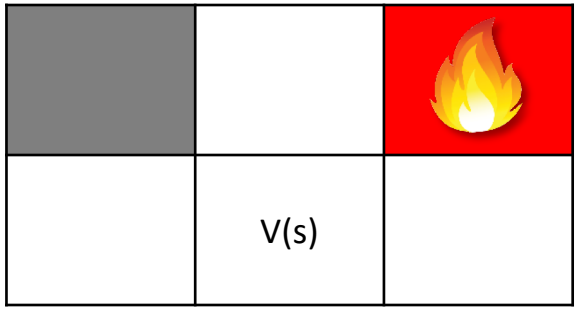




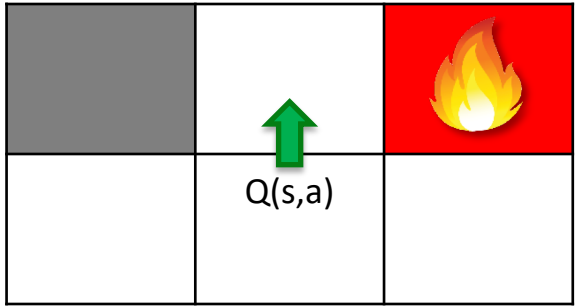
$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Q-Learning Intuição



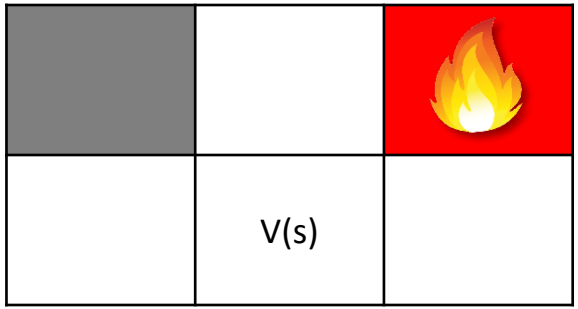
$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$



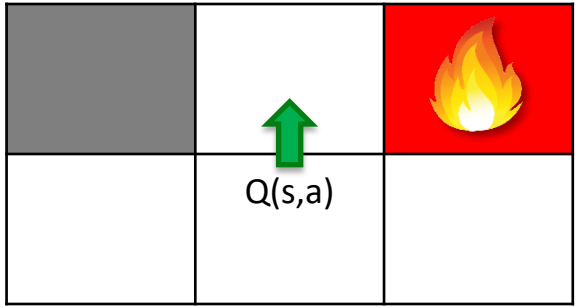
$$Q(s,a) =$$

© SuperDataScience & IA Expert

Q-Learning Intuição



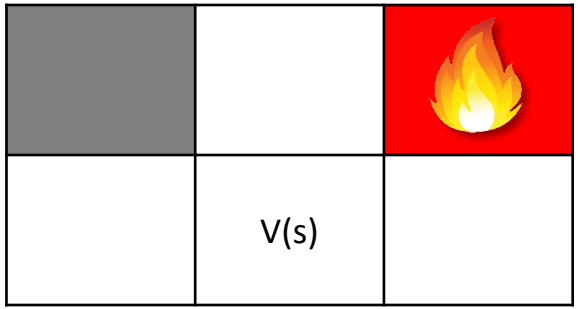
$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$



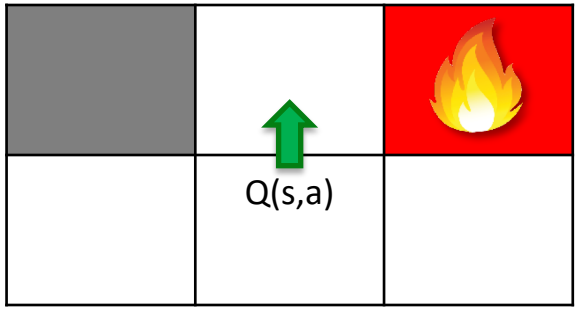
$$Q(s,a) = R(s,a) +$$

© SuperDataScience & IA Expert

Q-Learning Intuição



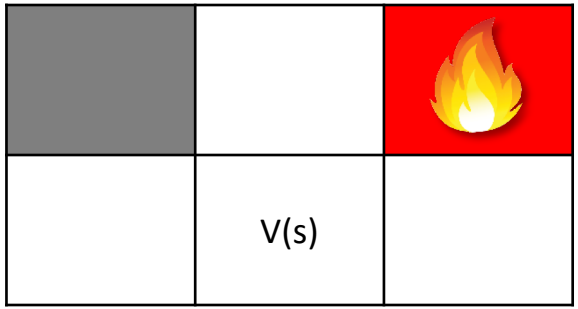
$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

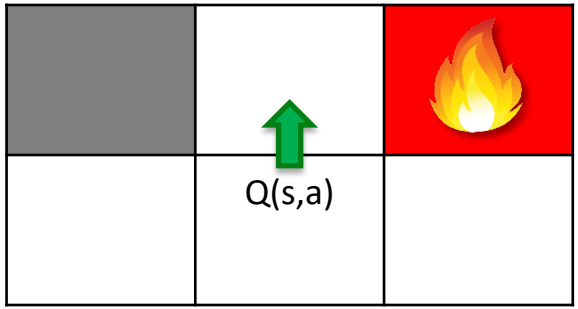


$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Q-Learning Intuição



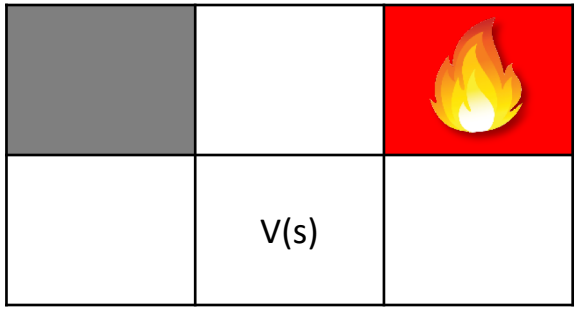


$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

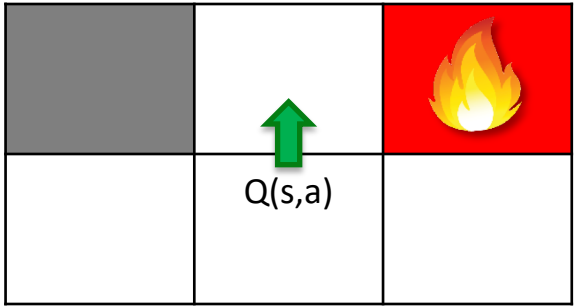
$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Q-Learning Intuição



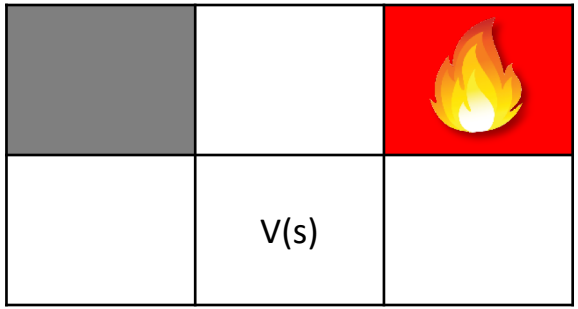
$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

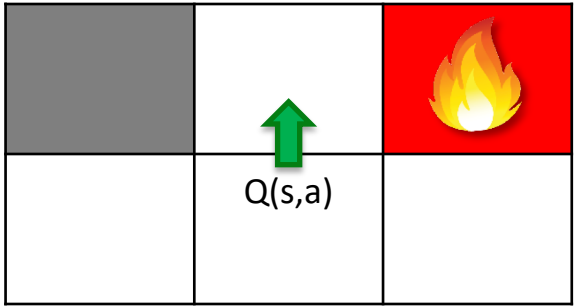


$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Q-Learning Intuição



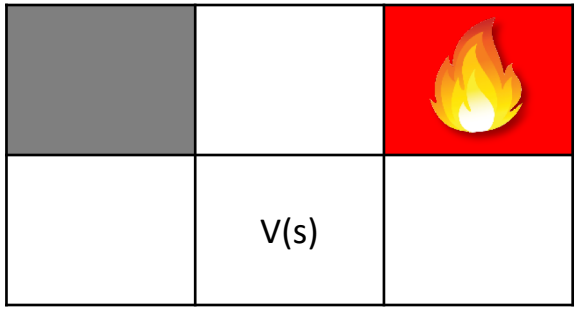


$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

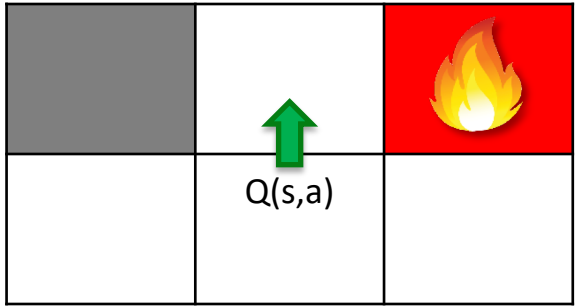
$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s')V(s') \right)$$

© SuperDataScience & IA Expert

Q-Learning Intuição



$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s') V(s') \right)$$



$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s') \max_{a'} Q(s',a') \right)$$

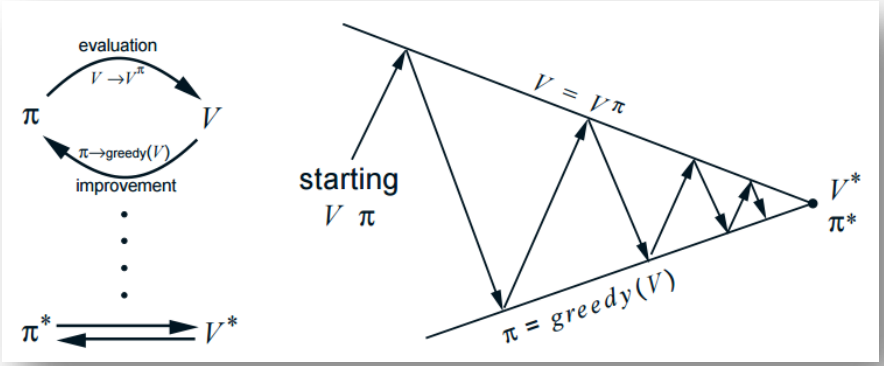
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Markov Decision Processes: Concepts and Algorithms

Martijn van Otterlo (2009)



Link:

https://pdfs.semanticscholar.org/968b/ab782e52faf0f7957ca0f38b9e9078454afe.pdf

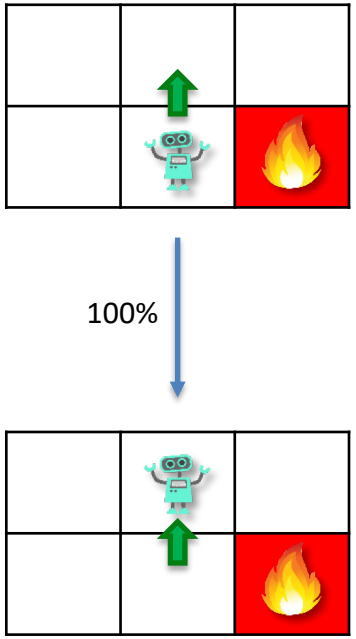
© SuperDataScience & IA Expert

Diferença Temporal

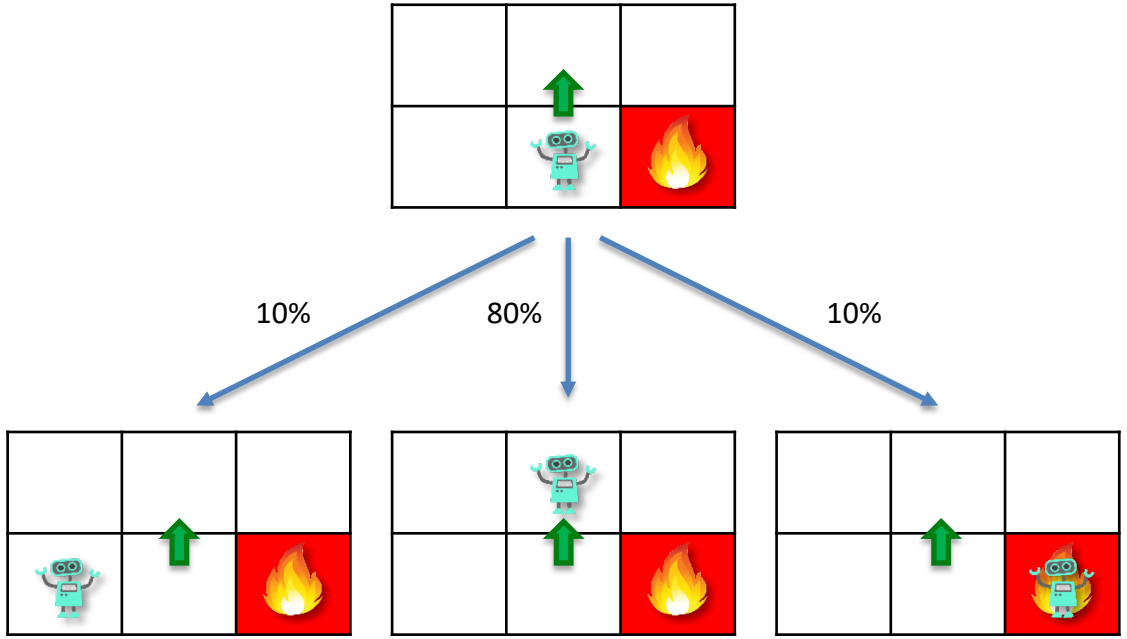
© SuperDataScience & IA Expert

Diferença Temporal

Busca Determinística



Busca Não-Determinística



© SuperDataScience & IA Expert

Diferença Temporal

V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

© SuperDataScience & IA Expert

Diferença Temporal

V=0.71	V=0.74	V=0.86	
V=0.63		V=0.39	
V=0.55	V=0.46	V=0.36	V=0.22

© SuperDataScience & IA Expert

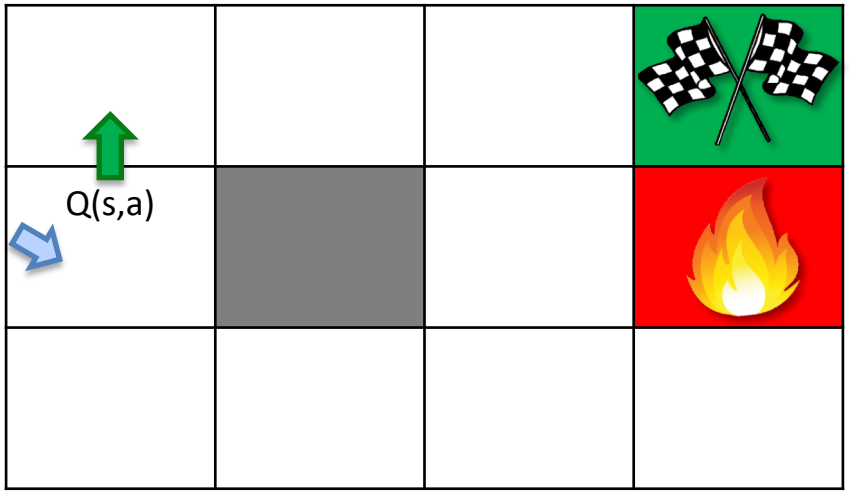
Diferença Temporal

$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s') \max_{a'} Q(s',a') \right)$$

$$Q(s,a) = R(s,a) + \gamma \max_{a'} Q(s',a')$$

© SuperDataScience & IA Expert

Diferença Temporal



Antes:

$$Q(s,a)$$

Depois:

$$R(s,a) + \gamma \max_{a'} Q(s',a')$$

$$TD(a,s) = \overbrace{R(s,a) + \gamma \max_{a'} Q(s',a')}^{} - \overbrace{Q(s,a)}^{}$$

© SuperDataScience & IA Expert

Diferença Temporal

$$TD(a,s) = R(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)$$

$$Q(s,a) = Q(s,a) + \alpha TD(a,s)$$

© SuperDataScience & IA Expert

Diferença Temporal

$$TD(a,s) = R(s,a) + \gamma \max_{a'} Q(s',a') - Q_{t-1}(s,a)$$

$$Q_t(s,a) = Q_{t-1}(s,a) + \alpha TD_t(a,s)$$

$$Q_t(s,a) = Q_{t-1}(s,a) + \alpha \left(R(s,a) + \gamma \max_{a'} Q(s',a') - Q_{t-1}(s,a) \right)$$

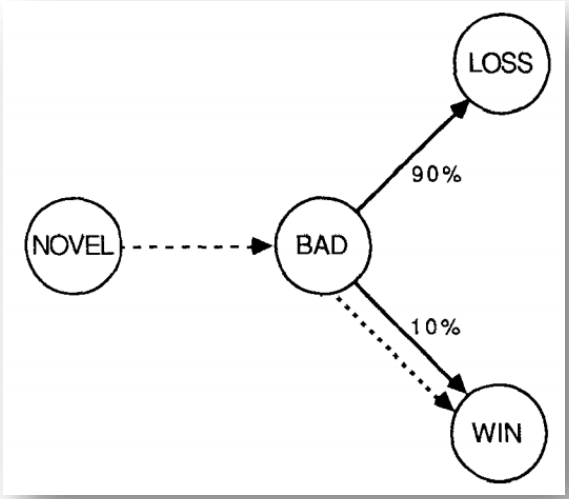
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Learning to Predict by the Methods of Temporal Differences

Richard Sutton (1988)



Link:

https://link.springer.com/article/10.1007/BF00115009

© SuperDataScience & IA Expert

Conteúdo

© SuperDataScience & IA Expert

Conteúdo

O que você aprenderá nessa seção:

- Intuição Deep Q-Learning (Aprendizagem)
- Intuição Deep Q-Learning (Ação)
- Experiência de Replay
- Políticas de Seleção de Ações

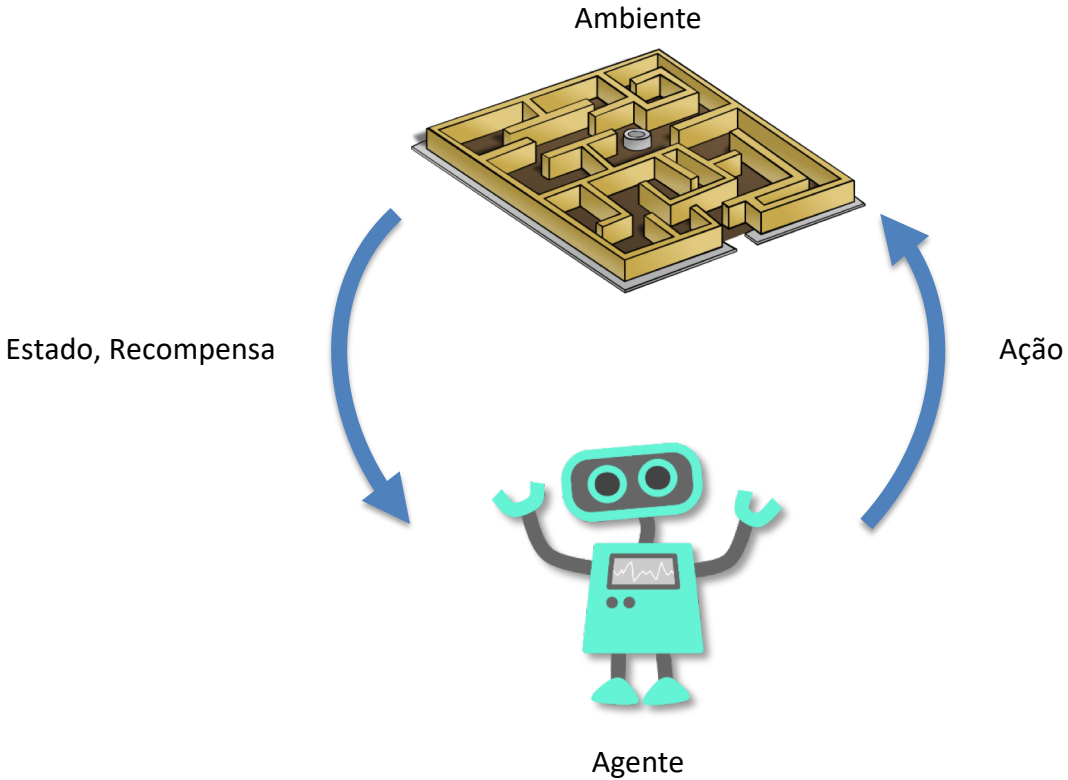
- Anexo 1: Redes Neurais Artificiais

© SuperDataScience & IA Expert

Intuição Deep Q-Learning

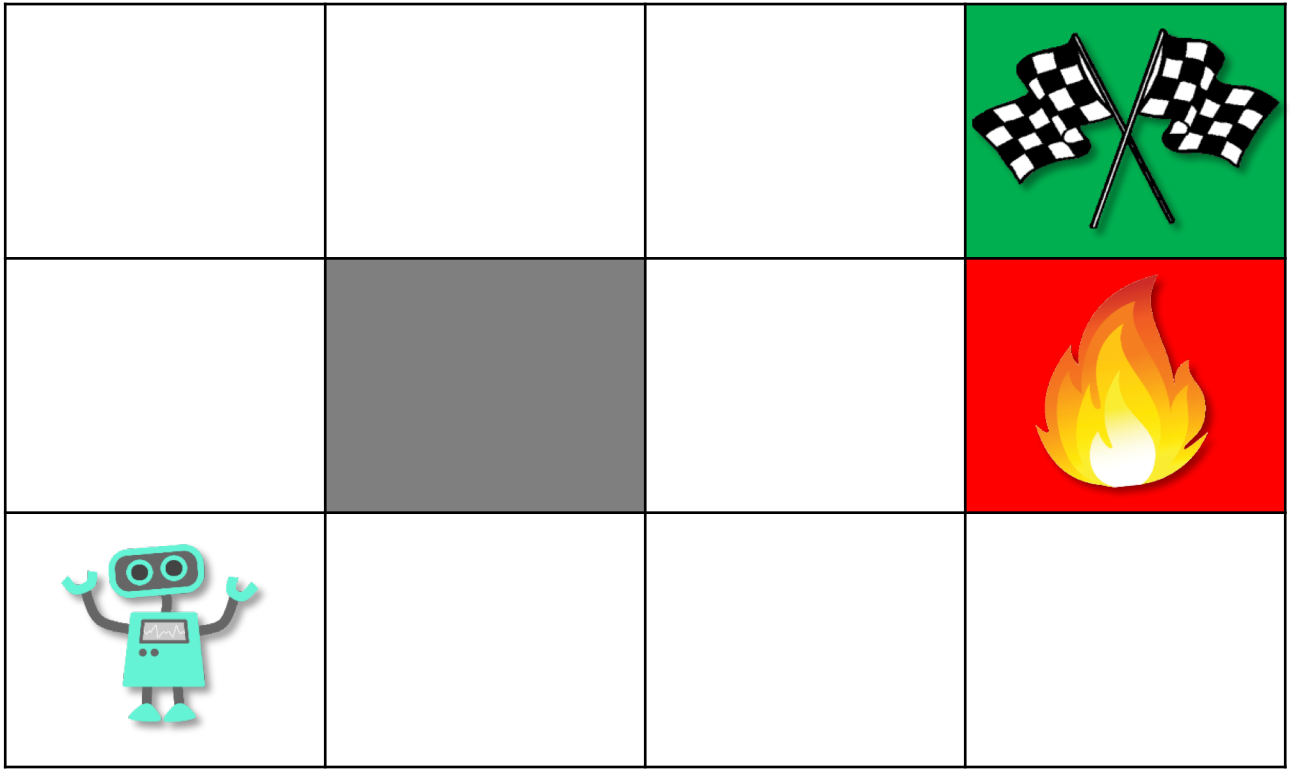
© SuperDataScience & IA Expert

Intuição Deep Q-Learning



© SuperDataScience & IA Expert

Intuição Deep Q-Learning



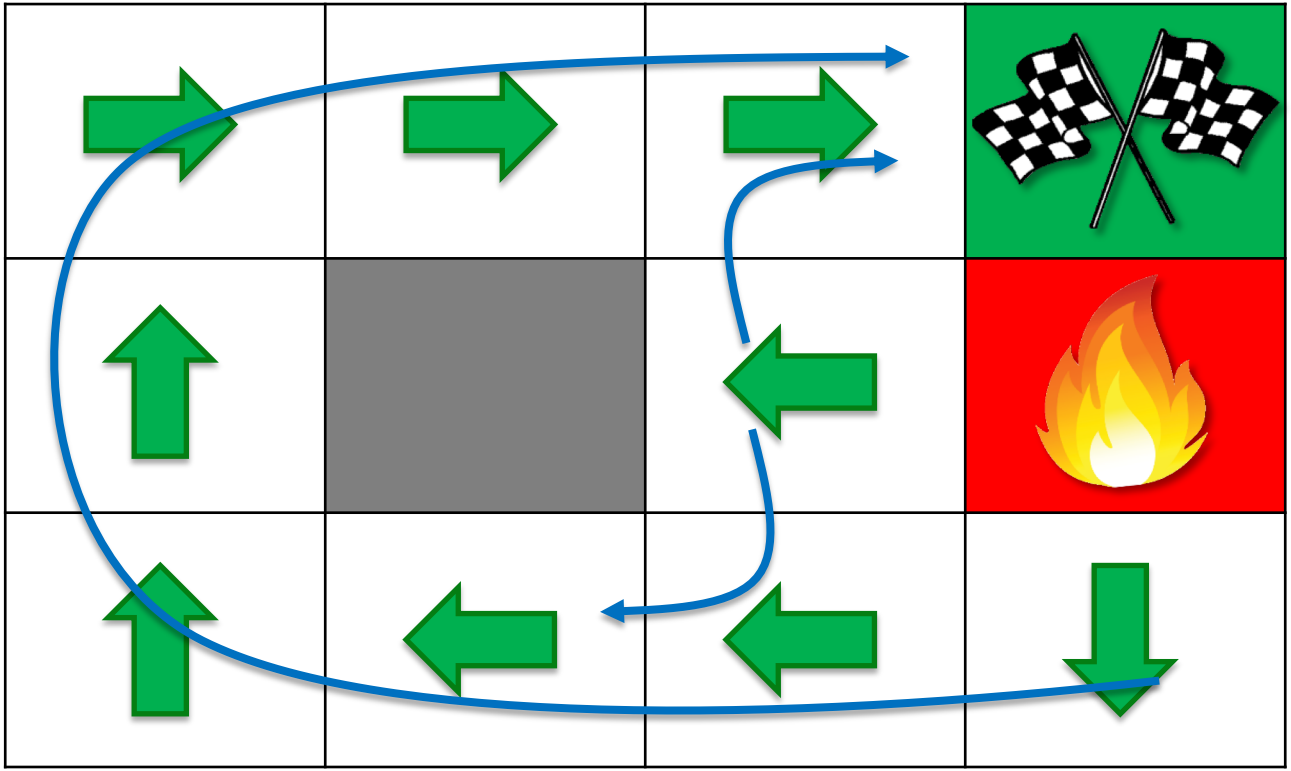
© SuperDataScience & IA Expert

Intuição Deep Q-Learning

V=0.71	V=0.74	V=0.86	
V=0.63		V=0.39	
V=0.55	V=0.46	V=0.36	V=0.22

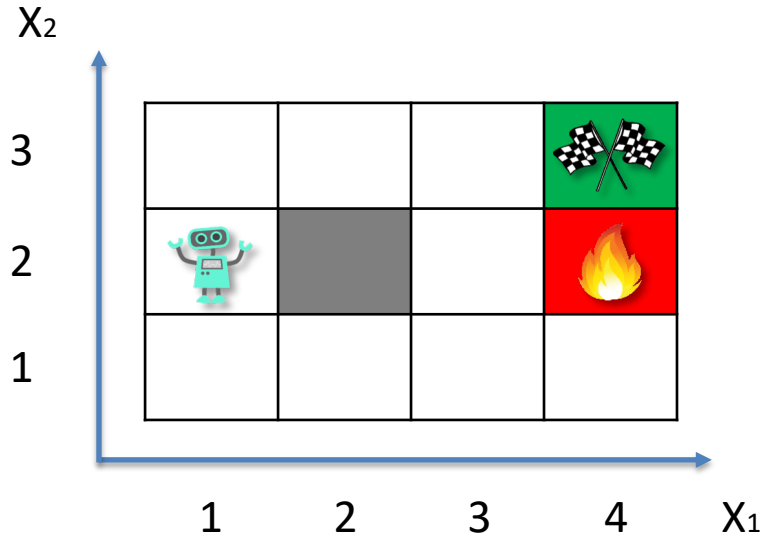
© SuperDataScience & IA Expert

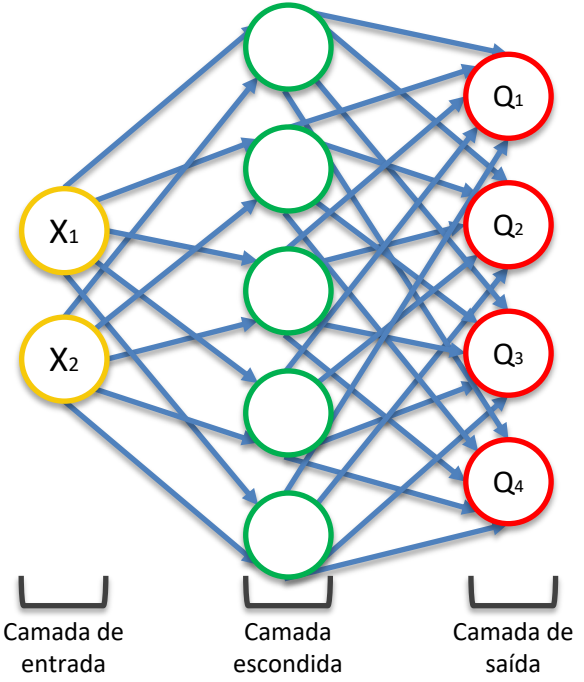
Intuição Deep Q-Learning



© SuperDataScience & IA Expert

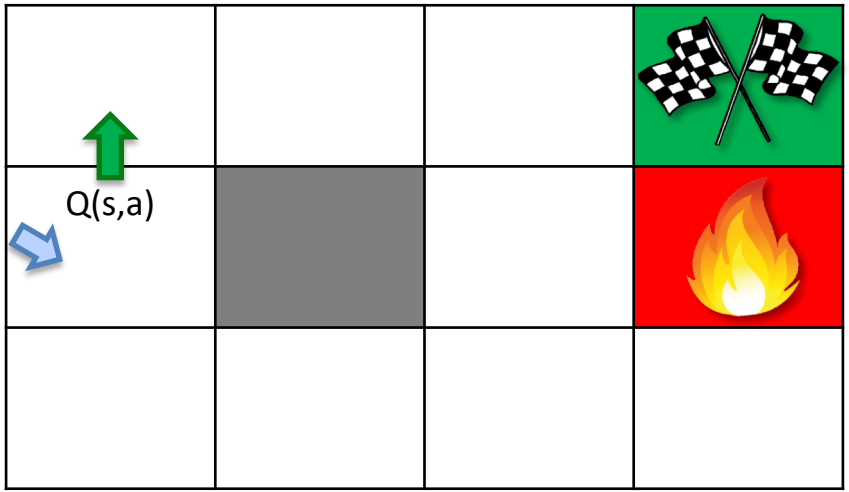
Intuição Deep Q-Learning





© SuperDataScience & IA Expert

Intuição Deep Q-Learning



Antes:

$Q(s,a)$

Depois:

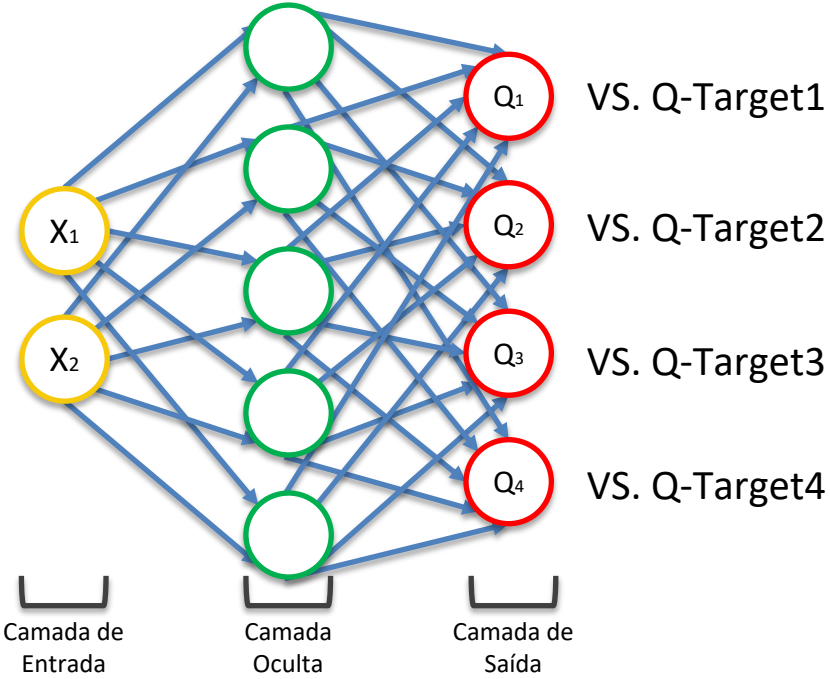
$R(s,a) + \gamma \max_{a'} Q(s',a')$

$$TD(a,s) = R(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)$$

© SuperDataScience & IA Expert

Intuição Deep Q-Learning

Aprendizagem:

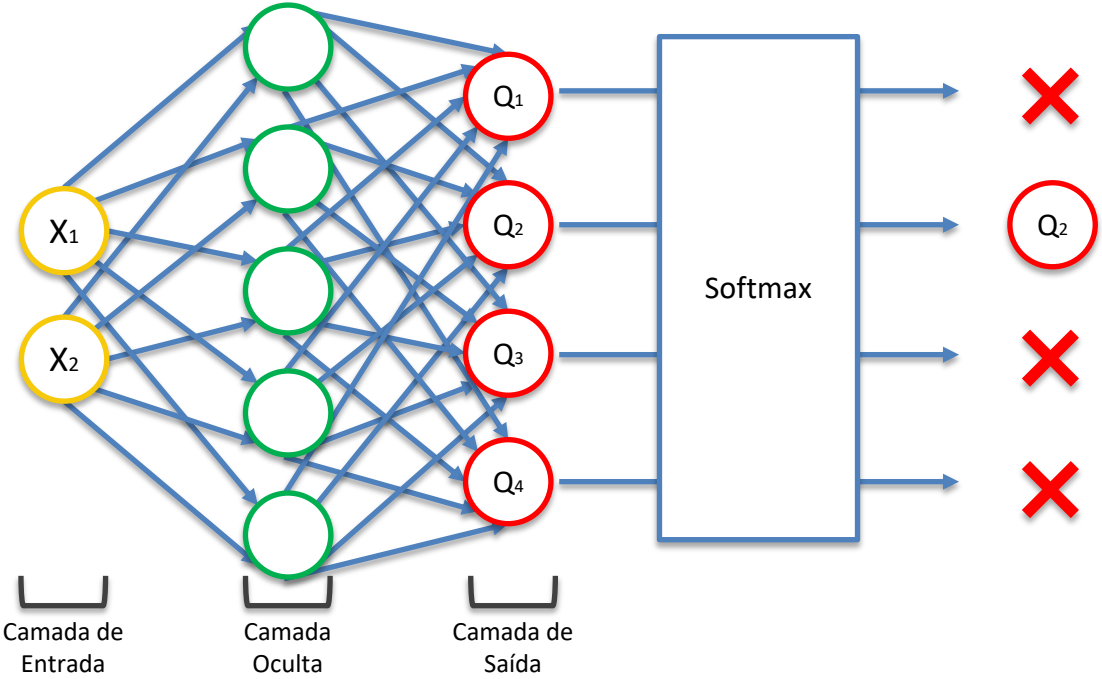


$$L = \sum (Q\text{-}Target\ - Q)^2$$

© SuperDataScience & IA Expert

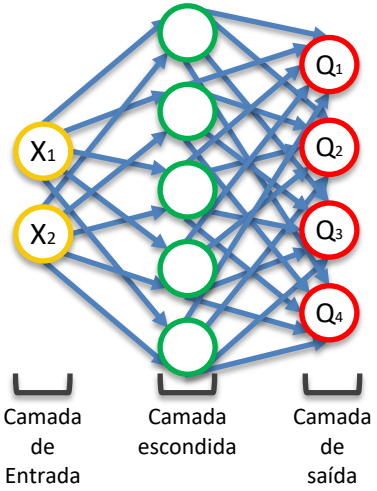
Intuição Deep Q-Learning

Ações:



© SuperDataScience & IA Expert

Intuição Deep Q-Learning



Aprendizagem:

VS. Q-Target1

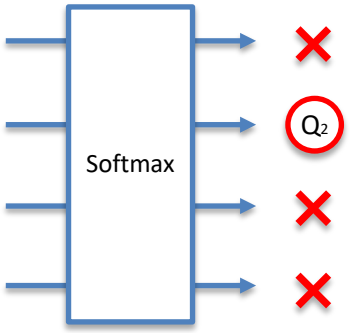
VS. Q-Target2

VS. Q-Target3

VS. Q-Target4

$$L = \sum (Q\text{–}Target\ - Q)^2$$

Ações:



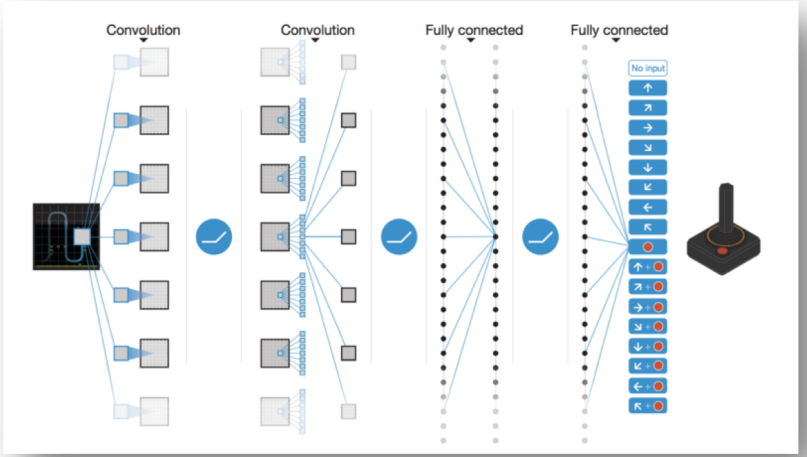
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Simple Reinforcement Learning with Tensorflow (Part 4)

Arthur Juliani (2016)



Link:

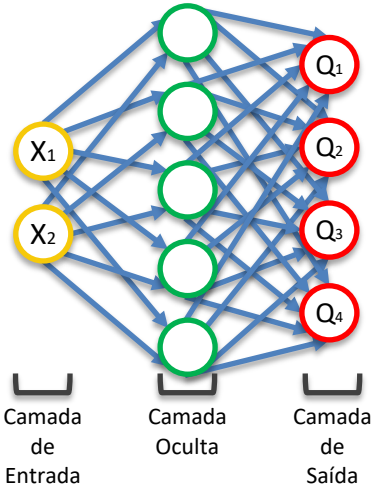
https://medium.com/@awjuliani/simple-reinforcement-learning-with-tensorflow-part-4-deep-q-networks-and-beyond-8438a3e2b8df

© SuperDataScience & IA Expert

Experiência de Replay

© SuperDataScience & IA Expert

Experiência de Replay



Aprendizagem:

VS. Q-Target1

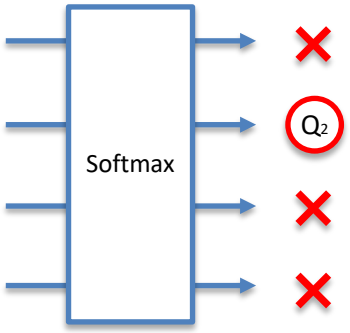
VS. Q-Target2

VS. Q-Target3

VS. Q-Target4

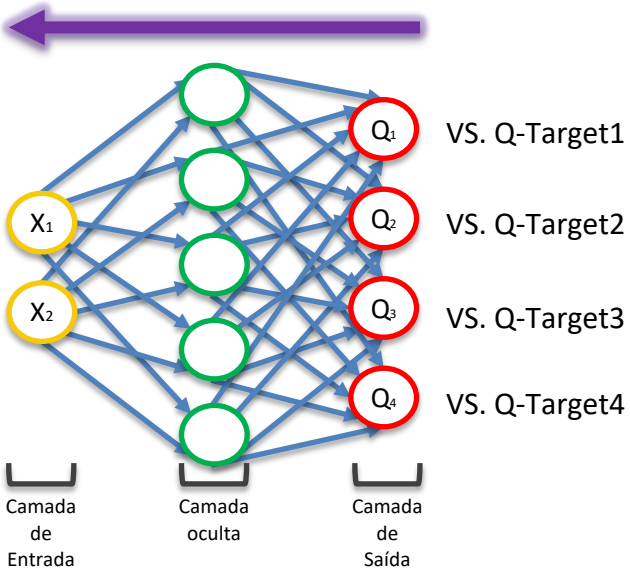
$$L = \sum (Q\text{–}Target\ - Q)^2$$

Ação:



© SuperDataScience & IA Expert

Experiência de Replay



$$L = \sum (Q\text{-}Target\ - Q)^2$$

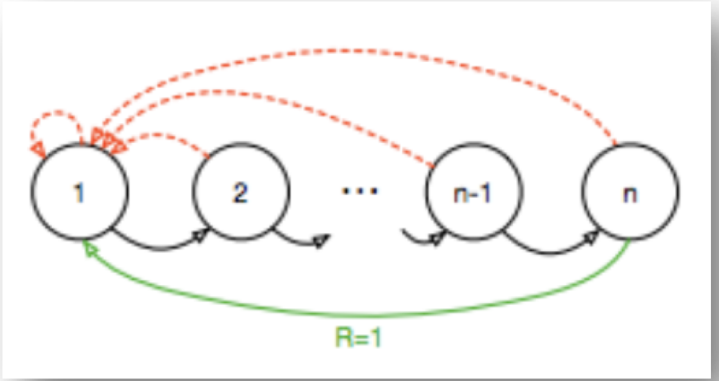
© SuperDataScience & IA Expert

Leitura Adicional

Leitura Adicional:

Prioritized Experience Replay

Tom Schaul et al.,
Google DeepMind (2016)



Link:

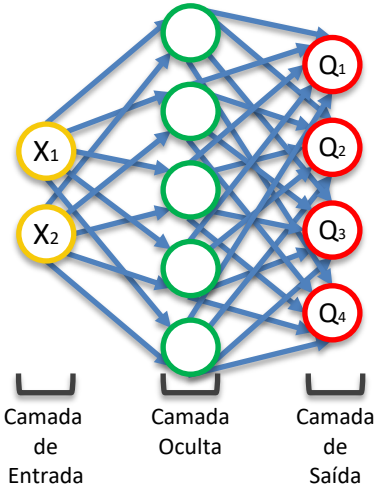
https://arxiv.org/pdf/1511.05952.pdf

© SuperDataScience & IA Expert

Política de Seleção de Ações

© SuperDataScience & IA Expert

Política de Seleção de Ações



Aprendizagem:

VS. Q-Target1

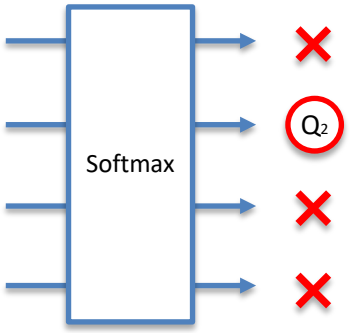
VS. Q-Target2

VS. Q-Target3

VS. Q-Target4

$$L = \sum (Q\text{–}Target\ - Q)^2$$

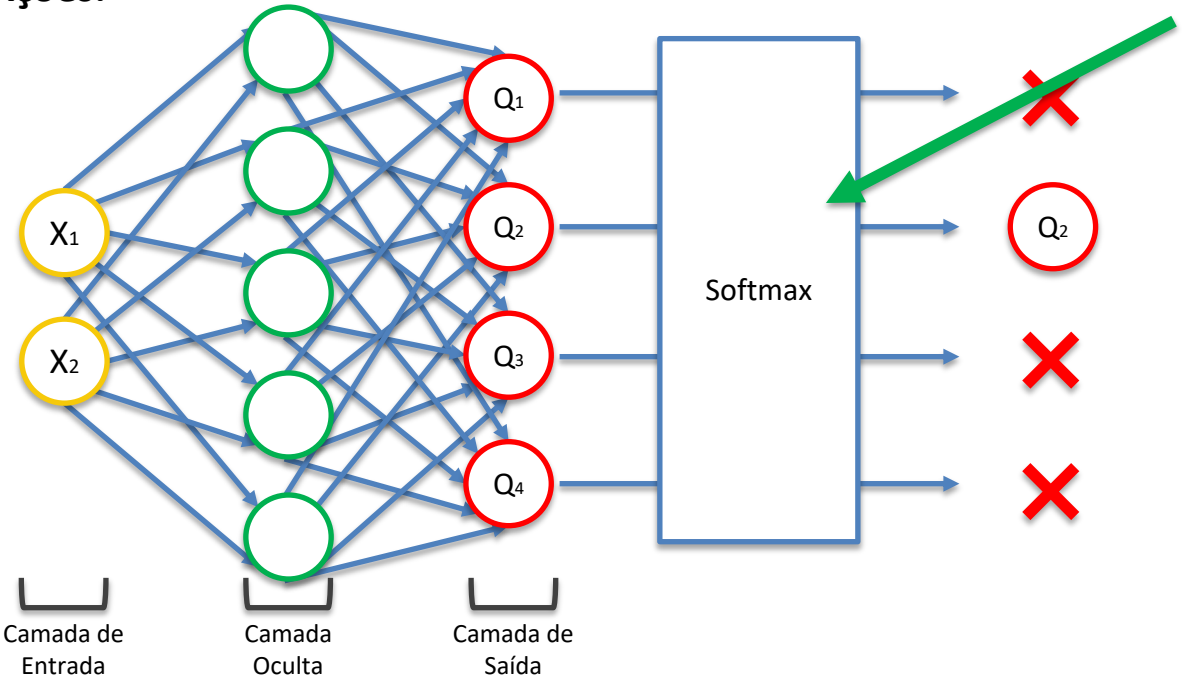
Ações:



© SuperDataScience & IA Expert

Política de Seleção de Ações

Ações:



Seleção de Ação:

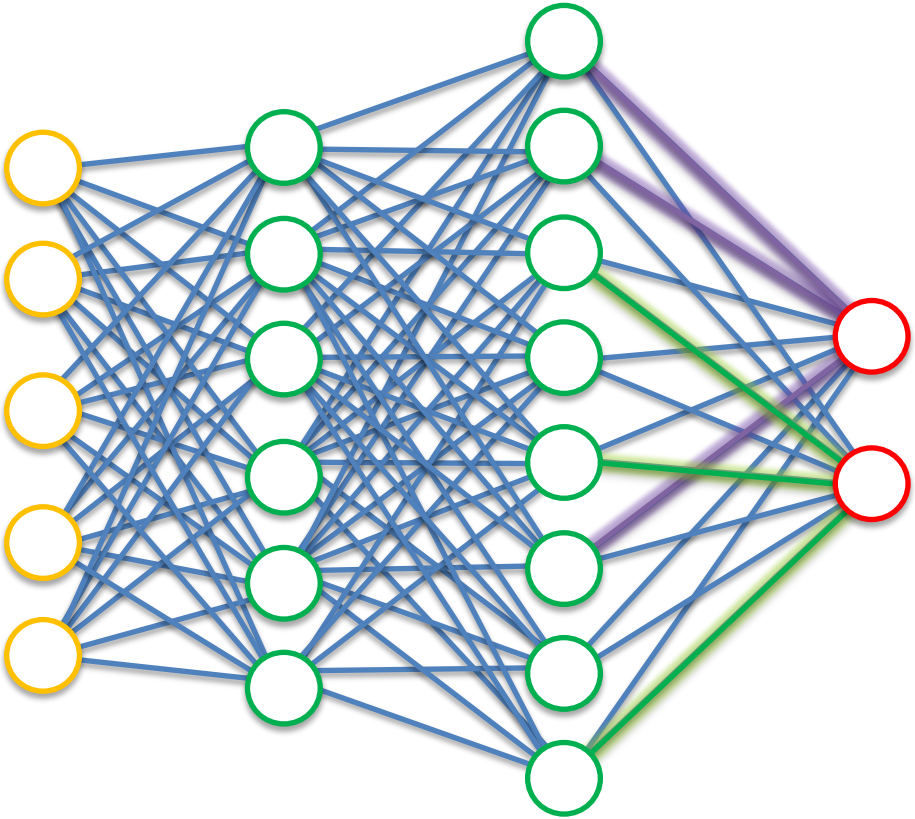
- ε-greedy
- ε-soft (1-ε)
- Softmax

Exploration
vs
Exploitation

© SuperDataScience & IA Expert

Política de Seleção de Ações

Flattening



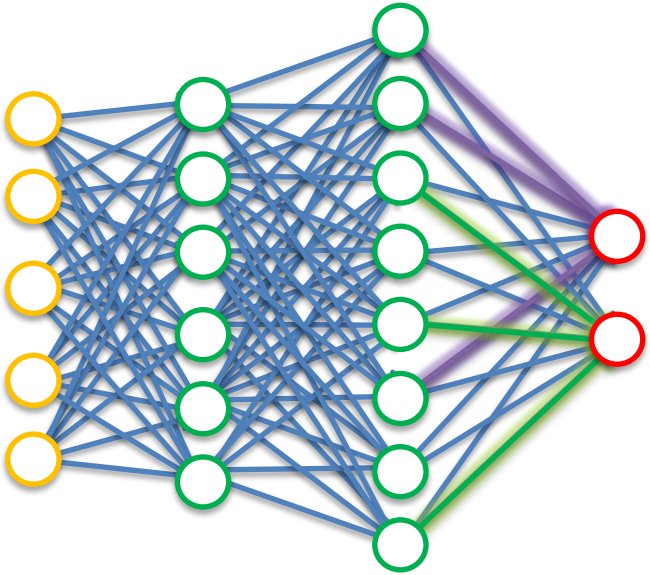
Cachorro 0.95

Gato 0.05

© SuperDataScience & IA Expert

Política de Seleção de Ações

Flattening



$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

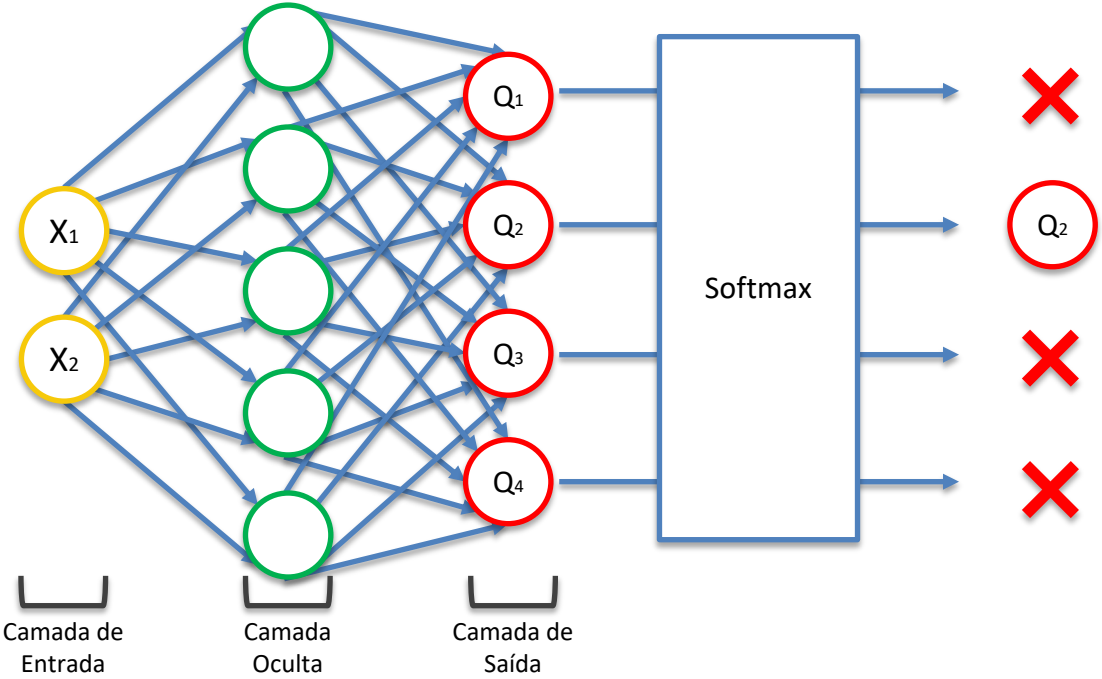
Cachorro → z_1 → 0.95

Gato → z_2 → 0.05

© SuperDataScience & IA Expert

Política de Seleção de Ações

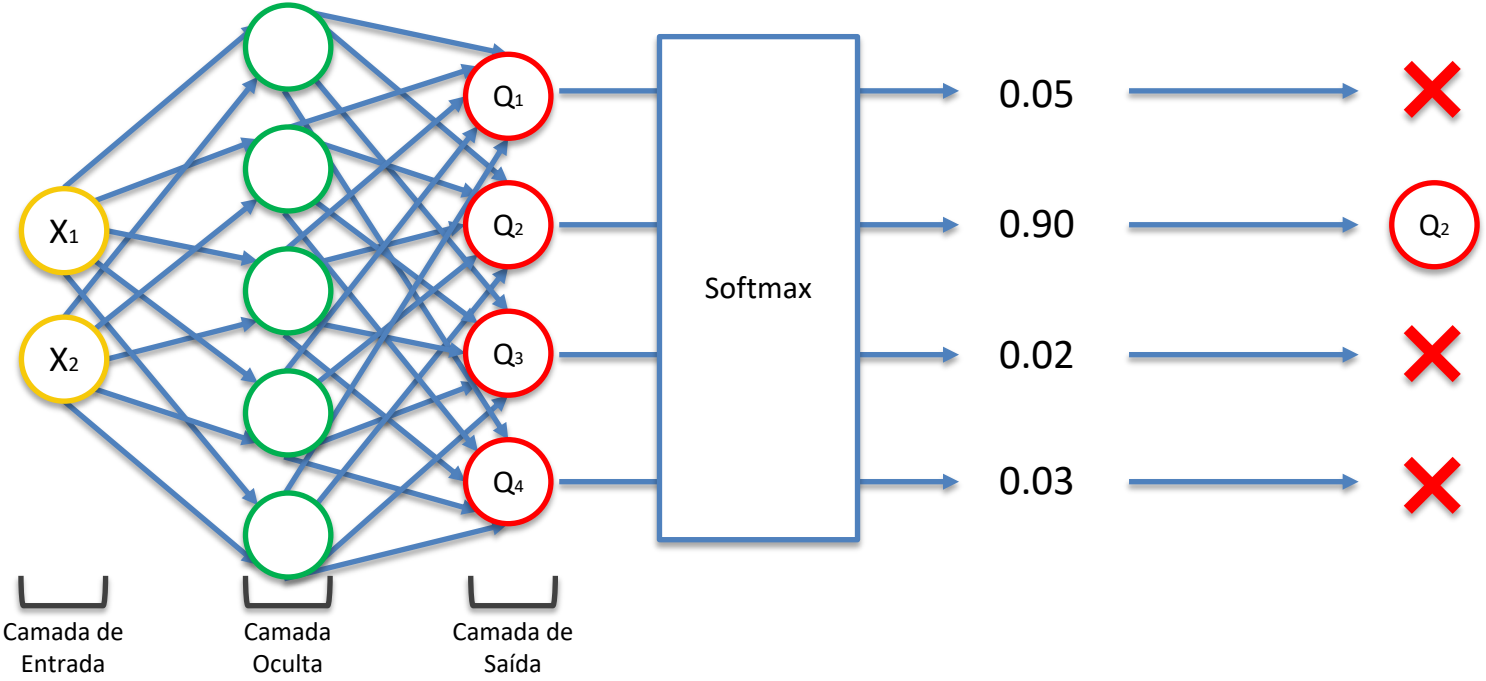
Ações:



© SuperDataScience & IA Expert

Política de Seleção de Ações

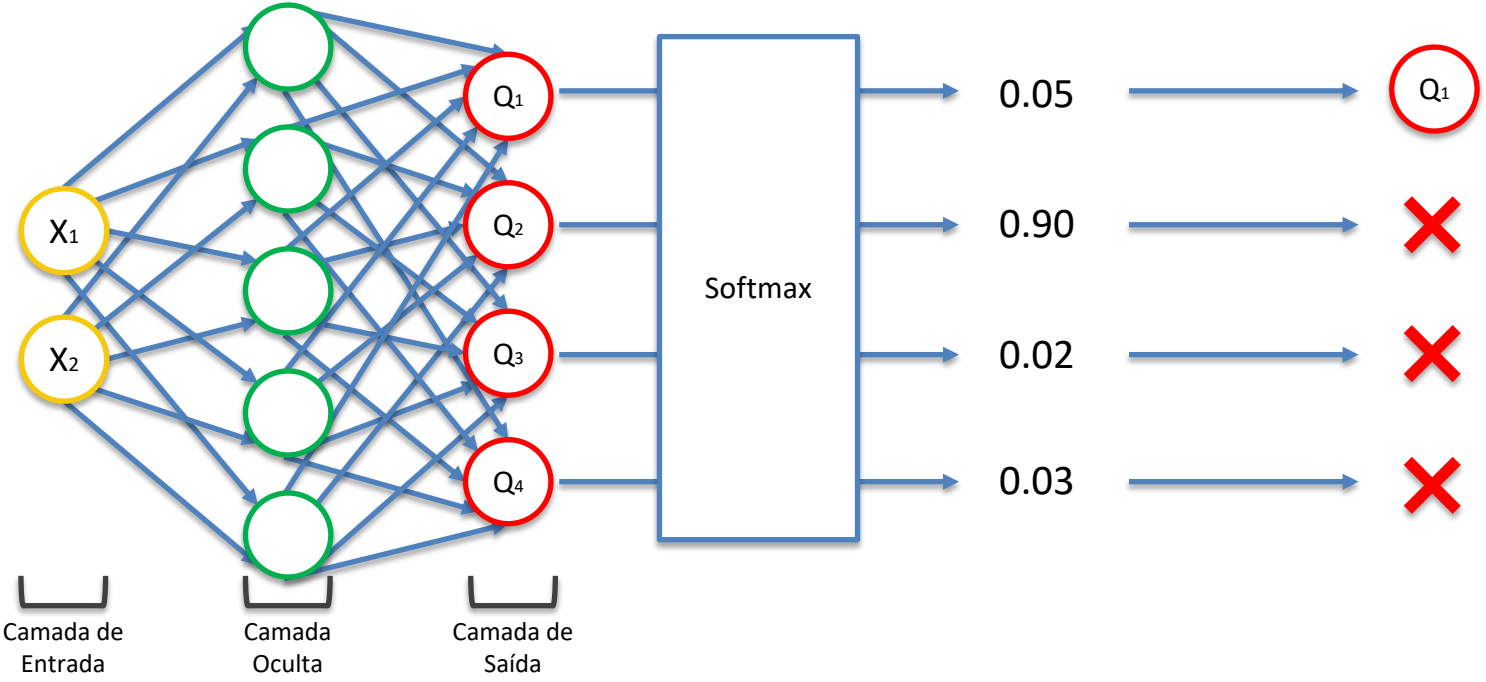
Ações:



© SuperDataScience & IA Expert

Política de Seleção de Ações

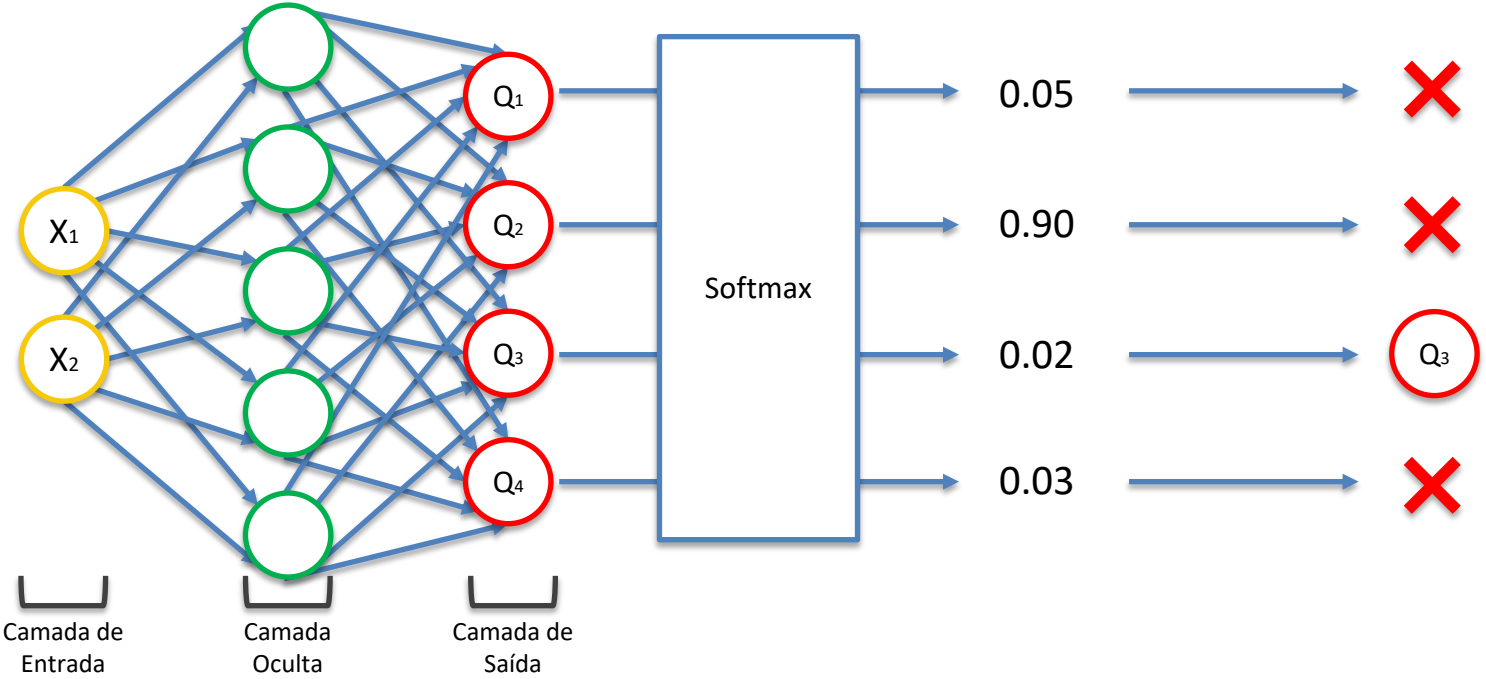
Ações:



© SuperDataScience & IA Expert

Política de Seleção de Ações

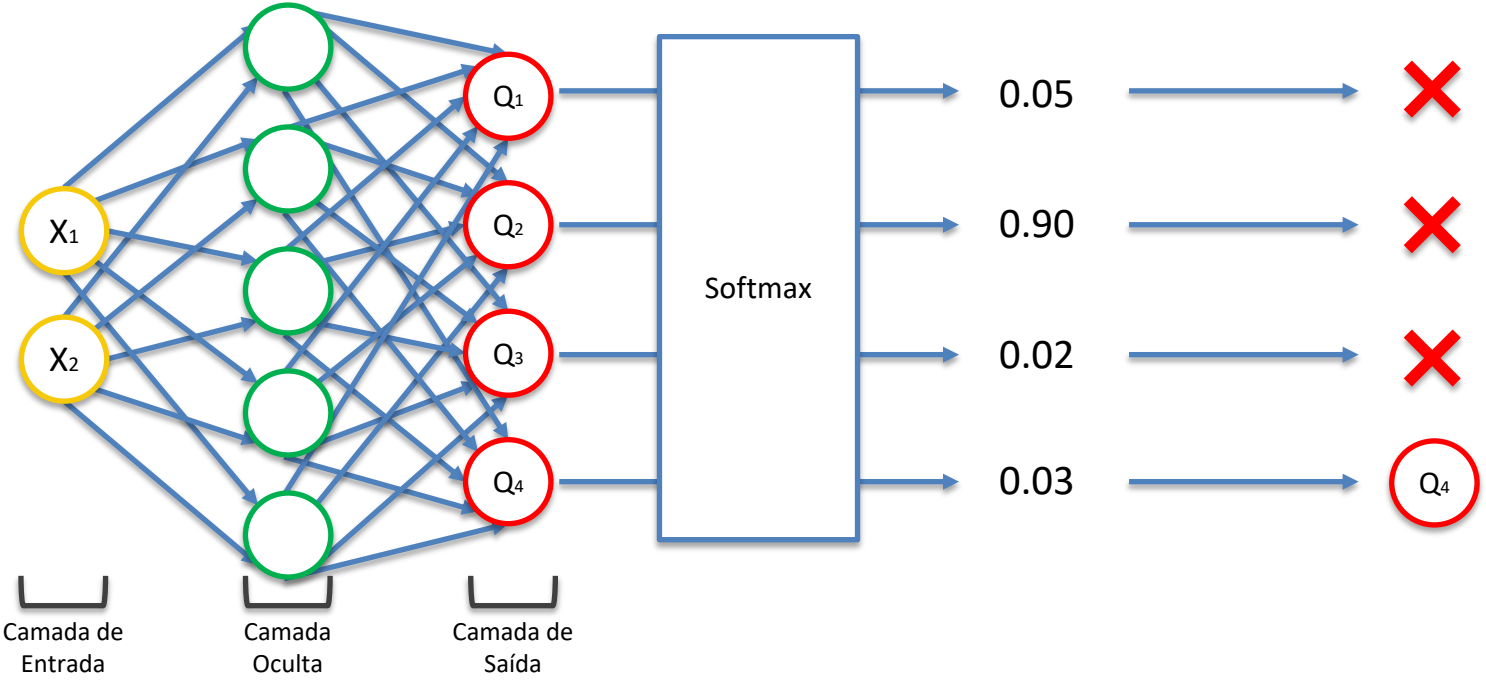
Ações:



© SuperDataScience & IA Expert

Política de Seleção de Ações

Ações:



© SuperDataScience & IA Expert

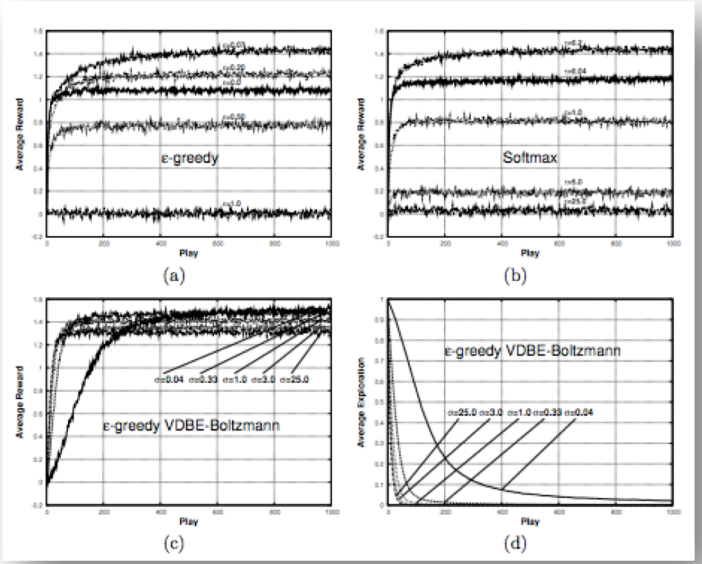
Leitura Adicional

Leitura Adicional:

Adaptive ε-greedy Exploration in Reinforcement Learning Based on Value Differences

Michel Tokic (2010)

Link:



http://tokic.com/www/tokicm/publikationen/papers/AdaptiveEpsilonGreedyExploration.pdf

© SuperDataScience & IA Expert

Conteúdo

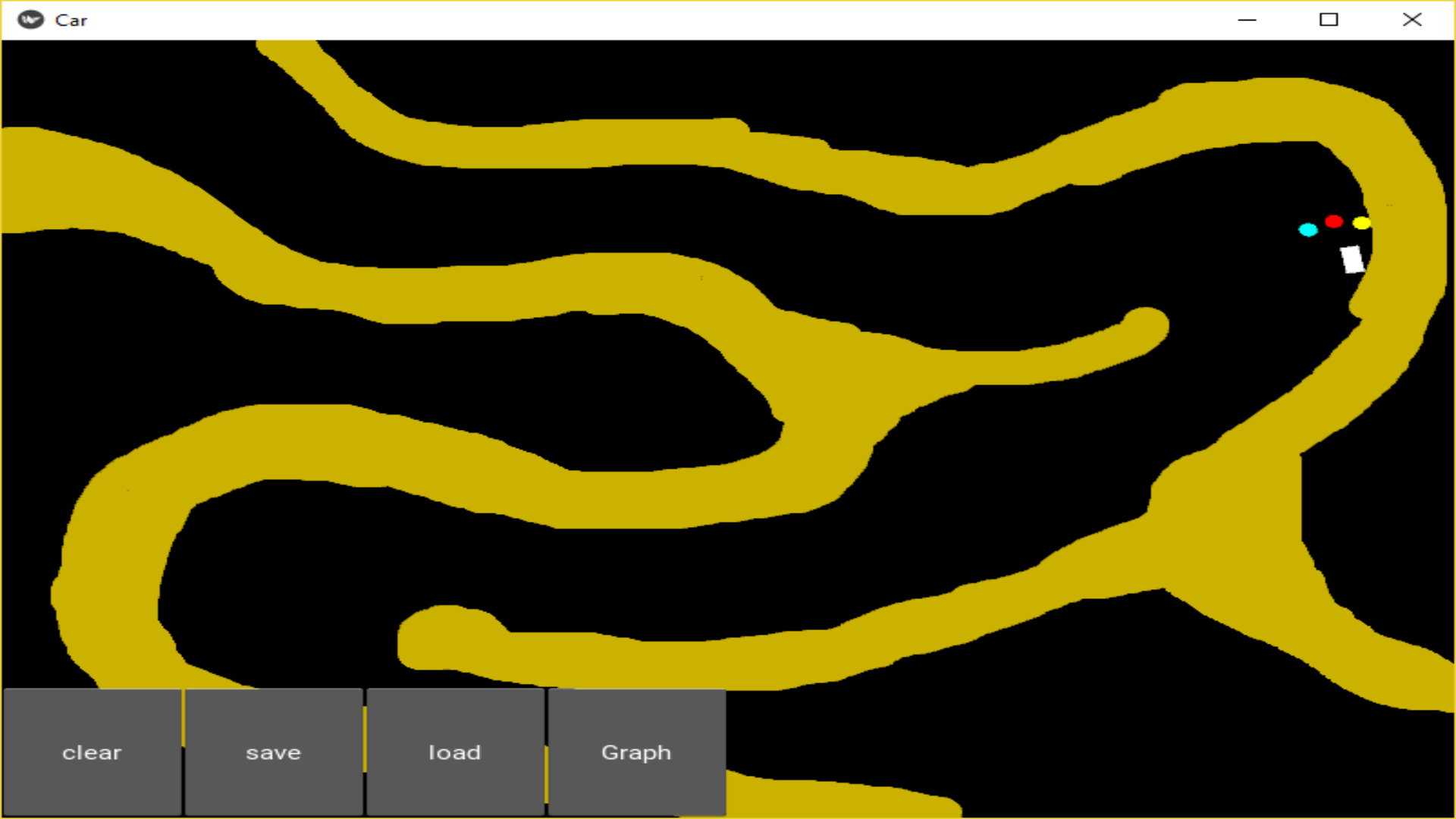
© SuperDataScience & IA Expert

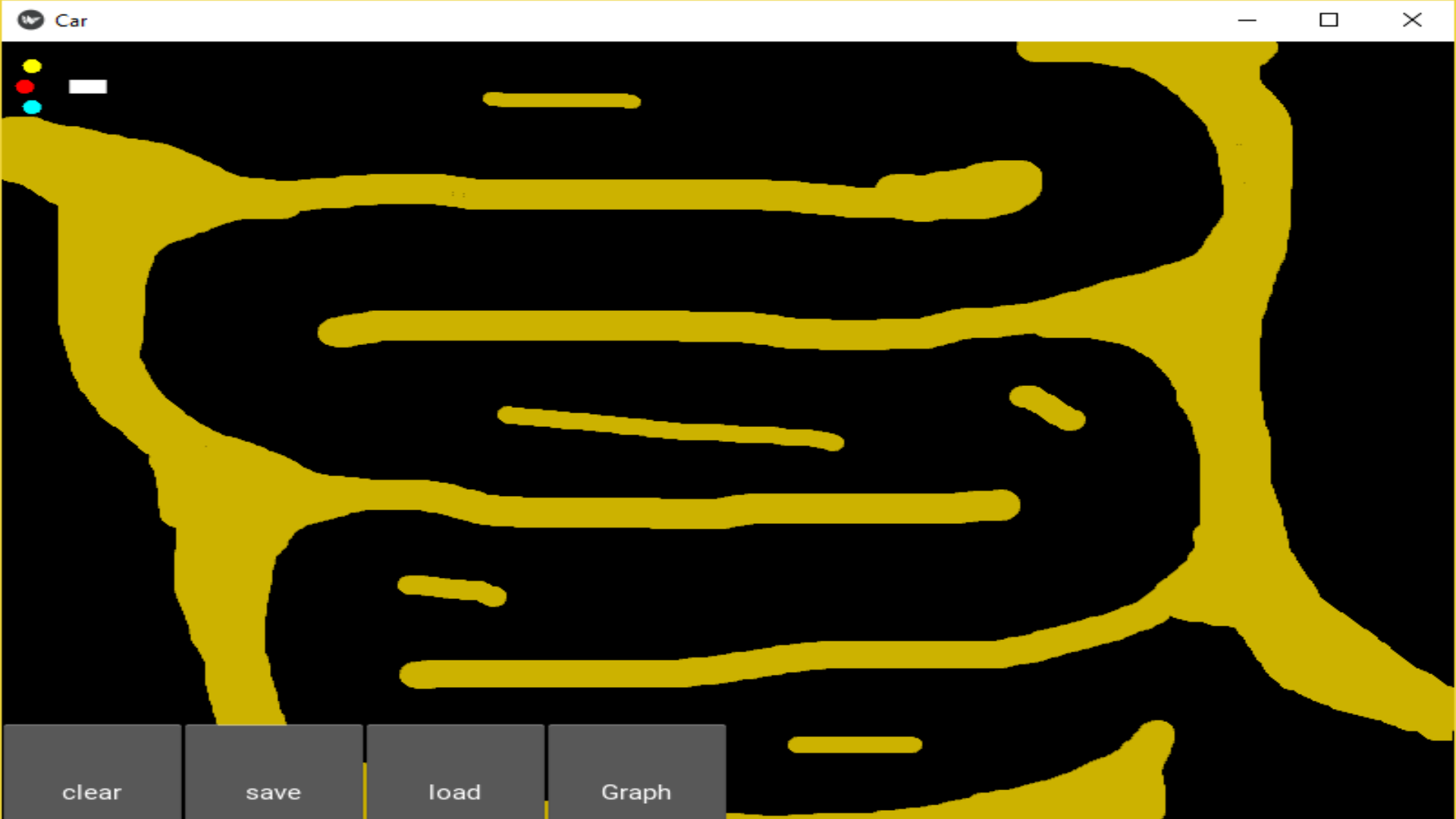
Conteúdo

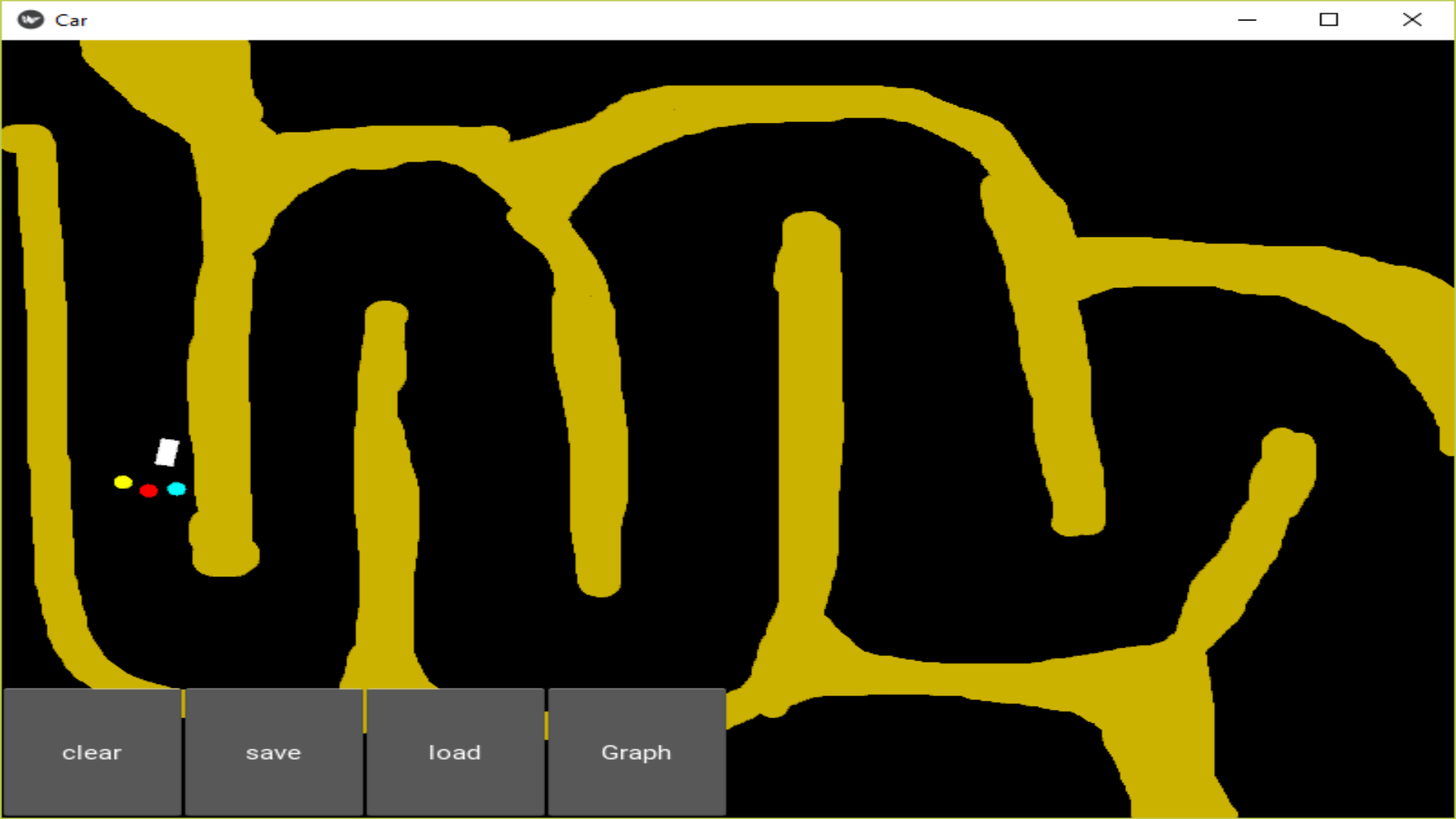
O que você aprenderá nessa seção:

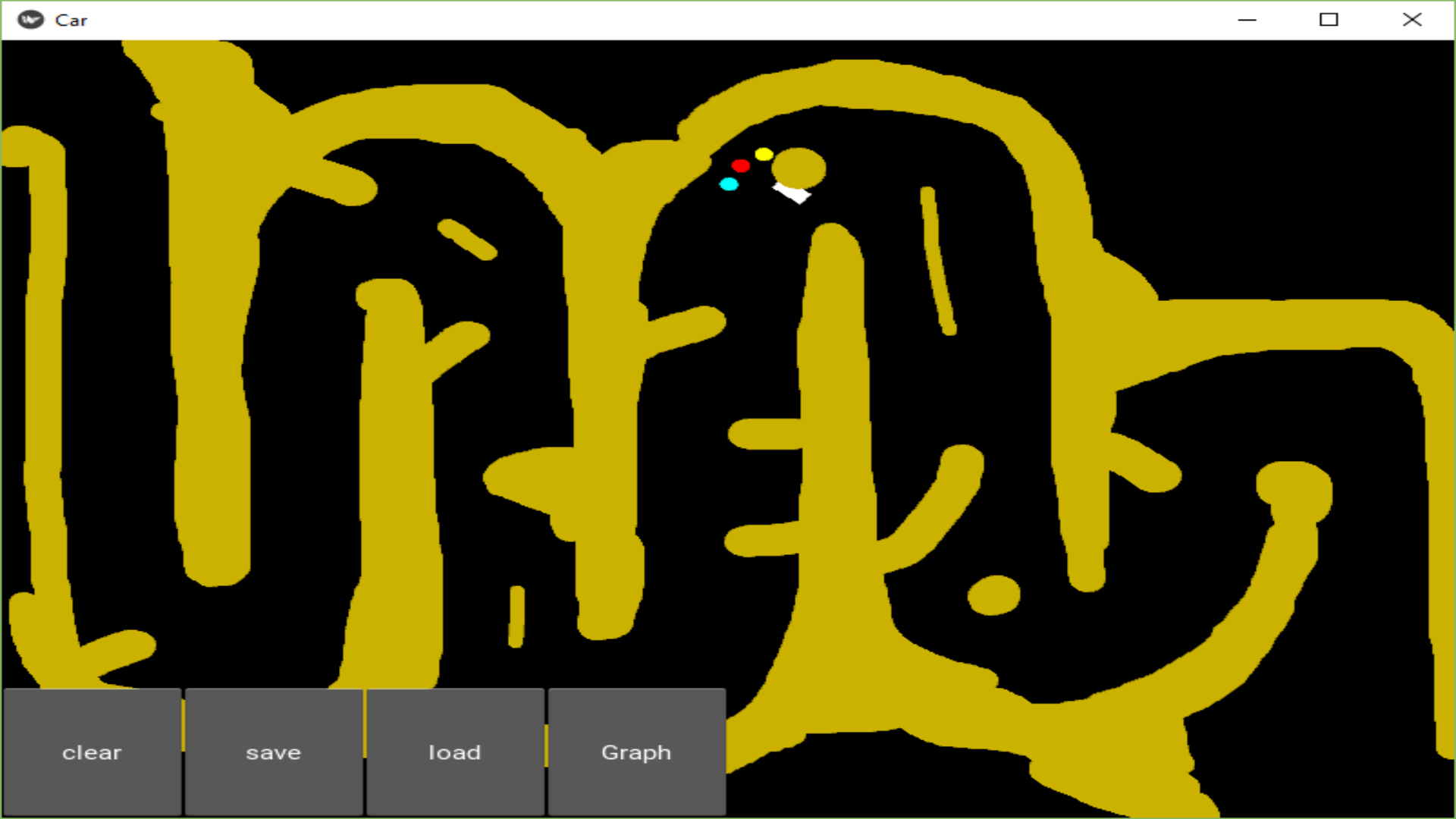
- Construção de um carro autônomo!
- Ambiente contendo o mapa
- IA para controlar o carro no mapa com Deep Q-Learning (PyTorch)
- Configuração do Ambiente

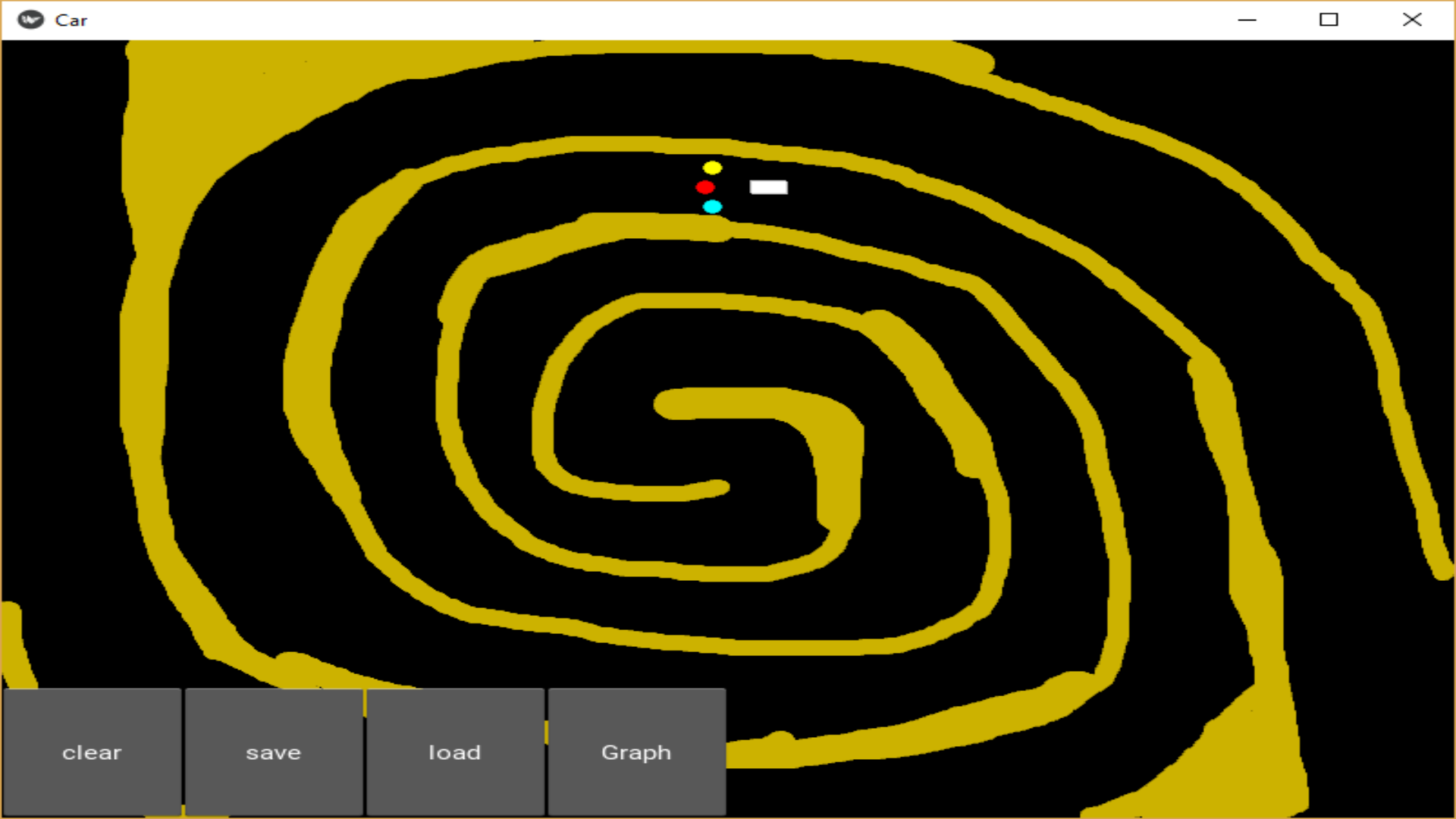
© SuperDataScience & IA Expert


Car
clear
save
load
Graph


Car
clear
save
load
Graph


Car
clear
save
load
Graph


Car
clear
save
load
Graph


Car
clear
save
load
Graph