

Machine Learning com Python

Prof. Luciano Galdino

APRENDIZAGEM POR REFORÇO

Aprendizado para tomar uma sequência de decisões até atingir uma meta sem ter parâmetros de referência.

O aprendizado se dá por tentativa e erro.

O agente percebe se está certo quando recebe uma recompensa e se está errado quando recebe uma punição.

Isto é, aprende tanto com erros como com acertos, assim como o ser humano.

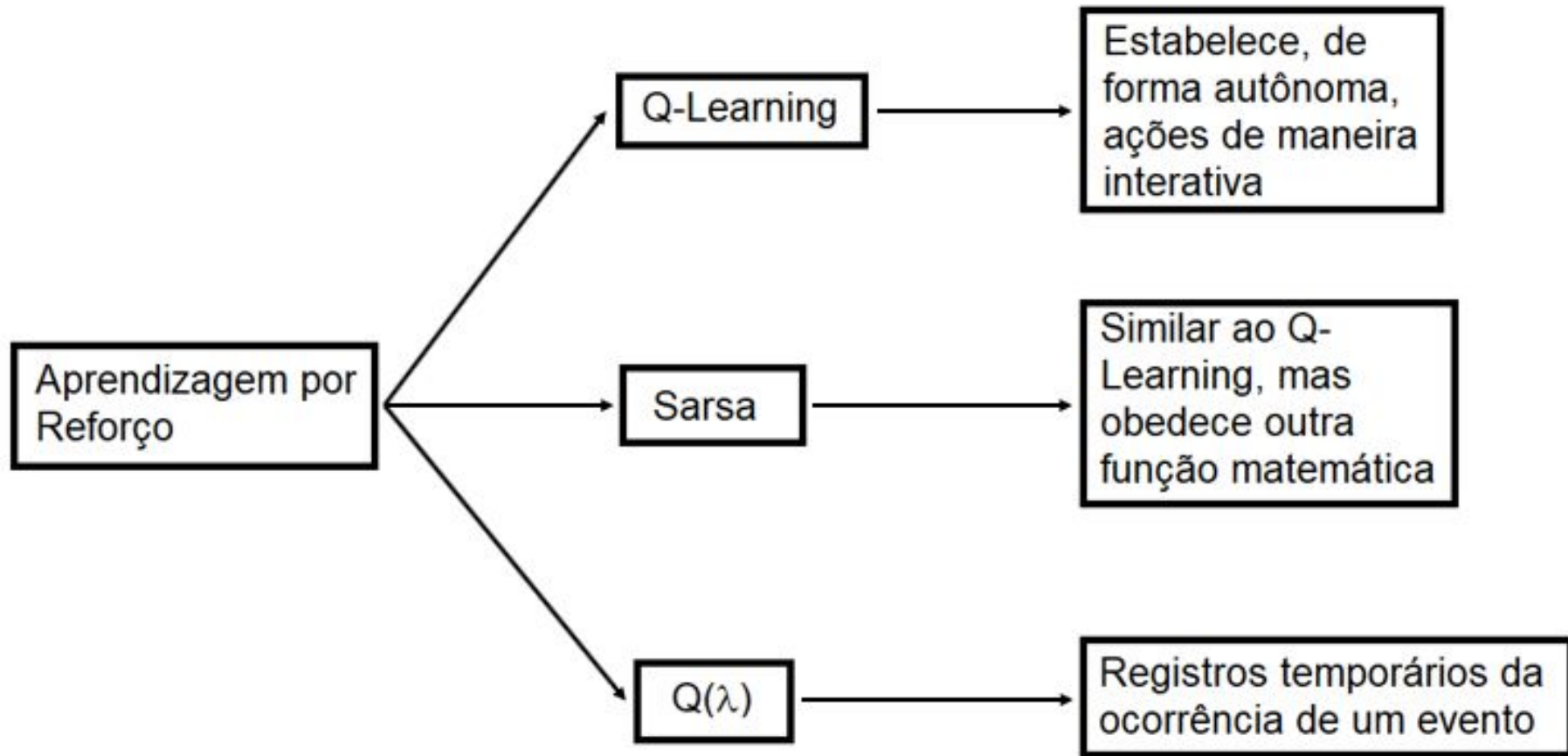
Algumas Aplicações

- Jogos
- Robótica
- Mercado Financeiro
- Aperfeiçoamento de carros autônomos
- Preços dinâmicos e personalizados
- Definição de estratégias para tratamento de doenças

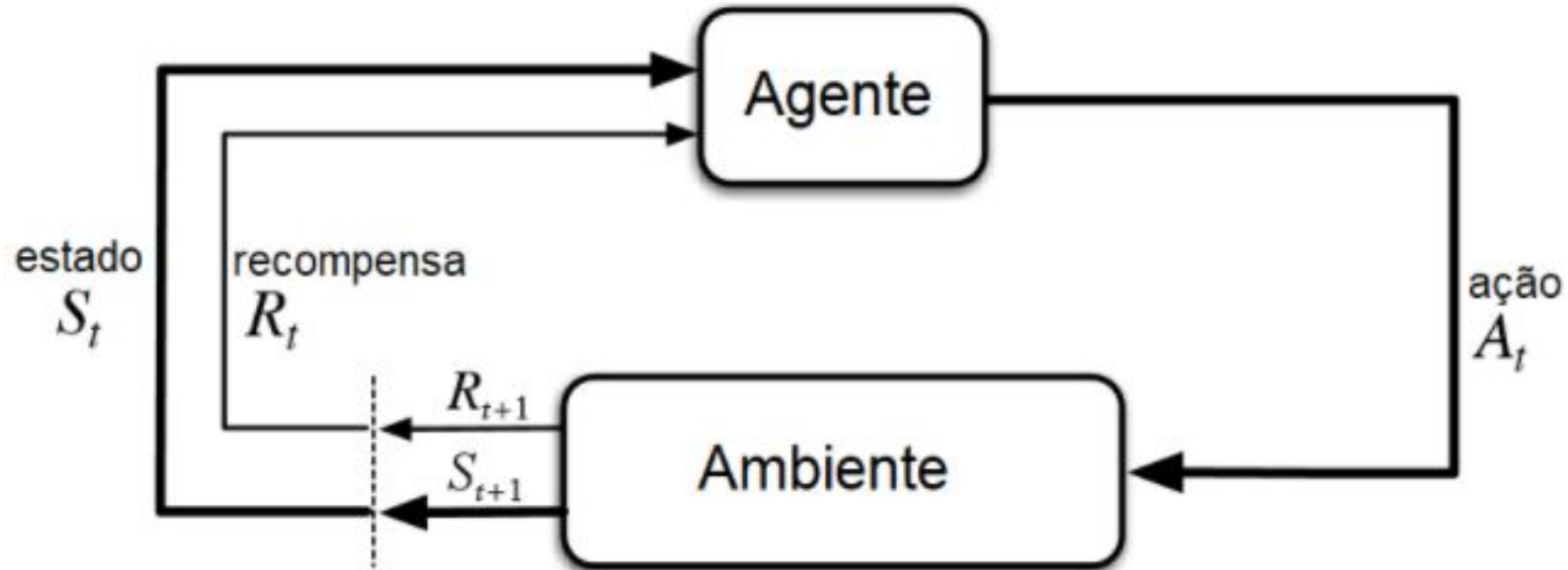
Nomenclaturas

- 1) **Agente:** Entidade que irá aprender (software e/ou hardware)
- 2) **Ambiente:** espaço de ação do agente (físico ou virtual).
- 3) **Estado (s):** situação atual do agente e ambiente.
- 4) **Política (π):** estratégia para a próxima ação.
- 5) **Recompensa (r):** referência para o agente sobre a sua ação.

Principais Algoritmos



Lógica da Aprendizagem



Utiliza o modelo matemático Markov Decision Process (MDP)

Equação de Bellman

$$Q(\text{state}, \text{action}) \leftarrow (1 - \alpha)Q(\text{state}, \text{action}) + \alpha \left(\text{reward} + \gamma \max_a Q(\text{next state}, \text{all actions}) \right)$$

α = taxa de aprendizagem

γ = fator de desconto.

Tabela Q:

	Action 1	Action 2	Action 3
State 1:	$Q = 1$	$Q = 10$	$Q = 12$
State 2:	$Q = 2$	$Q = 5$	$Q = 100$