|  |
| --- |
| **NOME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **RA:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

# 📋Atividade Complementar 3 – Temporal Diference- INSTRUÇÕES GERAIS:

1. Você pode executar esta atividade no Google Colab ou executar no seu computador. Execute as células com implementação antes de executar as tarefas.
2. Vedado acesso a celulares e consulta a sites externos como ChatGPT, Claude.ai, Gemini
3. Permitida a consulta ao material de aula e notebooks do Github

**📊 Formato de Entrega**

* **Use o template word (.docx) para download no SAVA e entregue um pdf no SAVA**
* **Data de Entrega : 31/10/2025 até 00:00**

**Para cada questão:**

1. **Print/screenshot** do resultado principal (gráfico ou tabela de win rates)
2. **Respostas** para cada pergunta procure esclarecer os conceitos empregados e apresentados em aula, como definições e pseudo códigos
3. **Observação adicional (opcional):** algo interessante que você notou

Objetivo Criar um Grid World Customizado e treinamento de algoritmos e análise de Algoritmos Temporal Diference.

# Q.1. Crie um GridWorld Customizado com as seguintes características (VALOR 1 ponto)

**Lay Out**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 |
| 0 |  |  | |  | |  | |  | |  | |  | | T1 R=+750 | | |
| 1 |  |  | |  | |  | |  | |  | |  | |  | | |
| 2 |  |  | |  | |  | |  | |  | |  | |  | | |
| 3 |  |  | |  | |  | |  | |  | |  | |  | | |
| 4 |  |  | |  | |  | |  | |  | |  | |  | | |
| 5 | S | C R=-100 | | C R = -100 | | C = -100 | | C = -100 | | C = 100 | | C = -100 | | T2 R=+1500 | | |

**Legenda:**

* S : Estado inicial típico (5,0)
* ■ : Parede em (1, 3), (1, 4), (1, 5),(2, 0), (2, 2), (2, 3), (2, 4),(2, 7),(3, 3), #(3, 4), (3, 5
* T1 : Terminal positivo em (0,3) Recompensa = 750.0
* T2 : Terminal positivo em (5,7) Recompensa = 1000.0
* C : Terminal negativo em (5, 1) recompensa = -100.0, #Precipio
* C : Terminal negativo em (5, 2) recompensa = -100.0, #Precipio
* C : Terminal negativo em (5, 3) recompensa = -100.0, #Precipio
* C : Terminal negativo em (5, 4) recompensa = -100.0, #Precipio
* C : Terminal negativo em (5, 5) recompensa = -100.0, #Precipio
* C : Terminal negativo em (5, 6) recompensa = -100.0, #Precipio

**Configuração**

* Tamanho: 6 linhas x 8 colunas
* Fator de desconto: γ = 0.95
* Ruído: 30% (0.3)
* Living reward: -10.00

Cole a imagem do ambiente e seus atributos gerados abaixo - (Dica: coloque o curso sobre a figura e clique com o botaõ direito do mouse e copie a imagem e cole no arquivo world)

* Use a função **create\_custom\_gridworld()** e **visualize\_gridworld()** e **print\_gridworld\_info()**

# Q.2 Treine o este ambiente com os seguintes parâmetros: (VALOR 1 ponto)

* Algoritmo: SARSA
* Episódios: 5000
* Alpha: 0.05
* Gamma: 0.9
* Episilon: 0.1

Use as seguintes funções para exibir os resultados:

* **Treinamento: Q\_sarsa, rewards\_sarsa = sarsa(gw\_custom, \*\*PARAMS, verbose=True)**
* Política: **visualize\_gridworld(gw\_custom, policy=policy, title=f"{name} - Política")**
* Valores Q: **visualize\_q\_values(Q, gw\_custom, title=f"{name} - Valores Q")**
* Q-values detalhados: **visualize\_q\_table\_detailed(Q, gw\_custom, title=f"{name} - Q-Values Detalhados")**

**Cole a imagem de cada um dos resultados obtidos abaixo, tome cuidado com o título do gráfico:**

# Q.3 Treine o este ambiente com os seguintes parâmetros: (VALOR 1 ponto)

* Algoritmo: Q-Learning
* Episódios: 5000
* Alpha: 0.05
* Gamma: 0.9
* Episilon: 0.1

Use as seguintes funções para exibir os resultados:

* **Treinamento: Q\_qlearning, rewards\_qlearning = q\_learning(gw\_custom, \*\*PARAMS, verbose=True)**
* Política: **visualize\_gridworld(gw\_custom, policy=policy, title=f"{name} - Política")**
* Valores Q: **visualize\_q\_values(Q, gw\_custom, title=f"{name} - Valores Q")**
* Q-values detalhados: **visualize\_q\_table\_detailed(Q, gw\_custom, title=f"{name} - Q-Values Detalhados")**

**Cole a imagem de cada um dos resultados obtidos abaixo, tome cuidado com o título do gráfico:**

# Q.4 Treine o este ambiente com os seguintes parâmetros: (VALOR 1 ponto)

* Algoritmo: EXPECTED SARSA
* Episódios: 5000
* Alpha: 0.05
* Gamma: 0.9
* Episilon: 0.1

Use as seguintes funções para exibir os resultados:

* **Treinamento: Q\_expected, rewards\_expected = expected\_sarsa(gw\_custom, \*\*PARAMS, verbose=True)**
* Política: **visualize\_gridworld(gw\_custom, policy=policy, title=f"{name} - Política")**
* Valores Q: **visualize\_q\_values(Q, gw\_custom, title=f"{name} - Valores Q")**
* Q-values detalhados: **visualize\_q\_table\_detailed(Q, gw\_custom, title=f"{name} - Q-Values Detalhados")**

**Cole a imagem de cada um dos resultados obtidos abaixo, tome cuidado com o título do gráfico:**

# Q.5 Compare os algoritmos treinados SARSA, Q-Learning e Expected SARSA (VALOR 2 pontos)

# Q.5 ITEM A Use as seguintes funções para exibir os resultados: (VALOR 0.5 ponto)

Texto

O conteúdo gerado por IA pode estar incorreto.

**Cole a imagem de cada um dos resultados obtidos abaixo, tome cuidado com o título do gráfico:**

# Q.5 Item B. RESPONDA À SEGUINTE QUESTÃO (VALOR 1.5 ponto)

**Os algoritmos TD são implementados com diferente variações no cálculo ds Q Values, essas variações resultam num comportamento esperado mais convervador para o algoritmo SARSA, um comportamento mais agressivo ou aproveitador para o Q-Learning e um comportamento ponderado para o Expected SARSA.**

**Usando as fórmulas do Valoe Estado Ação de cada algoritmo explique o que causa esse comportamento, mostre se os resultados obtidos no experimento anterior evidencia ou não esses comportamentos.**

**Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.**

**Uma imagem contendo Logotipo

O conteúdo gerado por IA pode estar incorreto.**

**Texto, Carta

O conteúdo gerado por IA pode estar incorreto.**

**Q.6 Item A. Elabore uma análise comparativa do comportamento do algoritmo SARSA para diferentes valores de α (alpha) [0.01, 0.1 e 0.5) (VALOR 1.5 ponto)**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Q.6 Item B. Explique com detalhes o que é o α (alpha) nos algoritmos de Diferença Temporal, esclarecendo qual seu efeito nas atualização de valor, demonstrando sua argumentação com os resultados obtidos no experimento do item anterior (Q.6 ITEM A).**

**Q.7. Ao analisar as políticas resultantes dos experimentos algumas ações podem ser inconsistentes. Avalie os resultados obtidos nas questões Q.2 até Q.4 e verifique ações que sugerem inconsistência e apresente aqui suas argumentações em caso positivo ou negativo, mostrando nos gráficos obtidos.**

**Imagine que nestes experimentos você verificou que uma ação não é consistente para uma politica ótima e precisa validar isso usando as ferramentas aprendidas, elabore um roteiro de como você faria está analise, não precisa escrever os script apenas as ferramentas algoritmos implementados, gráficos, análises etc..**

|  |  |  |
| --- | --- | --- |
| **QUESTÃO** | **VALOR** | **NOTA** |
| **Q.1** | **1.0** |  |
| **Q.2** | **0.5** |  |
| **Q.3** | **0.5** |  |
| **Q.4** | **0.5** |  |
| **Q.5 ITEM A** | **1.0** |  |
| **Q.5 ITEM B** | **2.0** |  |
| **Q.6 ITEM A** | **0.5** |  |
| **Q.6 ITEM B** | **1.5** |  |
| **Q. 7** | **1.5** |  |