Lista de exercício 2 – Aprendizado por Reforço

1. O que é um Processo de Decisão de Markov (MDP)?

a) Um modelo de aprendizado supervisionado.

b) Um modelo de aprendizado não supervisionado.

c) Um modelo que descreve a dinâmica sequencial de estados e ações, onde a próxima etapa depende apenas do estado atual e da ação atual.

d) Um modelo que não depende de probabilidades.

1. Em uma cadeia de Markov, o que é a propriedade de "estado de Markov"?

a) O estado atual é independente dos estados passados, dado o estado presente.

b) O estado atual depende apenas dos estados passados.

c) O estado atual depende apenas dos estados futuros.

d) O estado atual é independente de todos os outros estados.

1. Qual é o objetivo principal do Aprendizado por Reforço?

a) Minimizar o erro de previsão.

b) Maximizar o retorno cumulativo ao longo do tempo.

c) Aprender a partir de exemplos rotulados.

d) Maximizar a entropia do modelo.

1. O que é uma política em Aprendizado por Reforço?

a) Um conjunto de recompensas.

b) Uma estratégia que define a ação a ser tomada em cada estado.

c) Um tipo de reforço positivo.

d) Uma função de valor.

1. O que é "Living Penalty" em Aprendizado por Reforço?

a) Uma técnica de reforço que aumenta a recompensa ao longo do tempo.

b) Uma penalização que é aplicada a cada ação tomada, incentivando a tomada de ações rápidas.

c) Uma penalização que é aplicada a cada iteração, incentivando a exploração.

d) Uma técnica que remove todas as penalizações durante o treinamento.

1. O que é o algoritmo Q-Learning usado para fazer?

a) Previsão de estados futuros em um MDP.

b) Encontrar a política ótima em um MDP conhecido.

c) Treinar uma rede neural profunda.

d) Classificar os estados em um MDP.

1. Qual é a equação de atualização do Q-Learning?

a) Q(s, a) = R(s, a) + γ \* max(Q(s', a'))

b) Q(s, a) = R(s, a) + γ \* Q(s', a')

c) Q(s, a) = R(s, a) - γ \* max(Q(s', a'))

d) Q(s, a) = R(s, a) \* γ \* Q(s', a')

1. Qual é a principal desvantagem do Q-Learning?

a) É computacionalmente intensivo e requer muitos recursos.

b) Não pode lidar com ambientes estocásticos.

c) Requer uma política inicial definida.

d) Não pode ser usado para aprendizado profundo.

1. Qual é a diferença fundamental entre um Processo de Decisão de Markov (MDP) e uma Cadeia de Markov?

a) Em um MDP, não há ações, enquanto em uma Cadeia de Markov, ações são tomadas em cada estado.

b) Em uma Cadeia de Markov, não há estados, enquanto em um MDP, estados são definidos.

c) Um MDP inclui recompensas associadas a transições de estado, enquanto uma Cadeia de Markov não.

d) Não há diferença significativa entre eles; os termos são usados de forma intercambiável.

1. Qual é o principal objetivo da exploração em Aprendizado por Reforço?

a) Minimizar a recompensa imediata.

b) Maximizar a recompensa a longo prazo.

c) Garantir que o agente sempre escolha a ação ótima.

d) Reduzir a variação nas estimativas de valor.