Reinforcement Learning

1 Sobre aprendizado por reforço seguro, marque a alternativa que apresenta o objetivo principal deste tipo de aprendizado por reforço.

A Criar um algoritmo de aprendizado que seja parcialmente seguro durante o teste e altamente seguro durante o treinamento.

B Criar um algoritmo de aprendizado que seja inseguro durante o teste e seguro no treinamento.

C Criar um algoritmo de aprendizado que seja seguro durante o teste e inseguro durante o treinamento.

D Criar um algoritmo de aprendizado que seja seguro durante o teste e o treinamento.

2 A respeito da aprendizagem baseada em política, no contexto de Q-*learning*, julgue as afirmativas como falsas ou verdadeiras:

I. A política obrigatoriamente pode ser a política que de fato está sendo seguida durante o treinamento.

II. A aprendizagem Q baseada em política tenta aprender Q^{π}, a função Q para alguma política não projetada.

III. A aprendizagem Q baseada em política tenta aprender Q^{π}, a função Q para alguma política projetada.

IV. A política pode ser ou não a política que de fato está sendo seguida durante o treinamento.

Marque a alternativa que representa as afirmativas **verdadeiras**:

A II, IV

B I, IV

C II, III

D III, IV

3 Julgue os itens em verdadeiro ou falso:

I. A aprendizagem por reforço passiva faz uso de um conceito baseado em estados, em ambiente completamente observável.

II. Na aprendizagem por reforço passiva a política do agente é fixa, no estado s ele executa a ação.

III. O agente executa um conjunto de experiências no ambiente usando sua política.

IV. Em cada experiência, o agente começa num estado inicial e experimenta uma sequência de transições de estados até alcançar um dos estados terminais.

A V, V, F, V

B V, V, V, V

C V, F, F, V

D F, F, F, F

4 Um agente inserido em um processo de decisão de Markov:

I. Deve, para cada estado, escolher uma ação.

II. Verificar o estado em que o sistema se encontra.

III. Verificar uma política e efetivar uma ação.

IV. A ação executada pode ter um efeito sobre o ambiente e, assim, modificar o estado atual.

V. O agente verifica o novo estado para que possa tomar a próxima decisão.

A V, F, F, F, V

B F, F, V, V, F

C V, V, V, V, V

D V, F, V, F, V

5 Sobre os conceitos de aprendizagem por reforço, "o conjunto de todas as jogadas possíveis que o agente pode fazer". Assinale a alternativa **correta** que diz respeito a este conceito.

A Agente

B Ação

C Estado

D Ambiente

6 Em relação à Aprendizagem por Diferença Temporal, marque a alternativa **correta**.

A Os algoritmos de aprendizado por diferença temporal (DT) aprendem novas estimativas do valor com base em outras estimativas. O método DT não exige um modelo exato do sistema. Esse procura estimar valores de utilidade para cada estado do ambiente por recompensas oriundas das transições e de valores de estados sucessivos.

B O método DT exige um modelo exato do sistema.

C A aprendizagem ocorre indiretamente a partir da experiência.

D O método não procura estimar valores de utilidade para cada estado do ambiente por recompensas oriundas das transições e de valores de estados sucessivos.

7 A respeito dos comportamentos exploratórios clássicos na aprendizagem por reforço, marque a alternativa **correta**.

A Não assumem que o agente tem que explorar e aprender a pesar de ações diferentes e a agir de forma otimizada.

B Não assumem que o agente tem que explorar e aprender a pesar ações iguais e a agir de forma otimizada.

C Assumem que o agente tem que explorar e aprender a pesar de ações iguais e a agir de forma otimizada.

D Assumem que o agente tem que explorar e aprender a pesar de ações diferentes e a agir de forma otimizada.

8 *Reinforcement learning* ou aprendizado por reforço é um formalismo da inteligência artificial que permite a um agente aprender a partir da interação com o ambiente no qual ele está inserido.

A Esta aprendizagem se dá através do conhecimento sobre o estado do indivíduo no ambiente, das ações efetuadas neste e das mudanças de estado decorrentes das ações.

B Esta técnica não é indicada quando se deseja obter uma política ótima.

C Esta técnica conhece a priori a função que modela a política.

D Esta técnica o agente não deve interagir com seu ambiente diretamente para obter informações, que serão processadas por algoritmos apropriados.

9 Marque a alternativa que corresponde à definição de política ótima e função valor ótima de um MDP.

A Uma política ótima para um MDP é uma sequência de regras de decisão.

B Uma política ótima é, portanto, uma que fornece a utilidade que se espera mais alta.

C Uma política ótima é mapeada em um conjunto de ações, no qual cada ação tem a possibilidade de ser escolhida.

D A escolha de uma política ótima se sujeita a todo histórico de ações e estados do sistema até o presente momento.

10 A respeito do conceito de Programação Dinâmica Adaptativa, marque a alternativa **correta**.

A A ideia é que a utilidade de cada estado não seja a recompensa total que se espera a contar desse estado em diante.

B A ideia de Programação Dinâmica Adaptativa consiste em aprender o modelo de transição e a função de reforço empiricamente (ao invés das utilidades).

C Esse conceito consiste em não aprender o modelo de transição e a função de reforço empiricamente.

D A ideia é que a utilidade de cada estado é a recompensa total que se espera a contar desse estado em diante.