

Responda às questões apresentadas de forma clara e objectiva,
justificando sempre que necessário.

1. Considere as arquitecturas de agente anteriormente estudadas.
 - 1.1. ✓ Qual a relação entre memória e racionalidade numa arquitectura de agente deliberativa? Justifique.
 - 1.2. ✓ Qual o papel da noção de *comportamento* numa arquitectura de agente reactiva e qual a relação com a noção de *reação*? Justifique.
 - 1.3. Qual o papel das noções *esquema sensorial* e *esquema motor* e qual a sua relação? Justifique.
2. Pretende-se implementar um agente capaz de operar num ambiente caracterizado por um conjunto de estados S , por um conjunto de operadores O e pela função de transição de estado T , de seguida indicados.

$S = \{E1, E2, E3, E4, E5\}$

$O = \{o1, o2, o3\}$

$T(E1, o2) = E3$ ✓ ✓

$T(E2, o1) = E3$ ✓ ✓

$T(E2, o3) = E4$ ✓ ✓

$T(E3, o1) = E5$ ✓

$T(E3, o3) = E1$ ✓

$T(E4, o1) = E1$ ✓ ✓

$T(E4, o2) = E2$ ✓ ✓

Inicialmente o agente encontra-se no estado **E4** e tem por objectivo atingir o estado **E5**. Em situações de igualdade de escolha entre estados, têm maior prioridade os estados com maior índice. Ao nó inicial da árvore de procura corresponde a profundidade 0.

- 2.1. ✓ Tendo por base o método de ^{Ambos em procura em grafo} *procura em profundidade limitada* (limite de profundidade = 5), elabore e apresente a árvore de procura, bem como as respectivas estruturas de dados auxiliares. Indique a solução obtida.
- 2.2. ✓ Tendo por base o método de *procura de custo uniforme* e custo = 1 para todas as transições de estado, elabore e apresente a árvore de procura, bem como as respectivas estruturas de dados auxiliares. Indique a solução obtida.
- 2.3. ✓ Relacione os métodos de *procura bidireccional* e *procura em largura*, indicando as principais características de cada método, bem como as condições que devem ser garantidas para que essas características se verifiquem.

3. Pretende-se implementar um agente capaz de operar num ambiente descrito pelo modelo de seguida apresentado.

$S = \{E1, E2, E3, E4\}$	$T(E1, o1) = E2$	$R(E1, o1) = -1.0$
$A = \{o1, o2, o3, o4\}$	$T(E1, o4) = E3$	$R(E1, o4) = 10.0$
	$T(E2, o1) = E4$	$R(E2, o1) = -1.0$
	$T(E2, o3) = E1$	$R(E2, o3) = -1.0$
	$T(E3, o2) = E1$	$R(E3, o2) = -1.0$
	$T(E4, o3) = E2$	$R(E4, o3) = -1.0$
$T(s, a, s') = 0$, nas restantes situações		

Considere uma abordagem baseada em processos de decisão de Markov, e um factor de desconto temporal $\gamma = 0.9$.

- 3.1. ✓ Tendo por base uma política de selecção de acção *greedy*, resolva o processo de decisão de Markov correspondente à representação anterior, com um limiar de convergência de 2 ou até ao máximo de 5 iterações.
 - 3.2. ✓ Represente a política comportamental resultante da resolução da alínea anterior.
 - 3.3. ✓ Num processo de decisão de Markov, qual o papel do *modelo de transição de estado* e qual a relação com a noção de *operador* da procura em espaços de estados? Justifique.
4. Pretende-se implementar um agente capaz de aprender por reforço a agir no ambiente descrito na questão anterior. Considere um agente no estado inicial $s = E1$, $\alpha = 0.5$, $\gamma = 0.9$. Em situações de igualdade de escolha entre acções, o agente realiza a acção com maior índice.
- 4.1. Tendo por base o método de aprendizagem *Q-Learning* e uma política de selecção de acção *greedy*, simule a execução do algoritmo indicando, para cada iteração os estados, acções e recompensas envolvidas, bem como os valores da função $Q(s,a)$ alterados em cada iteração, para 3 iterações de aprendizagem.
 - 4.2. Represente a política comportamental resultante da resolução da alínea anterior.
 - 4.3. Na aprendizagem por reforço, qual o papel da *exploração*, como é regulada e como deve evoluir ao longo do tempo? Justifique.