

## Lista 2 – Processos de decisão de Markov

Considere um agente no ambiente 3×4 da Figura 1(a).

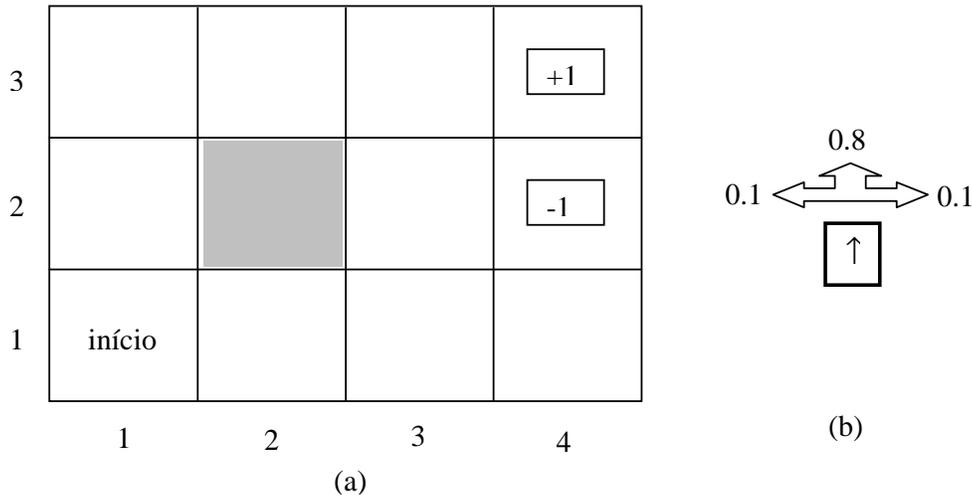


Figura 1: Ambiente e agente

A probabilidade de transição de estado  $P(S_{t+1}/S_t, x_t)$  é definida em 1(b), onde a seta indica a direção do movimento pretendida e a probabilidade de ocorrência do movimento (no exemplo da figura, a direção pretendida é *UP*, a probabilidade de se mover na direção *Up* é 0.8, *Left* é 0.1 e *Right* é 0.1). A partir do estado inicial [1, 1], o agente deve tomar uma decisão em cada passo  $t$ . A interação com o ambiente termina quando o agente atinge os estados marcados com +1 ou -1. Em cada local as decisões são *Up*, *Down*, *Left*, *Right*. Assume-se que o agente sempre sabe a posição onde está. O custo imediato em cada estado (exceto os marcados) é  $C_t(X_t, x_t) = -0.04$ . Supor  $\gamma = 1$ .

- determinar os valores de  $V_t(S_t) \forall S_t$  (exceto os marcados).
- determinar a política ótima utilizando o algoritmo iteração de valor.
- idem o item anterior utilizando o algoritmo iteração de estratégia.
- repetir itens (c) e (d) para  $C_t(X_t, x_t) = -0.02$ . Comparar com os resultados dos itens anteriores e interpretar.