Travaux dirigés d'
Apprentissage par renforcement

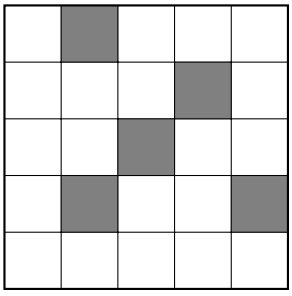
Université Paris–Saclay

Joon Kwon

mardi 28 novembre 2023

Exercice 1 (*Labyrinthe*). — Soit $n \geqslant 1$. On considère un labyrinthe carré de $n \times n$ cases. Chaque case est désignée par un couple $(i, j) \in \{1, \dots, n\}^2$ où i correspond à la ligne et j à la colonne. Les cases de départ et d'arrivée sont respectivement $(1, 1)$ et (n, n). Certaines cases sont murées et ne peuvent pas être accédées : on note $\mathscr{W}$ l'ensemble des cases murées. Le but du problème est de déterminer le chemin le plus court du départ à l'arrivée.

1) Modéliser le problème par un MDP.



2) Pour l'exemple donné en figure, et avec un taux d'escompte $\gamma = 1/2$, donner sans preuve une politique optimale π_* et les valeurs de la fonction état-valeur associée v_{π_*}.

Exercice 2 (*Différence de performance*). — Soit $(\mathcal{S}, \mathcal{A}, \mathcal{R}, p)$ un MDP fini et $0 < \gamma < 1$ le taux d'escompte. Soit π, π' deux politiques stationnaires. Pour $s \in \mathcal{S}$, on définit $d_{s,\pi} \in \mathbb{R}^{\mathcal{S}}$ comme suit :

$$d_{s,\pi}(s') = (1-\gamma) \sum_{t=0}^{+\infty} \gamma^t \mathbb{P}_{s,\pi} [\mathrm{S}_t = s'], \quad s' \in \mathcal{S}.$$

1) Montrer que $d_{s,\pi} \in \Delta(\mathcal{S})$.
2) Soit une variable aléatoire $\mathrm{S} \sim d_{s,\pi}$ et on définit $\alpha_\pi \in \mathbb{R}^{\mathcal{S} \times \mathcal{A}}$ par

$$\alpha_\pi(s', a) = q_\pi(s', a) - v_\pi(s'), \quad (s', a) \in \mathcal{S} \times \mathcal{A}.$$

Montrer que presque-sûrement :

$$v_\pi(s) - v_{\pi'}(s) = \frac{1}{1-\gamma} \mathbb{E}_{\mathrm{A} \sim \pi(\mathrm{S})} [\alpha_{\pi'}(\mathrm{S}, \mathrm{A})].$$