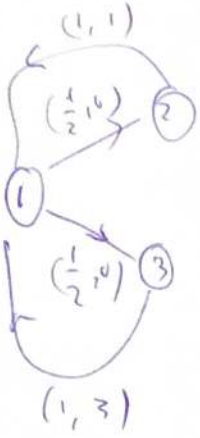
Exam
DPRL
Dec 2020
answers

1



a $\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$, $\phi = \langle \pi, r \rangle = 1$.

b
$$V(1) + \phi = \frac{1}{2}\left(V(2) + V(3)\right)$$
$$V(2) + \phi = 1 + V(1)$$
$$V(3) + \phi = 3 + V(1)$$

$V(1) = 0$, $\phi = 1 \Rightarrow V(2) = 0$, $V(3) = 2$

solutions: $\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\forall\, c \in \mathbb{R}$.

c $\langle V, \pi \rangle = 0 \iff \frac{1}{2} + c = 0 \iff c = -\frac{1}{2} \Rightarrow V^* = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1\frac{1}{2} \end{pmatrix}$

d MRC is periodic.

Replace P by $\delta P + (1-\delta) I$ for some $\delta \in (0,1)$ and solve.

2 $V_i(x_1, x_2) = \begin{cases} V_{i+1}(x_1, x_2) & \text{if } s_i > x_1 \text{ or } w_i > x_2 \\ \max\{r_i + V_{i+1}(x_1 - s_i, x_2 - w_i), V_{i+1}(x_1, x_2)\} & \text{otherwise} \end{cases}$

(x_1, x_2: remaining size/weight)

3 a $\pi_< = (\frac{1}{2}, \frac{1}{2})$, $\pi_> = (\frac{1}{3}, \frac{2}{3})$, $\phi = \frac{2}{3}$

b $V(1) + \phi = V(2)$

$V(2) + 2\phi = 2 + V(1)$

c $V(1) = 0$, $\phi = \frac{2}{3}$, $V(2) = \frac{2}{3}$

4. a+b

$$V(1) = \frac{1}{4}\left(V(2) + V(3)\right) \quad (0.5\beta)$$

$$V(2) = 1 + \frac{1}{2}V(1) \quad (\beta)$$

$$V(3) = 3 + \frac{1}{2}V(1) \quad (\beta)$$

$$\Rightarrow V(1) = \frac{1}{4}\left(4 + V(1)\right)$$

$$V(1) = \frac{4}{3}$$

$$V(2) = 1\frac{2}{3} \qquad V(3) = 3\frac{2}{3}$$

yes, unique, see th.

c.

$$V(1) = \frac{1}{4}\left(V(2) + V(3)\right)$$

$$V(2) = \max\left\{1 + \frac{1}{2}V(1),\ V(1)\right\}$$

$$V(3) = \max\left\{3 + \frac{1}{2}V(1),\ V(1)\right\}$$

$$m_1(1) = \frac{2}{3}(1-\beta)V(1) = \frac{2}{3}$$

d. Solution of a is still opt with restart, thus once you play this arm you will never stop.

5a Why a good exploration/exploitation tradeoff is absolutely needed in practice for some tasks when using planning?
--> To have reasonable computational requirements by reducing the breath and depth of search as compared to random search.

b Cite two techniques that can be used to reduce the *depth* of search.
--> Monte-Carlo rollouts and value function estimates (e.g. based on heuristics or based on neural networks).

c What is the reality gap?
--> The agent may not be able to interact with the true environment but only with an inaccurate simulation of it. In that case, the reality gap refers to the difference that exists between the simulation and the reality in which we would like to learn the policy.

d Cite two potential advantages of learning a *distribution* of discounted sum of future rewards instead of $Q^*(x,a)$.
--> It can allow for selecting a policy that does not only take into account the expectation (e.g. a risk averse policy), and it provides some auxiliary tasks that might help learning in practice when using function approximators.