DPRL
Exam
Feb 2021

1 a. aperiodic & comm

b. $\pi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, $\phi = 2\frac{1}{2}$

c. $V(1)=0$. $V(1) = 1 + \frac{1}{2}V(2) \Rightarrow V(2)=3$

$V(2) + \phi = 2 + \frac{1}{2}V(3)$ $\quad V(3)=7$

$V(3) + \phi = 3 + \frac{1}{2}V(2) + \frac{1}{2}V(4)$ $\quad V(4)=10$

$V(4) + \phi = 4 + \frac{1}{2}V(3) + \frac{1}{2}V(4)$ check

d. $\operatorname{argmax} \left\{ 1 + \frac{1}{2}V(1) + \frac{1}{2}V(2),\ r^* + V(4) \right\}$

$= \operatorname{argmax} \left\{ 2\frac{1}{2},\ r^* + 10 \right\} = 2$ iff $r^* > -7.5$

e. $\max\ r^* \pi(1,2) + \pi(1,1) + 2\pi(2) + 3\pi(3) + 4\pi(4)$

s.t. $\pi(1,2) + \pi(1,1) = \frac{1}{2}\pi(2) + \frac{1}{2}\pi(1,1)$

$\pi(2) = \frac{1}{2}\pi(1,1) + \frac{1}{2}\pi(3)$

$\pi(3) = \frac{1}{2}\pi(2) + \frac{1}{2}\pi(4)$

$\pi(4) = \pi(1,2) + \frac{1}{2}\pi(3) + \frac{1}{2}\pi(4)$

$\pi(1,2), \pi(1,1), \pi(2), \pi(3), \pi(4) \geq 0$

$\Sigma = 1$

2a. $V(1) = 1 + \frac{1}{4} V(1) + \frac{1}{4} V(2)$ $\times 4$

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

$4V(1) = 6 + \frac{3}{2} V(1) \Rightarrow V(1) = \frac{12}{5}$

$V(2) = \frac{16}{5}$

b. state 1, with restart:

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

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

solution: $\left(\frac{12}{5}, \frac{16}{5}\right) \Rightarrow m(1) = \frac{1}{2} \frac{12}{5} = \frac{6}{5}$

state 2, with restart:

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

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

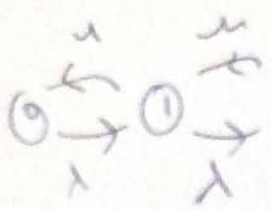
$V(1) = V(2) = 2 + \frac{1}{2} V(1) = 4. \Rightarrow m(2) = \frac{1}{2} \cdot 4 = 2$

c. ~~Pull all bandits~~

Every step: 1. Pull a bandit in state 2

2. If none, pull a bandit in state 1

never touch the "1.1 - bandit".

3a.  ... or 

b. ~~[illegible]~~

$$V(x) + \frac{1}{\lambda+\mu}\phi = \frac{x}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}V(x+1) + \frac{\mu}{\lambda+\mu}V((x-1)^+)$$

c. solution: $\phi = \frac{\lambda}{\mu-\lambda}$, $V(x) = \frac{x(x+1)}{2(\mu-\lambda)}$

e.g., $x=0$: $\frac{1}{\lambda+\mu}\frac{\lambda}{\mu-\lambda} = \frac{\lambda}{\lambda+\mu}\frac{2}{2(\mu-\lambda)} + 0$ ✓

d. one-step iteration: consider $\Delta_i = V_i(x+1) - V_i(x)$ for queue i.

Choose queue with lowest Δ_i

(choice of λ_i: can depend on optimisation problem)

Exam Question:
a What is the principle of "optimism under uncertainty"?
--> It is a principle that aim at encouraging the exploration of the parts of the environment where more uncertainty is present (usually =less data).

b. Cite three elements on which it is possible to work as a data scientist to improve the generalization capabilities of RL algorithms.
--> (1) an abstract representation that discards non-essential features, (2) the objective function (e.g., reward shaping, tuning the training discount factor) and (3) the learning algorithm (type of function approximator and model-free vs model-based).

c. When estimating the expected return of following a given policy, Monte Carlo rollouts can be used. Cite one advantage and one drawback of using Monte Carlo rollouts.
--> (+) The Monte-Carlo estimator is an unbiased well-behaved estimate.
(-) the main drawback is that the estimate requires on-policy rollouts and can exhibit high variance (Many rollouts are thus needed).

d. In Q-learning, what are the two potential characteristics of the state space that make is it necessary to use function approximators?
--> It is important when there is a large and/or continuous state space.