

RL-Theory -2

Vishal Sarmah
2017209.

1.

s	a	s'	r	$P(s' s,a)$
high	search	h	r_s	α
h	s	l	r_s	$1-\alpha$
h	w	h	r_w	1
l	s	h	-3	$1-\beta$
l	s	l	r_s	β
l	w	l	r_w	1
l	r	h	0	1

For s , $h \rightarrow$ high
 $l \rightarrow$ low

For a , $s \rightarrow$ search
 $w \rightarrow$ wait
 $r \rightarrow$ recharge

For s' , $h \rightarrow$ high
 $l \rightarrow$ low

3. 3.15

Signs of the rewards are also important since, it would affect overall expected sum reward.

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$$

Adding constant 'c' to each reward,

$$G'_t = (R_{t+1} + c) + \gamma(R_{t+2} + c) + \dots$$

$$\Rightarrow G_t' = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$$

$$+ c + c\gamma + c\gamma^2 + \dots$$

$$\Rightarrow G_t' = G_t + \frac{c}{1-\gamma} \quad \text{assuming } 0 \leq \gamma < 1$$

$$\Rightarrow V_c = \frac{c}{1-\gamma}$$

Hence, adding constant to each term leads to addition of a constant to the overall sum.

3.11

For episodic task,

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{T-t-1} R_{T+1}$$

$$= \sum_{k=t+1}^T \gamma^{k-t-1} R_k$$

Adding a constant 'c' to all rewards, would again similar to 3.10 would add a constant to ~~the~~ all the returns.

Since, we are adding a constant to all rewards it would be similar to assigning new rewards. Thus, task remains unchanged.

eg:- If in a gridworld,

from (0,1) to (2,3) reward is 3, and 0 for all actions.
 Then we add +2 to it making the reward +5 from (0,1) to (2,3) and +2 otherwise.

all funcⁿ would ~~be~~ ~~seen~~
But relatively, ~~action values~~ ~~will~~ remain similar.
Thus, task won't change

5. $V^*(s)$ is the optimal state-value function
 $Q^*(s,a)$ is the optimal ~~opt~~ action-value fn

$$\text{thus, } V^*(s) = \max_{a \in A(s)} \{Q^*(s,a)\}$$

Since, we ~~need~~ ~~opt~~ would greedily choose
optimal. action. For. optimal state value
function.