RL Assignment-2

Ashish Sethi
MT-18024

Question - 2

Results

Initial Value Function

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Value Function after policy iteration

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.7	-0.4
-1	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

Question - 4

Results

Initial Value Function

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Value Function after policy iteration

22	24.4	22.	19.4	17.5
19.8	22.	19.8	17.8	16
17.8	19.8	17.8	16	14.4
16	17.8	16	14.4	13.0
14.4	16.0	14.4	13.0	11.7

Question - 6

Policy Improvment

Intial Random Policy

3	2	3	3
2	2	3	2
0	1	2	3
3	0	0	0

Final Policy

0	0	0	0
1	0	0	3
1	0	2	3
1	2	2	0

Value Iteration

Intial Policy

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Final Policy

0	0	0	0
1	0	0	3
1	0	2	3
1	2	2	0

RL Assignment-2
Ashish Sethi
MT18024

Question-1

s	a	s'	$p(s' \mid s,a)$	$r(s,a,s')$	$p(s',r \mid s,a)$
high	search	high	α	r_{search}	$0.1\alpha r_{search}$
high	search	low	$1-\alpha$	r_{search}	$-0.1(1-\alpha) r_{search}$
low	search	high	$1-\beta$	-3	0
low	search	low	β	r_{search}	$-0.1\beta r_s$
high	wait	high	1	r_{wait}	$0.1 r_w$
high	wait	low	0	—	—
low	wait	high	0	—	—
low	wait	low	1	r_{wait}	$0.1 r_w$
low	recharge	high	1	0	—
low	recharge	low	0	—	—

As we know that -

$$r(s,a,s') = \sum_{r \in R} \frac{r\, p(s',r \mid s,a)}{p(s' \mid s,a)}$$

$$p(s',r \mid s,a) = \frac{r(s,a,s')\, p(s' \mid s,a)}{\sum_{r \in R} r}$$

(1). $p(s', r \mid s, a) = \dfrac{r_{search} \cdot \alpha}{10}$

$= 0.1\, r_{search}\, \alpha.$

(2) $-\dfrac{(1-\alpha)\, r_{search}}{10}$

$= -0.1(1-\alpha)\, r_s$

Similarly we can calculate the other values as well.

Date
Page

Q3 Exercise 3.15

we know that

$$V^{\pi}(s) = E_{\pi}\{R_{t} \mid s_t = s\}$$

$$= E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^{k} r_{\cancel{t+}k+1} \mid s_t = s\right\}$$

Add constant in reward.

$$\bar{r}_{t+k+1} = r_{t+k+1} + C$$

$$V^{\pi}(s) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} \bar{r}_{t+k+1} \mid s_t = s\right]$$

$$= E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^{k} r_{\cancel{k+t}+1} \mid s_t = s\right\}$$

$$+ E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^{k} c \mid s_t = s\right\}$$

classmate

Date

Page

$$\hat{V}^{\pi}(s) = V^{\pi}(s) + c \sum_{k=0}^{\infty} \gamma^{k}$$

$$\boxed{\hat{V}^{\pi}(s) = V^{\pi}(s) + \frac{c}{1-\gamma}}$$

Ex - 3.16

Adding a constant at episodic tasks

$$V_{\pi}(s) = E\left[R_{t+1} + \gamma R_{t+2} \cdots \gamma^{k-1} R_{t+k} \,\middle|\, S_t = s\right]$$

$$= E\left[(R_{t+1} + c) + \gamma(R_{t+2} + c) + \gamma^{k-1}(R_{t+k} + c)\right]$$

$$c\left[1 + \gamma + \gamma^{2} \cdots \gamma^{k-1}\right] + V_{\pi}(s)$$

$$\boxed{\frac{c(1-\gamma^{k})}{1-\gamma} + V_{\pi}(s)}$$

classmate

Date

Page

Question - 5

$$V_\pi(s) = E_\pi\left[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s\right]$$

$$= E_{\pi^*}\left[R_{t+1} + \gamma V_*(S_{t+1}) \mid S_t = s\right]$$

$$= \max_a E\left[R_{t+1} + \gamma V_*(S_{t+1}) \mid S_t = s\right]$$

$$= \max_a \sum_r \sum_{s'} (r + \gamma V_*(s')) \, p(s' r \mid s, a)$$

$$V_*(s) = \max_a q_*(s, a)$$

Question - 6

The bug which causes the policy to keep on switching in case of finding multiple policies.

In this case

if $V_\pi(s) \le \theta$

don't update the $V_\pi(s)$

θ increase by a small value to break continous update of state

Question 4.

Non linear solution of bellman equation using linear programming

$$V(s) \geq R(s) + \gamma \max_{a \in A} \sum P(s'|s,a) V(s')$$

via $|A|$ is linear constraints.

$$V(s) \geq R(s) + \gamma \sum_{s' \in S} P(s'|s,a) V(s')$$

$\forall\ a \in A$

Now using linear program.

s.t to $V(s) \geq R(s) + \gamma \sum_{s' \in S} p(s'|s,a) V(s')$

$\forall\ a \in A,\ s \in S$

~~Theorem~~

suppos

$$V(s) \geq R(s) + \gamma \max_{a \in A} \sum_{s' \in S} P(s'|s,a) V(s')$$

In objective, we can optimize any positive linear function of $V(s)$ and result above will be true

$$\text{minimize} \sum_{s} d(s) V(s)$$

s.t. $V(s) \geq R(s) + \gamma \sum_{s' \in S} P(s'|s,a) V(s')$

$\forall\ a \in A,\ s \in S$.

$d(s)$ is distribution over states

Adding dual variables $\mu(s, a)$

Maximize $\sum_{s \in S} R(s) \sum_{a \in A} \mu(s, a)$.

s.t $\sum_{a \in A} \mu'(s', a) = d(s') + \gamma \sum_{s \in S} \sum_{a \in A} P(s'|s,a) \mu(s, a)$

$\forall s' \in S$

$\mu(s, a)$