(S - silent, Q - query)

Q 13. MDP

1.

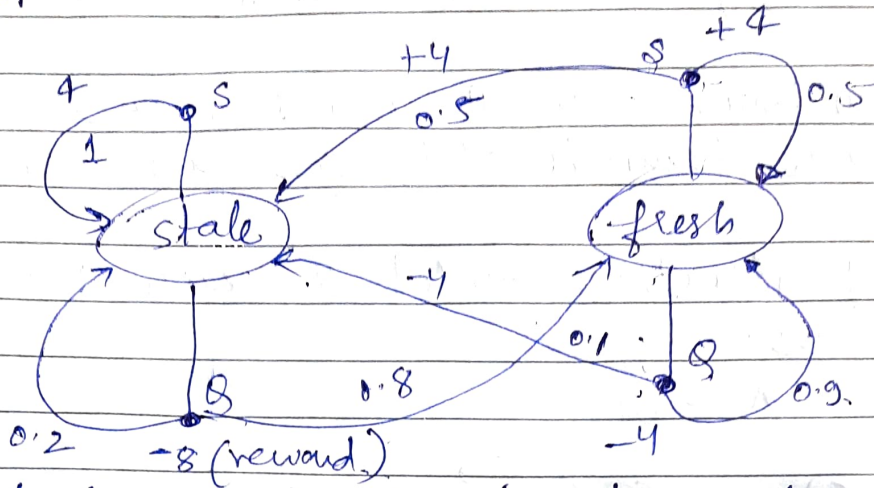


table drawn in the jupyter coded file.

3. discounting factor $\gamma = 0.5$ (Y)

$S_0\ A_0\ R_1\ S_1\ A_1\ R_2\ S_2\ A_2\ R_3$

at 3rd time step.

If information, $R_3 = \{-10, S_2 = \text{stale}\}$
$\{+10, S_2 = \text{fresh}\}$

Let S = Stale, F = fresh

~~V_q (state)~~ = value functions.

(A) $V_q(S) = \frac{1}{2}\left(0.8\left(-8 + 0.5 \cdot V_q(F)\right)\right) + \frac{1}{2}\left(0.2\left(-8 + \underset{1}{\overset{0.5}{V_q(S)}}\right)\right)$

$+ \frac{1}{2}\left(4 + 0.5\, V_q(S)\right)\Big) \times 1$

(B) $V_q(F) = \frac{1}{2}\left(0.9\left(-4 + \underset{1}{\overset{0.5}{V_q(F)}}\right)\right) + \frac{1}{2}\left(0.1\left(-4 + \underset{1}{\overset{0.5x}{V_q(S)}}\right)\right)$

$+ \frac{1}{2}\left(0.5\left(4 + \underset{1}{\overset{0.5}{V_q(F)}}\right)\right) + \frac{1}{2}\left(0.5\left(4 + 0.5 \times V_q(S)\right)\right.$

let $V_q(S) = V_q(F) = 0$

Date
Page

$$V_q(S) = \frac{1}{2}\left(0.8(-8) + 0.2(-8) + 4\right)$$

$$V_q(F) = \frac{1}{2}\left(0.9(-4) + 0.1(-4) + 0.5(4) + 0.5(4)\right)$$

$V_q(S) = -2$, theta = 0.3
$V_q(F) = 0$, delta(Δ_F) = 0
delta (Δ_S) = 0. or some minimum value.
loop while (1):
loop for $s \in (S, F)$, $V_q(s) = V_q(F) = 0$.
From (A) and (B)
calculati value functions
~~delta~~
$\Delta_S \leftarrow \max(\Delta_S, |v - V(s)|)$
Similarly calculate for,
$\Delta_F \leftarrow \max(\Delta_F, |v - V(F)|)$
condition:
loop until $\Delta < 0$.

So, for next iteration we have,
$V_q(s) = -2$, $V_q(F) = 0$

~~$V_q(S) = \frac{1}{2}\{0.8(-8) + 0.2(-8 + (-2)) + 4 +$~~

time step 2

$$V_q(S) = \frac{1}{2}\left(0.8(-8) + 0.2\left(-8 + 0.5(-2)\right) + 4 + 0.5(-2)\right)$$

$$V_q(F) = \frac{1}{2}\left(0.9(-4) + 0.1\left(-4 + 0.5(-2)\right) + 0.5(4) + 0.5(4 + 0.5(-2))\right)$$

$V_q(S) = -2.6$	$\Delta_S = 0.6$ $\Delta = \max(\Delta_S, \Delta_F)$
$V_q(F) = -0.3$	$\Delta_F = 0.3$
	$\Delta = 0.6$

$V_q(S) = -2.6$, $V_q(F) = -0.3$

3. time step 3

$$V_q(S) = \frac{1}{2}\left[0.8\left(-8+0.5(-0.3)\right) + 0.2\left(-8+0.5(-2.6)\right) + 4 + 0.5(-2.6)\right]$$

$$V_q(F) = \frac{1}{2}\left[0.9\left(-4+0.5(-0.3)\right) + 0.1\left(-4+0.5(-2.6)\right) + 0.5\left(4+0.5(-0.3)\right) + 0.5\left(4+0.5(-2.6)\right)\right]$$

$V_q(S) = -2.84$ $\quad \Delta_S = \max\left(\Delta_{S_p}, |V - V_q(S)|\right)$

$V_q(F) = -0.5$ $\quad \Delta_F = \max\left(\Delta_{F_p}, |V - V_q(F)|\right)$

$\Delta = 0.24$ [max updated value between transition state]

Therefore after 3 iteration, the Δ is less than taken $\theta = 0.3$,

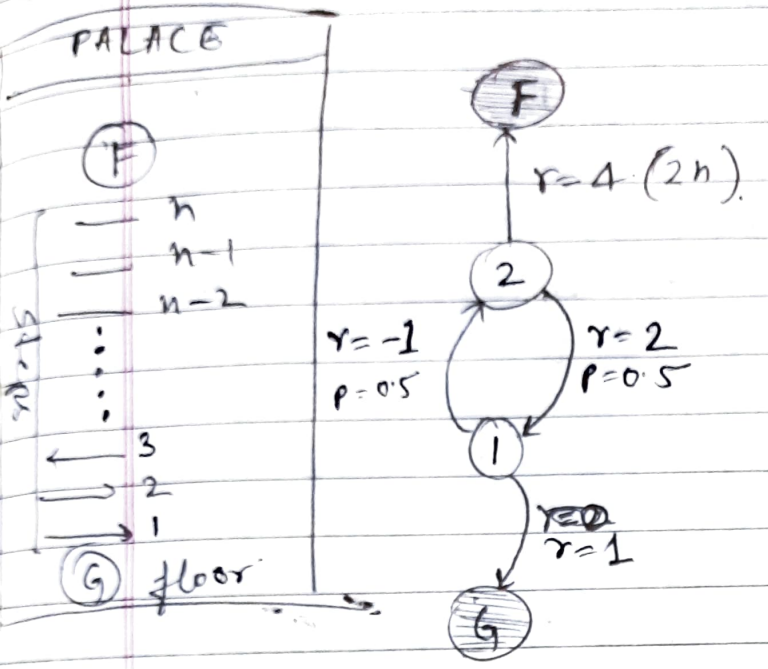
Policy improvement step

So, to choose policy, (optimal, we pick the we action state with max value.

$\pi^*(s) = \max\left(V_q(S), V_q(F)\right)$

$\pi(s) = V_q(F)$ is maximum, hence choosing Fresh information will have optimal policy.

Q15 - MDP for n = 2, Stair well.



MDP.

discount factor $\gamma = 1$, where r represents reward.
Following policy π (where $\pi(up|s) = 0.5$
$\pi(down|s) = 0.5$),
robot can start in any of the step (1 or 2).
So, given we have

$$V_\pi(s) = \sum_\pi \pi(a|s) \sum_{s,r} p(s', r|s, a)[r + \gamma V_\pi(s')]$$

we have

$$V_\pi(s=1) = \frac{1}{2}\left(0.5(-1 + V_\pi(2)) + \frac{1}{2}\left(0.5(1 + V_\pi(G))\right)\right.$$

$$V_\pi(s=2) = \frac{1}{2}\left(0.5(2 + V_\pi(1)) + \frac{1}{2}\left(0.5(4 + V_\pi(F))\right)\right.$$

Considering the values of terminal states as 0.

$$V_\pi(G) = V_\pi(F) = 0 \quad \text{---} \quad (1)$$

Solving the equations we have.

$$V_q(1) = 0.25\,V_q(2) + V_q(G) + 0.25$$

$$V_q(2) = 1.5 + 0.25\,V_q(1) + 0.25\,V_q(F)$$

From (1)

$$V_q(1) = 0.25\,V_q(2) \quad \text{—— (A)}$$

$$V_q(2) = 1.5 + 0.25\,V_q(1) \quad \text{—— (B)}$$

Solving (A) and (B) we get—

$$V_q(1) = 1.6$$

$$V_q(2) = 1.9$$

Transition from step (1) to step (2) and step (2) to step (1). give the state value for steps.

optimal policy for actions up and down. from step (1) and (2) respectively.

$$\pi(s) = \text{argmax}(V_q(1), V_q(2))$$

$$\pi(s) = V_q(2)$$

So agent must start from step 2 for single iteration to get max return.

Similarly for iterations > 1, have multiple possibilities.

(1) starting from state (step (1))

(1) → (2) → (1)	(2) → (1) → (2)
(1) → (2) → (F)	(2) → (1) → (G)
(1) → (G) → (G)	(2) → (F) → (F)

Possible state transitions for two time steps.

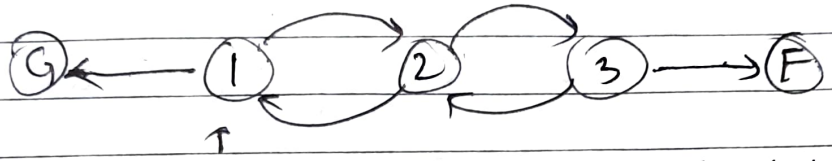
$$S_t \rightarrow S_{t+1} \rightarrow S_{t+2}$$

When no. of steps increases, $n > 2$.,
possibility of initial - starting state increases,

As minimum only (1) iteration is needed to reach terminal states, for each step $n = 2$.

No. of iteration needed to reach terminal state increases..

Eg – $n = 3$



Starts for 1st step, closer to G, only 1 iteration, for, F, ① → ② → ③ → F, 3 iterations needed, Similarly,

No. of iterations increases with n.
Possible no. of transition with 3 iteration.

From ①,

① → ② → ① → G
① → ② → ① → ②
① → ② → ③ → ②
① → ② → ② → F

From ②

② → ① → ② → ①
② → ① → ② → ③
② → ③ → ② → ①
② → ③ → ② → ③

Similarly. for step ③, therefore it increases. w. with n power -

# of iteration $\simeq 2^n + n$ (approx),