Question 3)

Ex 3.15)

NO, sign's are not important ~~only~~ Interval between them is important

$rE \rightarrow$ reward of hitting edge
$rA \rightarrow$ reward for leaving box A or B.
$rB \rightarrow$ " " " B
$rO \rightarrow$ " " other action's

Suitable order
$rE < rO < rA, rB$

$V_* = R_{t+1} + \gamma R_{t+2} \text{———} \gamma^{n-1} R_{t+n}$

$\cancel{V} = R_{t+1} + \gamma R_{t+2} \text{———} \gamma^{n-1} R_{t+n} + c[1 + \gamma + \gamma^2 + \text{———} \gamma^{n-1}]$

$V_c =$
$= V_* + c[1+\gamma+\gamma^2 + \text{———} \gamma^{n-1}]$

$V_c = c[1+\gamma+\gamma^2 \text{———} \gamma^{n-1}]$

$V_c = c + c\gamma + c\gamma^2 + \text{———} + c\gamma^{n-1}$
$-\gamma V_c = c\gamma + c\gamma^2 + \text{———} + c\gamma^n$
$V_c(1-\gamma) = c(\gamma^n - 1)$

$V_c = \frac{c(\gamma^n - 1)}{1-\gamma} \cdot \frac{c}{1-\gamma}$ [since $\lim_{n \to \infty} \gamma^n \to 0$ $|\gamma| < 1$]

Ex 3.16 >

It will change the Maze running
Because in episodic task n is
finite.

from equation's of 3.15

$$\cancel{V_*^{original}} \; V_*^{modified} = V_*^{original} + \frac{c(\gamma^n - 1)}{1-\gamma}$$

so for different state's n will
have different Value's.
Highest for starting state and
lowest for least visited state.

Q 5)

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

$$= \cancel{\sum_a q} \sum_a \pi(a|s) \, q_\pi(s, a)$$

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

Ans 8 >

Yes, R_{t+2} depends of S_t, A_t

Let :- S, R, A be the set of state, Reward action.

S_1', S_2' ——— S_n' ~~be~~ represent state's after jumping to next state from initial state (s) by taking action ~~A~~ a.

$r(R_{t+2} \mid S_t = s, A = a)$

$r(.) \to$ this represent PMF of R_{t+2}

Below expression represent probability being in S_i' from state s after taking action a.

$$p(s_i') = \sum_{r} p(s_i', r \mid s, a)$$

$p(.)$ is PMF. state is random variable

$$E(s_i') = \sum \pi(a \mid s_i') \left(\sum_{s''} \sum_{r} r\, p(s'', r \mid s_i', a) \right)$$

$E(.)$ is ~~PMF~~ Expected reward in state s_i'

$$E[R_{t+2} \mid S_t = s, A_t = a] = \sum_{i=1}^{N} p(s_i') E(s_i')$$

Since $p(s_i')$ depends on S_t and A_t so, R_{t+2} also depend on S_t and A_t.

'N' is total number of states

Ans 9)

$$E[R_{t+2} \mid S_t = s, A_t = a] =$$

$$\sum_{i=1}^{N} p(s_i') E(s_i') \quad \text{from Ans 8.}$$

Ans 10)

$\pi(a|s)$ $a \in A$, $s \in S$ is given

$p(s', r \mid s, a)$ is also given.

$$V_\pi(s) = E(G_t \mid S_t = s)$$

$$= \sum_a \pi(a|s) . E(G_t \mid S_t = s, A_t = a)$$

Solving for $E[G_t \mid S_t = s, A_t = a]$

$$= E[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a]$$

$$= E[R_{t+1} \mid S_t = s, A_t = a] + \gamma E[G_{t+1} \mid S_t = s, A_t = a$$

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

— (ii)

Replace $E[G_t \mid S_t = s, A_t = a]$ by ~~from~~ eq (ii)

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

Ans 11)

$= 2$

$= 0.5(-1) = -0.5$

$= (0.5)^2(10) = 2.5$

$= (0.5)^3(-3) = -0.375$

Let $E(G_t)$ be x

$x = c + \gamma c + \gamma^2 c \text{ --- } \gamma^{n-1} c$ — (i)

$x\gamma = \gamma c + \gamma^2 c + \quad \gamma^n c$ — (ii)

subtract (ii) from (i)

$x(1-\gamma) = c - \gamma^n c$

$$x = E[G_t] = \frac{c(1-\gamma^n)}{1-\gamma}$$

Since $|\gamma| < 1$ and $n \to \infty$ so, $\lim_{n \to \infty} \gamma^n \to 0$

$$E[G_\infty] = \frac{c}{1-\gamma}$$

Ans 12) We have $V_*(s)$ $s \in S$.

for any state s in S.

$$\cancel{\arg\max_{a} \sum_{s'} \sum_{r} r[p(s',r|s,a) + \gamma V_*(s')]}$$

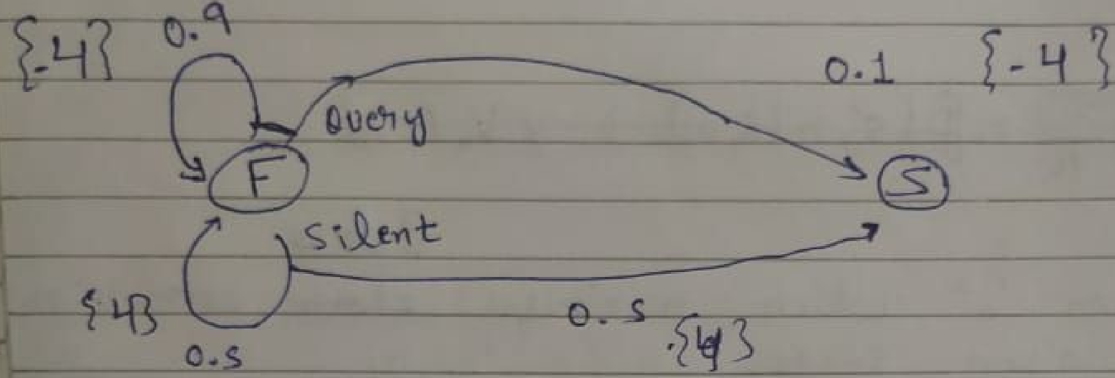
take action 'a' which maximises ~~above~~ Below expression for a given state.

~~if new $V_*(s)$ is greather than old $V_*(s)$ then replace old with new one.~~

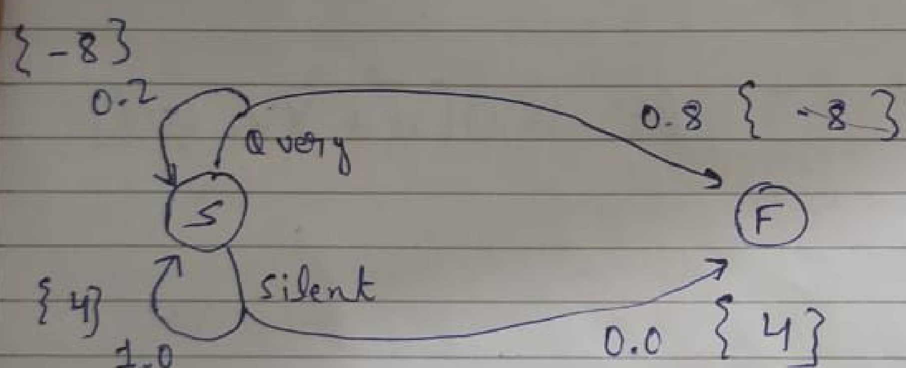
$$\arg\max_{a} \sum_{s'} \sum_{r} p(s',r|s,a)[r + \gamma V_*(s')]$$

Q 13

states = { fresh , stale }



{ I } represent ~~reword~~ reward



cost is converted into reward by changing sign.