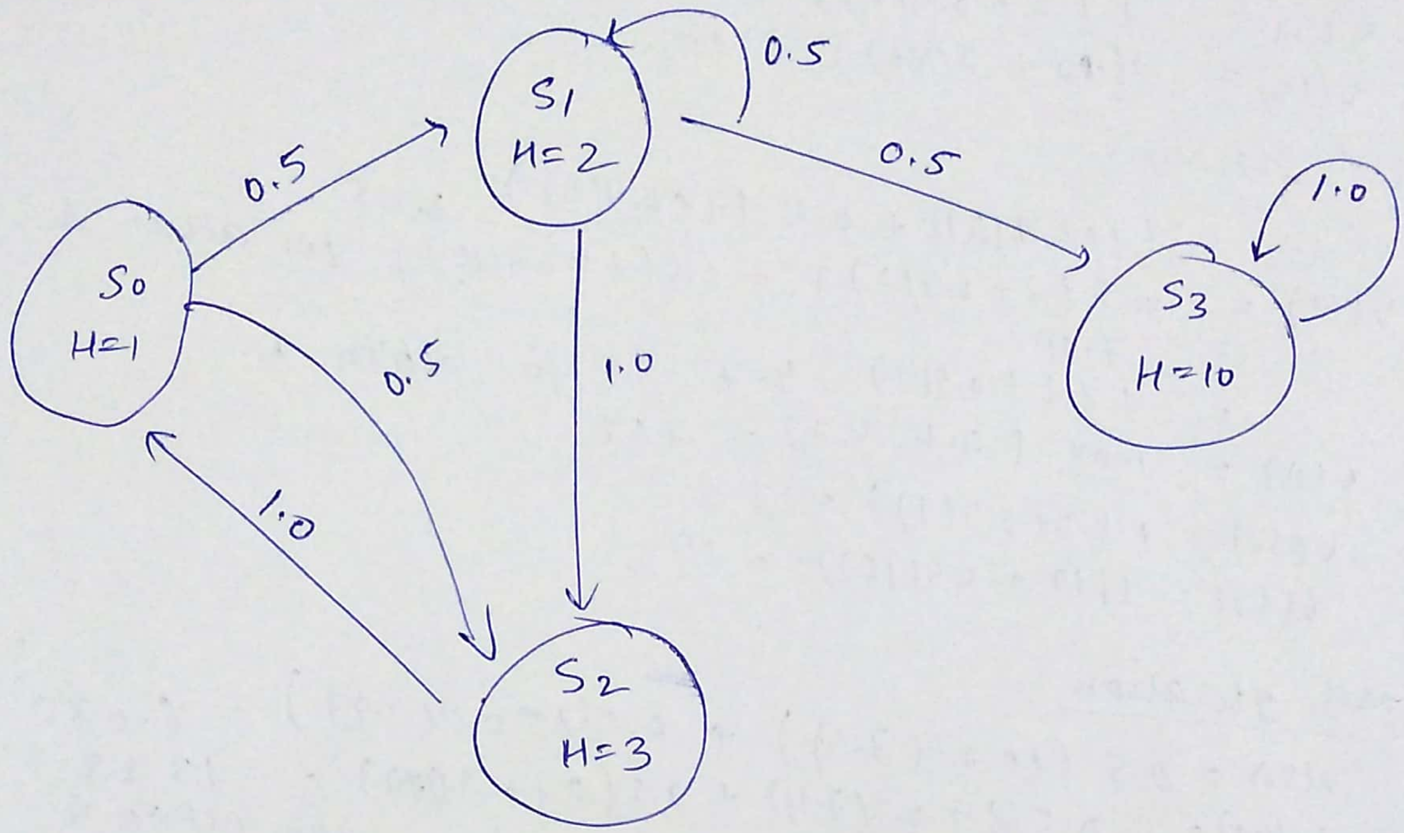
HOMEWORK-4 THEORY

SHUBHAM KUMAR
2015098

QUES 1) a)



Markov Decision Process (MDP)

we are given,
S : set of states
A : set of actions
$T_t(s, a, s') = P(s_{t+1} = s' \mid s_t = s, a_t = a)$
$R_t(s, a, s') =$ reward for $(s_{t+1} = s', s_t = s, a_t = a)$

our goal: find a policy (π) that maximises the expected sum of rewards, i.e.

$$\pi^* = \arg\max_{\pi} E\left[\sum_{t=0}^{H} R_t(S_t, A_t, S_{t+1}) \mid \pi\right]$$

now, we'll see how changes occur at the different iterations:- ②

1st iteration

$V(s0) = 0.5(1 + 0.9(0)) + 0.5(1 + 0.9(0)) = 1$
$V(s1) = 0.5(2 + 0.9(0)) + 0.5(2 + 0.9(0)) = 2$
$V(s2) = 1(3 + 0.9(0)) = 3.$
$V(s3) = 1(10 + 0.9(0)) = 10.$

2nd iteration

$V(s0) = 0.5(1 + 0.9(2)) + 0.5(1 + 0.9(3)) = 3.25$
$V(s1) = 0.5(2 + 0.9(2)) + 0.5(2 + 0.9(10))$ for action a
$= 7.4$
$= 1(2 + 0.9(3)) = 4.7$ for action b.
$V(s1) = \max(7.4, 4.7) = 7.4$
$V(s2) = 1(3 + 0.9(1)) = 3.9$
$V(s3) = 1(10 + 0.9(10)) = 19.$

3rd iteration

$V(s0) = 0.5(1 + 0.9(7.4)) + 0.5(1 + 0.9(3.9)) = 6.085$
$V(s1) = 0.5(2 + 0.9(7.4)) + 0.5(2 + 0.9(19)) = 13.88$ for action a
$= 1(2 + 0.9(3.9)) = 5.51$ for action b.
$= \max(13.88, 5.51) = 13.88$
$V(s2) = 1(3 + 0.9(3.25)) = 5.925.$
$V(s3) = 1(10 + 0.9(19)) = 27.1$

ques 1) b) Find qvalue!

For action a: $0.5(2 + 0.9(13.88)) + 0.5(0 + 0.9(27.1))$
$= 20.441$

For action b: $1(2 + 0.9(5.925))$
$= 7.3325$

so, ACTION-A is the optimal policy.

c) i) FALSE
convergence will not be there due to cyclic MDP. 3.

ii) FALSE
MDP will never converge ($\gamma = 1$) :-

$$\| V^*_{i+1} - V^* \| < 2\epsilon\gamma \,/\, (1-\gamma) = 0$$

apply $\gamma = 1$, then $1-\gamma$ becomes 0
now, this is not possible, agent will look for reward at ∞ distance.

iii) TRUE
next state = Reward, $V^*_{i+1} = \sum_{max} T[R(s, a, s')]$
Hence, it will converge.

iv) TRUE
A cyclic MDP converges for any $\gamma \in (0,1)$

v) FALSE
There is no end point and noise.

QUES 2.
No of uncompressed bits $= 24N^2$
no of bits reqd by receiver $= N^2 \lg k + 24k$ (k: no of clusters)

$$\therefore \text{compression Ratio} = \frac{24N^2}{N^2 \log_2 k + 24k}$$