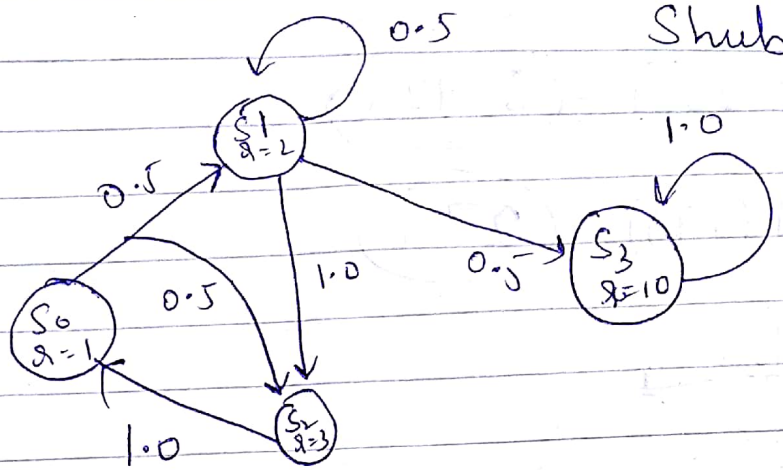
ML Assignment - 4

Shubham Khanna
2015179

Q.1



(a) $V(s') = 0, \forall s' \in \{S_0, S_2, S_3, S_4\}$

for Iter → 1

$V(S_0) = 0.5(1 + 0.9(0)) + 0.5(1 + 0.9(0)) = 1$

$V(S_1) = 0.5(2 + 0.1(0) + 2 + 0.9(0)) = 2$

$V(S_2) = 1(3 + 0.9(0)) = 3$

$V(S_3) = 1(10 + 0.9(0)) = 10$

for Iter → 2

$V(S_0) = 0.5[1 + 0.9(2)) + 0.5(1 + 0.9(3)) = 3.25$

$V(S_1) = \text{Max}(0.5(2 + 0.9(2)) + 0.5(2 + 0.9(10)), 1(2 + 0.9(3))$

$= \text{Max}(4.7, 7.4) = 7.4$

$V(S_2) = 1(3 + 0.9(1)) = 3.9$

$V(S_3) = 1(10 + 0.9(10)) = 19$

for Iter → 3

$V(S_0) = 0.5(1 + 0.9(7.4)) + 0.5(1 + 0.9(3.9))$

$= 6.085$

$V(S_1) = \text{Max}(0.5(2 + 0.9(7.4)) + 0.5(4.7), 1(2 + 0.9(3.9))$

$= \text{Max}(13.88, 5.51) = (13.88)$

$V(S_2) = 1(3 + 0.9(3.25)) = (5.925)$

$V(S_3) = 1(10 + 0.9(19)) = (27.1)$

(b) Optimal Policy for S_1

$$\underset{A}{\text{argmax}} \left(0.5(2 + 0.9(13.88)) + 0.5(2 + 0.9(27.1)), \right.$$
$$\left. 1(2 + 0.9(5.925)) \right)$$

$= \text{argmax}(20.44, 7.3325)$

$\therefore$ first

$\therefore$ We choose actions to form optimal policy.

(c) (i) False : A cyclic MDP cannot converge

(ii) False : for $\gamma = 1$, MDP won't converge ever since agent will try to look for higher reward till infinity. Value

(iii) True as the for next State will be equal to only the estimate of max Reward.

$$V_i' = \max \sum \left(T_{s,a,s'} (R(s,a,s')) \right)$$

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

(v) False There is no exit & noise.

(2) No. of uncompressed bits = $N^2 * 8 \times 3$

$= 24N^2$

No. of compressed " $= N^2 \times \log_2 k + 24k$

$\therefore$ Compression ratio $= \frac{24N^2}{N^2(\log_2 k + 24k)}$

where k = no. of clusters