Name - T. Vijay

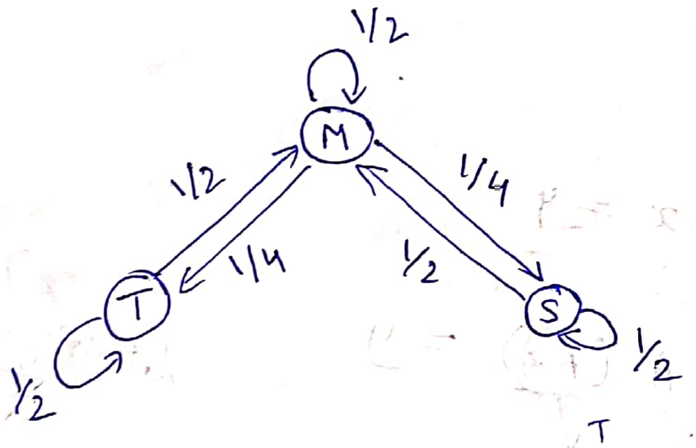
Roll.no - CS17BTECH11040

Assignment No.1

Q1

a) States S: Tall (TT), Medium (TS or ST), short (SS)

Fig: Transition Probabilities



Transition probability Matrix P =

$$\begin{array}{c|c} & \begin{array}{ccc} T & M & S \end{array} \\ \begin{array}{c} T \\ M \\ S \end{array} & \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$

b) We know that ~~step~~ n^{th} step transition matrix is given by P^n

$\therefore$ Probabilities of Tall, short, medium offspring belonging to first generation = second row of P matrix

$$= \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$$

(Tall) (Medium) (Short)

Second generation

$$p^2 = \begin{bmatrix} 0.375, & 0.5, & 0.125 \\ 0.25, & 0.5, & 0.25 \\ 0.125, & 0.5, & 0.375 \end{bmatrix} \longrightarrow \text{Second row}$$

$\therefore P_{Tall} = 0.25$, $P_{medium} = 0.5$, $P_{short} = 0.25$

Similarly for Third generation

$P_{Tall} = 0.25$, $P_{medium} = 0.5$, $P_{short} = 0.25$

c) Second row of $P^n \ \forall n \in N = [0.25, 0.5, 0.25]$

$\therefore P_{Tall} = 0.25$, $P_{medium} = 0.5$, $P_{short} = 0.25$

Q2

(a) States: S, 1, 3, 5, 6, 7, 8, W [~~other~~ We can't stay on other states]

Transition Matrix

$$\begin{array}{c} \\ S \\ 1 \\ 3 \\ 5 \\ 6 \\ 7 \\ 8 \\ W \end{array} \begin{array}{c} \begin{array}{cccccccc} S & 1 & 3 & 5 & 6 & 7 & 8 & W \end{array} \\ \begin{bmatrix} 0 & 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 & 0 \\ 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 0 & 0 & 0 & 1/2 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

(b) W is the absorbing state as we cannot leave state W after entering it.

(c) We take the reward = -1 at state, so the final value of each state will indicate the no. of ~~steps~~ expected steps but the value will be negative.

From Bellman eq^n, we have $V = R + \delta P V$

$\therefore V = (I - \delta P)^{-1} R$

We take $\delta = 1$, P = Probability transision matrix
(Discount factor) (We will exclude the last row and column from matrix obtained)

$R = [-1, -1, -1, -1, -1, -1, -1]^T_{1\times 7}$

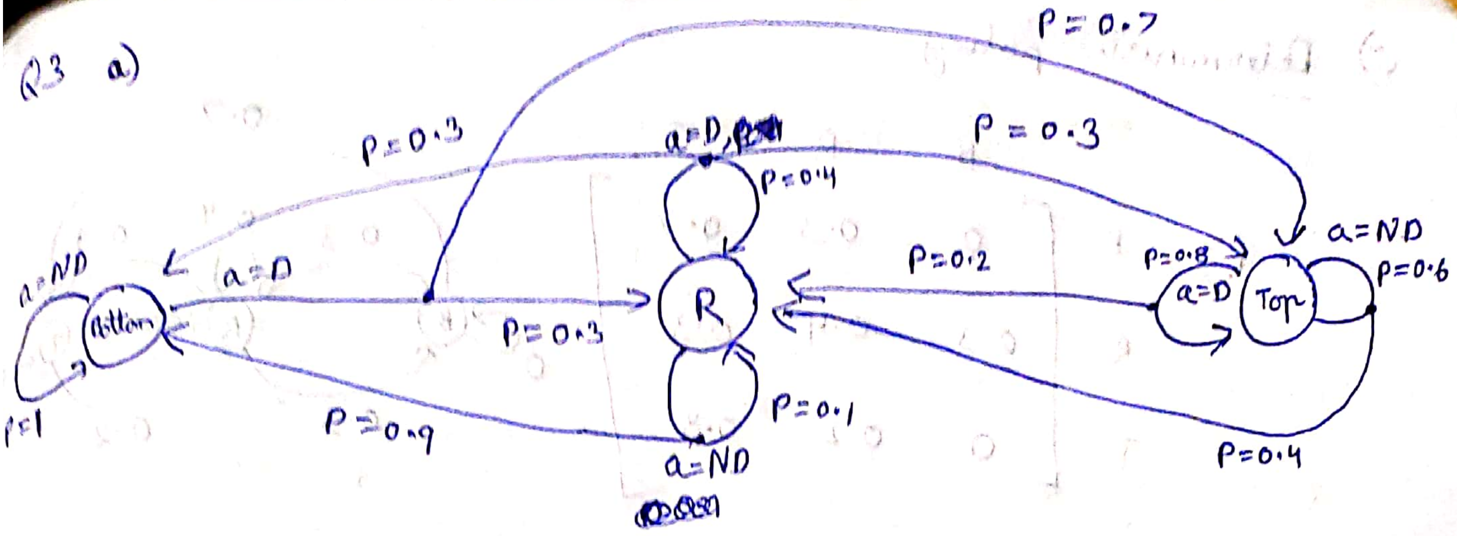
$P_{7\times 7}$ ← Excluding last row and column from P (Q2 a)

$\therefore V = [7.083, 7, 6.67, 6.67, 5.33, 5.33, 5.33]^T$

$\therefore$ V(i) represents the no. of steps from state i

Scanned by CamScanner

Q3 a)



Actions: D → Driving

ND → not Driving

b) Deterministic policy

$$\pi(s) = \begin{cases} \text{Drive}, & s = \text{top} \\ \text{Drive}, & s = \text{rolling} \\ \text{Drive}, & s = \text{Bottom} \end{cases}$$

Stochastic policy

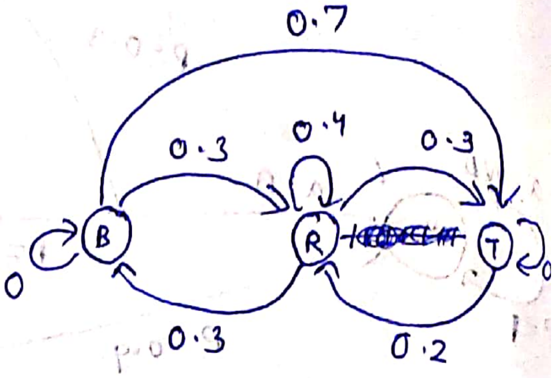
$$\pi(a \mid \text{bottom}) = \begin{cases} 0.5, & a = D \\ 0.5, & a = ND \end{cases}$$

$$\pi(a \mid \text{Top}) = \begin{cases} 0.5, & a = D \\ 0.5, & a = ND \end{cases}$$

$$\pi(a \mid \text{Rolling}) = \begin{cases} 0.5, & a = D \\ 0.5, & a = ND \end{cases}$$

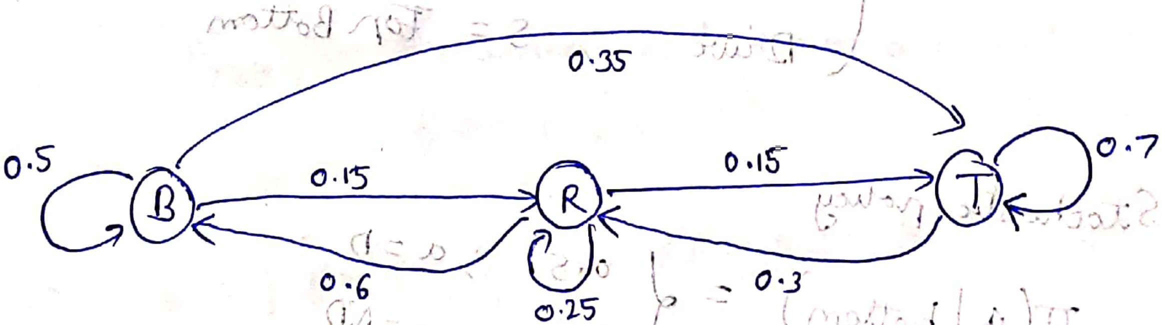
c) Deterministic policy

$$P = \begin{array}{c} \\ B \\ R \\ T \end{array} \begin{array}{c} \begin{array}{ccc} B & R & T \end{array} \\ \begin{bmatrix} 0 & 0.3 & 0.7 \\ 0.3 & 0.4 & 0.3 \\ 0 & 0.2 & 0.8 \end{bmatrix} \end{array}$$



~~State~~ Stochastic policy

$$P = \begin{array}{c} \\ B \\ R \\ T \end{array} \begin{array}{c} \begin{array}{ccc} B & R & T \end{array} \\ \begin{bmatrix} 0.5 & 0.15 & 0.35 \\ 0.6 & 0.25 & 0.15 \\ 0 & 0.3 & 0.7 \end{bmatrix} \end{array}$$



d) $\pi(a \mid S_t = S, S_{t-1} = S) = \begin{cases} 0.5 & a = \text{Drive} \\ 0.5 & a = \text{Not driving} \end{cases}$

$\pi(a \mid S_t = S, S_{t-1} \neq S) = \begin{cases} 0.7 & a = D \\ 0.3 & a = ND \end{cases}$

$(n-p) \times q$

$n - p \geq q$

Q4 a)

$$P^{\pi_1} = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & 0 & 0.9 & 0.1 & 0 \\ b & 0.1 & 0 & 0 & 0.9 \\ c & 0.9 & 0 & 0 & 0.1 \\ d & 0 & 0 & 0 & 1 \end{array}$$

$$P^{\pi_2} = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & 0 & 0.1 & 0.9 & 0 \\ b & 0.9 & 0 & 0 & 0.1 \\ c & 0.1 & 0 & 0 & 0.9 \\ d & 0 & 0 & 0 & 1 \end{array}$$

$$P^{\pi_3} = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & 0 & 0.42 & 0.58 & 0 \\ b & 0.1 & 0 & 0 & 0.9 \\ c & 0.1 & 0 & 0 & 0.9 \\ d & 0 & 0 & 0 & 1 \end{array}$$

$R^{\pi} = [-10 \quad -10 \quad -10 \quad 100]^T$ for all policies because reward for each action is same.

Let ~~δ~~ $\gamma = 0.9$

$$V^{\pi} = (I - \gamma P^{\pi})^{-1} R^{\pi}$$

a)

$V^{\pi_1} = [63.2 \quad 84.69 \quad 51.08 \quad 109.89]^T$

$V^{\pi_2} = [63.2 \quad 51.08 \quad 84.69 \quad 109.89]^T$

$V^{\pi_3} = [66.49 \quad 84.99 \quad 84.99 \quad 109.89]^T$

b) π_3 is the best policy because for every state V^{π_3} has better values than the values of other policies

c) π_1 and π_2 are not comparable because values of V^{π_1} is greater than V^{π_2} for some states while V^{π_2} is greater than V^{π_1} for some other states

$\therefore$ π_1 and π_2 cannot be compared.

Q5 We will use value iteration to find the optimal value function

Step I : We initialize $V_1(s)$ for $s \in S$, to small value ϵ

Then for all states, we find

$$V_{k+1}(s) = \max_a \left[\sum_{s' \in S} P_{ss'}^{a} \left(R_{ss'}^{a} + \gamma V_k(s') \right) \right]$$

If $(|V_{k+1}(s) - V_k(s)| < \epsilon$ $\forall s \in S$

we go to the next step

else we repeat the previous step

Step ②: For all state $s \in S$

$$\pi_*(s) = \underset{a}{\operatorname{argmax}}\ V_*(s)$$

After using the above process we will get the following optimal function

$$\pi(a|s) = \pi_s(a|s) \quad \forall a \ \forall s$$

Intitution

① The above value function is optimal because for each state s, we consider the corresponding optimal state function and then find an optimal policy that achieve this value function and use the probabilities of action associated at this state

② Another Intitution is that if we apply policy iteration on the above policy, we won't get any better policy because the current policy achieves the max value at state. Hence it can't be improved.

Q6. a) The optimal policy is to go right at each state.

The optimal value function has value 10 at each state because $\gamma = 1$ (far-sighted)

b) $\gamma = 0.9$ $V = [5.9, 6.56, 7.29, 8.1, 9, 10]$

$\gamma = 0.5$ $V = [0.31, 0.62, 1.25, 2.5, 5, 10]$

$\gamma = 0.1$ $V = [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10]$

The optimal policy still remains the same i.e to go right at each step.

Decrease in Discount factor leads to decrease in optimal value but ~~optimal~~ value of each state increases as we go to the right.

c) For all non-positive values of C, the optimal policy will still remain the same i.e to go right because we don't want to accumulate negative rewards.

When $1 \leq C \leq 10$, the optimal policy is to move right when ~~[illegible]~~ γ is small.

when $c > 10$, then the optimal policy is to ~~move from S_1 to S_5 and then back~~

turn left when the agent reaches state S_5.

This will lead to an infinite process.

d) The new policy

$$\hat{V}^{\pi} = (1-\gamma P)^{-1}(R+C)$$

where $C = [c \quad c \quad c \; - - -]^{T}_{1\times n}$

$$\therefore \hat{V}^{\pi} = V^{\pi} + (1-\gamma P)^{-1} C$$

where V^{π} is value function for any policy π.

Q7

a) * When γ is low we need to prefer close exit as we are not far sighted

* When γ is high we will prefer distant exit as we are far sighted

* When η is high, we will avoid cliff because we don't have the control over the path taken

* When η is low, we can risk the cliff because we have control over the our actions

① Prefer close exit but risk cliff

→ $\gamma = 0.1$ and noise $= 0$

② Prefer distant exit but risk cliff

→ $\gamma = 0.9$ and noise $= 0$

③ Prefer close exit by avoiding the cliff

→ $\gamma = 0.1$ and noise $= 0.5$

④ Prefer the distant exit by avoiding the cliff

→ $\gamma = 0.9$ and noise $= 0.5$

Q 8 Given : $L(V) = \max_{a \in A} [R^a + \gamma P^a V]$ —— ①

Since , V_* is fixed point of operator L, we have

$L(V^*) = V^*$

and,

$|V_{k+1} - V^*|_\infty = |L(V_k) - L(V^*)|_\infty$ (From ①)

$= |\max_a \{R + \gamma P^a V_k\} - \max_a \{R + \gamma P^a V^*\}|_\infty$

$\leq \max_a |\{R + \gamma P^a V_k\} - \{R + \gamma P^a V^*\}|_\infty$

$= \gamma \, |P^a (V_k - V^*)|_\infty$

$\leq \gamma \, |V_k - V^*|_\infty$

i.e $|V_{k+1} - V^*|_\infty \leq \gamma \, |V_k - V^*|_\infty$

and recursively $|V_{k+1} - V^*|_\infty \leq \gamma^k |V_1 - V^*|_\infty$

Hence Proved

Scanned by CamScanner