Topics in Reinforcement learning
(Quiz 1, 45 mins)

Each Question is for 10 marks

1. Obtain the stationary distribution for a Markov chain with state space $S = \{1,2\}$ with the following transition probability matrix $P = \begin{bmatrix} 0 & 1 \\ 0.1 & 0.9 \end{bmatrix}$. Now suppose you get a reward of 10 everytime you visit state 1 and reward of 0 otherwise. Assuming a discount factor $\beta = 0.5$, obtain the value function ($V(s)$ for $s = 1, 2$) associated with this Markov reward process.

2. Consider a finite horizon, finite state and finite action MDP. Prove the following Bellman Optimality Equation

$$V_t(s) = \max_{a \in \mathcal{A}} \{r(s,a) + E_{s,a}[V_{t+1}(S')]\}.$$

   Using the above, write the Bellman Optimality equations in terms of the state action value function $Q_t(s,a)$. Give justification.

3. Find the shortest path (Figure 1 next page) from node 1 to node 8 using deterministic dynamic programing. Number on the edge denotes cost of that edge. Identify $V_t(s_t)$ for all t and applicable s_t and use it to obtain the shortest path, i.e., $V(1)$. Also obtain $\pi^* = (\pi_1^*, \pi_2^*, \dots \pi_4^*)$

4. Consider the MDP as in Fig 1 (next page) for a time horizon of 3 units, i.e., $t = 0, 1, 2$. Identify $V_t(s_t)$ for all t and s_t and use it to obtain $V(s_1)$ and $V(s_2)$. Also obtain π^*.

Figure 1: Shortest path problem

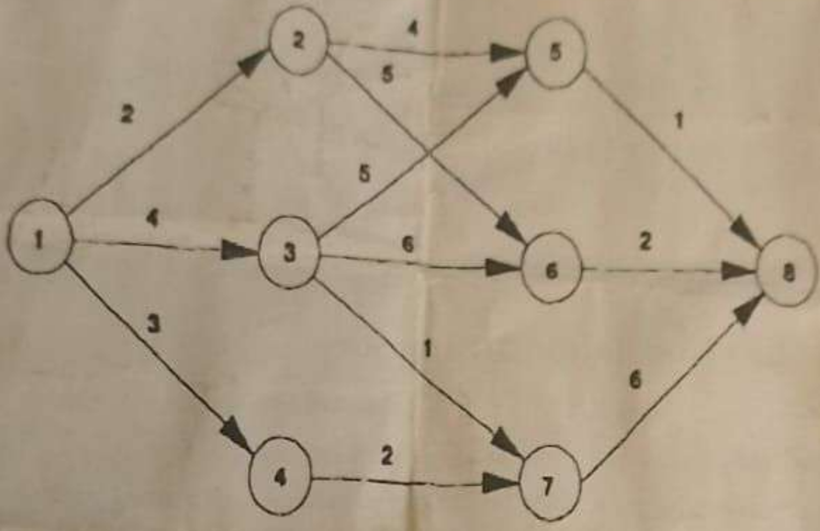

Figure 2: 2 state MDP