__/__/____

Quiz-1.

1.
$$\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.7 & 0.5 & 0.4 \\ 0.6 & 0.8 & 0.9 \end{bmatrix}$$

$P(S_{t+1} = S1 \mid S_t = S_3) = 0.6.$
$P(S_{t+1} = S2 \mid S_t = S_3) = 0.8$
$P(S_{t+1} = S2 \mid S_t = S_1, S_{t-1} = S_1) =$
Current & previous state S_1. $= 0.2.$

2. The future state depends only on the current state and the current action, not on the sequence of states and actions that preceded it.
∴ a) S_t
c) A_t.

3. The numerical signal that the agent receives from the environment at each step is called the "reward" whereas "return" is the total reward the agent receives over a long run.
Long term return can be expressed as a sum of individual rewards.
$$G_t = R_{t+1} + R_{t+2} + \dots + R_t.$$

4. As γ approaches 1, return objective takes future rewards into account more strongly.

As γ is set to 0, future rewards are completely ignored in the calculation of long-term return.

5. In episodic tasks, the simulation will come to an end. If γ is set to 0, agent only considers immediate rewards. As there is a definite start and end to the task, γ needn't always be 0.

Eg: Navigation in a Maze.

6. $G_t = R_{t+1} + R_{t+2} + \ldots R_t$.

As $R_{t+1}, R_{t+2} \ldots R_t$ are random variables, the stochastic nature of rewards by State transitions contributes to the randomness of G_t.
$\therefore G_t$ is a random variable.