1. Write out the MP/MRP/MDP/Policy definitions and MRP/MDP Value Function definitions in your own style/notation

- Markov Process

> **Markov Process** is a tuple $\langle S, P \rangle$. S is a finite set of states and P is the state transition probability matrix where entry $P_{ss'} = \mathbb{P}[S_{t+1} = s' | S_t = s]$

- Markov Reward Process

> **Markov Reward Process** is a tuple $\langle S, P, \mathcal{R}, \gamma \rangle$. S and P hold the same definition as in MP, $\mathcal{R}$ is a reward function of states: $\mathcal{R}(s) = \mathbb{E}[R_{t+1} | S_t = s]$. Note that $\mathcal{R}$ is time invariant. $\gamma \in [0, 1]$ is discount factor

- Value Function in MRP

> **Value function** with respect to state s in MRP is the expected return starting from that state: $v(s) = \mathbb{E}[G_t | S_t = s]$, where return G_t is the total discounted reward starting from time t: $G_t = \sum_{i=0}^{\infty} \gamma^i R_{t+i+1}$ given a state sequence $\{S_t, S_{t+1}, \cdots\}$.

In order to compute $v(s)$, we apply the Bellman equation:

$$\begin{aligned} v(s) &= \mathbb{E}[\sum_{i=0}^{\infty} \gamma^i R_{t+i+1} | S_t = s] \\ &= \mathbb{E}[R_{t+1} | S_t = s] + \mathbb{E}[G_t - R_{t+1} | S_t = s] \\ &= R(s) + \gamma \sum_{s' \in S} P_{ss'} v(s') \end{aligned}$$

Concatenate all states in a vector(e.g. $V = [v(s_1), v(s_2), \cdots]^\intercal$) we can rewrite the above equation as:

$$V = R + \gamma P V$$

- Markov Decision Process

Markov Decision Process is a tuple $\langle S, A, P, \mathcal{R}, \gamma \rangle$, where each element in the tuple is defined as follows:

- S: a finite set of states
- A: a finite set of actions
- P: state transition probability matrix, $P^a_{ss'} = \mathbb{P}[S_{t+1} = s' | S_t = s, A_t = a]$
- R: reward function, $R^a_s = \mathbb{E}[R_{t+1} | S_t = s, A_t = a]$
- γ: discount factor

- Policy

Policy in MDP is a distribution over actions given states: $\pi(a|s) = \mathbb{P}(A_t = a | S_t = s)$

- Value Function in MDP
  - **State Value function** in MDP is the expected return from state s and follow policy π: $v_\pi(s) = \mathbb{E}_\pi[G_t | S_t = s]$
  - **Action Value function** in MDP is the expected return from state s, apply action a and then follow policy π: $q_\pi(s, a) = \mathbb{E}_\pi[G_t | S_t = s, A_t = a]$

Intuitively we can get one value function from another:

$$v_\pi(s) = \sum_{a \in A} \pi(a|s) q_\pi(s, a)$$

$$q_\pi(s, a) = R^a_s + \sum_{s' \in S} P^a_{ss'} v_\pi(s')$$

Using the above equations we can get the Bellman equation for state value function and action value function (individually):

$$v_\pi(s) = \sum_{a \in A} \pi(a|s) \left(R^a_s + \sum_{s' \in S} P^a_{ss'} v_\pi(s') \right)$$

$$q_\pi(s, a) = R^a_s + \sum_{s' \in S} P^a_{ss'} \left(\sum_{a' \in A} \pi(a'|s') q_\pi(s', a') \right)$$