Bellman Equation

for a Markov Reward Process (MRP)

The Bellman Equation for MRPs:

$$V(s) = R(s) + \gamma \sum_{s'} P(s'|s)V(s')$$

Where:

- v(s) is the value function of state s, representing the expected discounted future reward.
- R(s) is the immediate reward received at state s.
- γ is the discount factor, determining the importance of future rewards relative to immediate ones.
- p(s′ ∣ s) is the transition probability from state s to state s′.

For MDP

Prediction

State-Value Function

The state-value function, denoted as π(a∣s), represents the expected discounted future reward starting from state S and following a given policy π. **In other words, it quantifies how good it is to be in state S and follow policy π thereafter.**

Bellman Expectation Equation

The Bellman Expectation Equation for the state-value function is

$$v_\pi(s) = \sum_{a \in A} \pi(a|s) \left(R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_\pi(s') \right)$$

Where:

- v(s): This is the **value function** for state (s) under policy (\pi)

- π(a|s) This is the **policy** (\pi), which gives the probability of taking action (a) when in state (s).

- R(s|a): This is the **immediate reward** received after taking action (a) in state (s).

- γ: This is the **discount factor**, a number between 0 and 1 that determines the importance of future rewards.

- p(ss′|a): This is the **transition probability**, which represents the probability of moving from state (s) to state (s′) after taking action (a).\

- v(s′): This is the **value function** for the next state (s′) under policy π, indicating the expected return from state (s′).

Action-Value Function:

The action-value function, represents the expected discounted future reward starting from state , taking action A, and then following policy . It quantifies how good it is to be in state , take action , and follow policy thereafter.

Bellman Expectation Equation for Action-Value Function:

The Bellman Expectation Equation for the action-value function is:

$$q_\pi(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_\pi(s')$$

Where:

- q(s, a): Represents the action-value function.
- R(s|a): This is the **immediate reward** received after taking action (a) in state (s).
- γ: This is the **discount factor**, a number between 0 and 1 that determines the importance of future rewards.
- p(ss′|a): This is the **transition probability**, which represents the probability of moving from state (s) to state (s′) after taking action (a).
- v(s′): This is the **value function** for the next state (s′) under policy π, indicating the expected return from state (s′).

Action-Value Function with Respect to Itself

$$q_\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \sum_{a' \in A} \pi(a'|s') q_\pi(s',a')$$

Control

Bellman Optimality Equation for Control

$$V_\pi(s) = \max_a q_\pi(s,a)$$