FPT Education
FPT UNIVERSITY

Bellman Equation

Objectives

- Derive the Bellman equation for state value functions
- Define the Bellman equation for action value functions
- Understand how Bellman equations relate current and future values.

Bellman Equation

- The Bellman equation is a fundamental concept in dynamic programming and reinforcement learning.
- It expresses the relationship between the value of a state (or state-action pair) and the value of its successor states.
- The Bellman equation plays a crucial role in many RL algorithms, as it provides a recursive definition for computing value functions.

Bellman Equation Types

- Bellman Expectation Equation:
  - The Bellman expectation equation expresses the relationship between the value of a state (or state-action pair) and the expected immediate reward plus the discounted value of the successor states.
  - $$V^{\pi}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma V^{\pi}(s')]$$
  - p(s′,r|s,a) is the probability of transitioning to state s′and receiving reward r when taking action a in state s, and π(a|s) is the policy's probability of selecting action a in state s. γ is the discount factor which determines the importance of

Bellman Equation Types

- Bellman Expectation Equation:
  - The Bellman expectation equation expresses the relationship between the value of a state (or state-action pair) and the expected immediate reward plus the discounted value of the successor states.
  - $$Q^{\pi}(s,a) = \sum_{s',r} p(s',r|s,a)\left[r + \gamma \sum_{a'} \pi(a'|s') Q^{\pi}(s',a')\right]$$
  - p(s′,r|s,a) is the probability of transitioning to state s′and receiving reward r when taking action a in state s, and π(a|s) is the policy's probability of selecting action a in state s. γ is the discount factor which determines the importance of

Bellman Equation Types

- Bellman Optimality Equation:
  - The Bellman optimality equation expresses the optimal value of a state (or state-action pair) in terms of the maximum expected immediate reward plus the discounted value of the successor states.
  - For the state value function V∗(s), it is defined as:

$$V^*(s) = \max_a \sum_{s',r} p(s',r|s,a)[r + \gamma V^*(s')]$$

  - V∗(s) represents the optimal value of state s under the optimal policy

Bellman Equation Types

- Bellman Optimality Equation:
  - The Bellman optimality equation expresses the optimal value of a state (or state-action pair) in terms of the maximum expected immediate reward plus the discounted value of the successor states.
  - For the action value function Q∗(s,a), it is defined as:

$$Q^*(s,a) = \sum_{s',r} p(s',r|s,a)[r + \gamma \max_{a'} Q^*(s',a')]$$

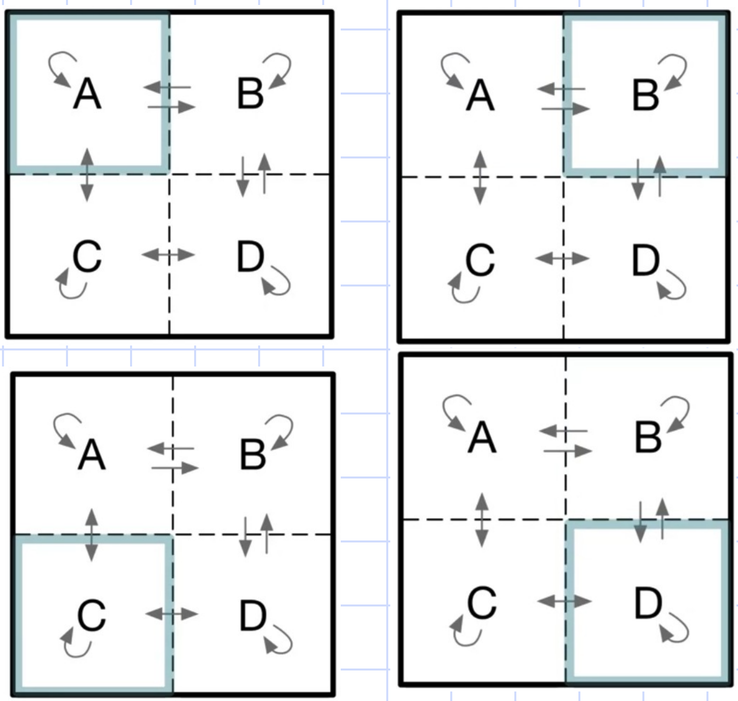
  - Q∗(s,a) represents the optimal value of taking action a in state s under the optimal policy.

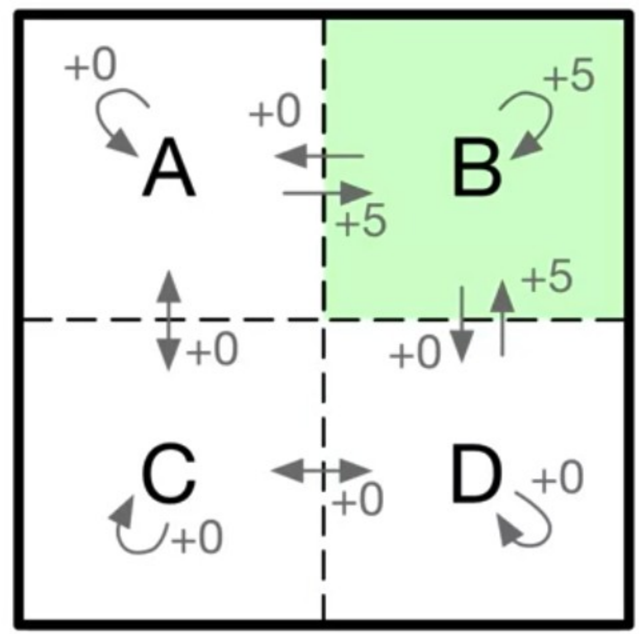
Bellman Equation

- It define how the value functions relate to each other and to the dynamics of the environment.
- RL algorithms leverage these equations to iteratively improve value function estimates and derive optimal policies.
- Bellman equations to compute value functions.

Bellman Equation

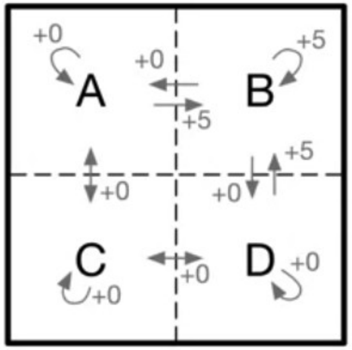
- Example
  - Start from C → A→ B→ D. The reward is 0 everywhere except for any time the agent lands in state B.



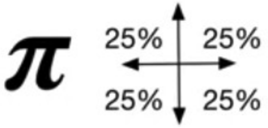


Bellman Equation

- Example
  - Using the Bellman equation, we can write down an expression for the value of state A in terms of the sum of the four possible actions and the resulting possible

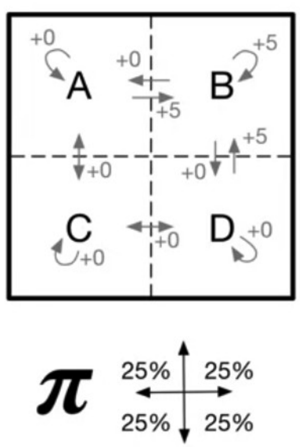


$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$



Bellman Equation

- Example
  - The expression further in this case, because for each action there's only one possible associated next state and reward.
  - That's the sum over s prime and r reduces to a single value (s prime and r do still depend on the selected action, and the curre

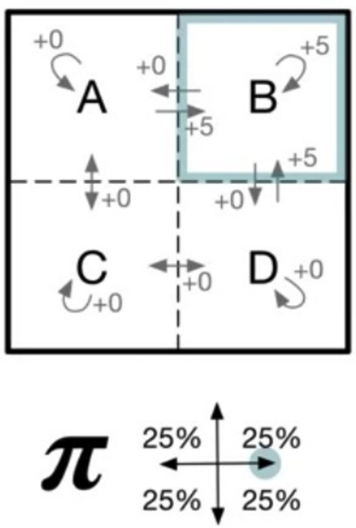


$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = \sum_a \pi(a \mid A)\left(r + 0.7 V_\pi(s')\right)$$

Bellman Equation

- Example
  - If we go right from state A, we land in state B, and receive a reward of +5. This happens one quarter of the time under the random policy.



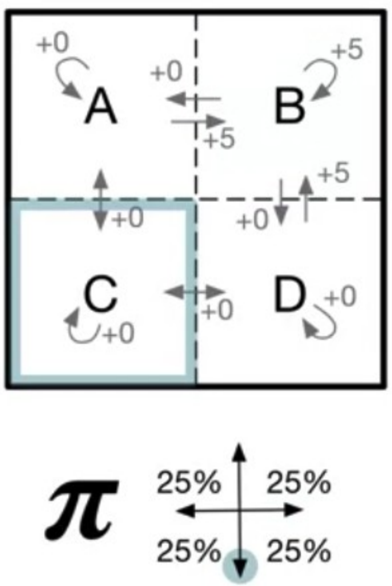
$$V_\pi(s) = \sum_a \pi(a|s) \sum_r \sum_{s'} p(s',r|s,a)[r+\gamma V_\pi(s')]$$

$$V_\pi(A) = \sum_a \pi(a|A)\big(r + 0.7 V_\pi(s')\big)$$

$$V_\pi(A) = \frac{1}{4}(5 + 0.7V_\pi(B))$$

Bellman Equation

- Example
  - If we go down, we land in state C, and receive no immediate reward. → this occurs one-quarter of the time



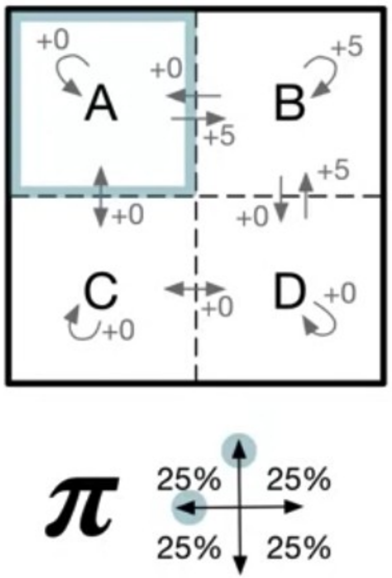
$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = \sum_a \pi(a \mid A)\left(r + 0.7 V_\pi(s')\right)$$

$$V_\pi(A) = \frac{1}{4}(5 + 0.7 V_\pi(B)) + \frac{1}{4} 0.7 V_\pi(C)$$

Bellman Equation

- Example:
  - If you go either up or left, we will land back in state A again. Each of the actions, up and left, again, occur one-



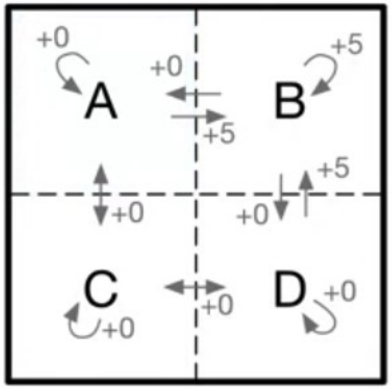
$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = \sum_a \pi(a \mid A)\left(r + 0.7V_\pi(s')\right)$$

$$V_\pi(A) = \frac{1}{4}(5 + 0.7V_\pi(B)) + \frac{1}{4}0.7V_\pi(C) + \frac{1}{2}0.7V_\pi(A)$$

Bellman Equation

- Example
  - Finally, we arrived at the expression shown here for the value of state A.

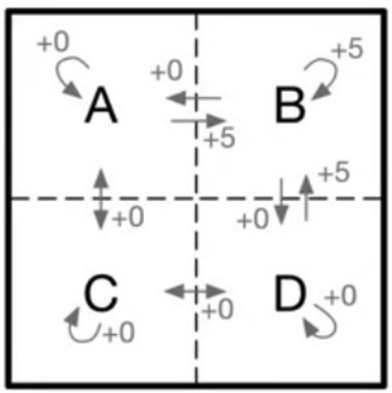


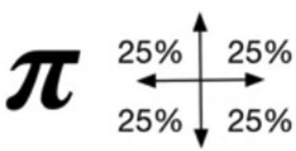
$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = \frac{1}{4}(5 + 0.7V_\pi(B)) + \frac{1}{4}0.7V_\pi(C) + \frac{1}{2}0.7V_\pi(A)$$

Bellman Equation

- Example
  - Equation for each of the other states, B, C, and D.





$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = \frac{1}{4}(5 + 0.7V_\pi(B)) + \frac{1}{4}0.7V_\pi(C) + \frac{1}{2}0.7V_\pi(A)$$

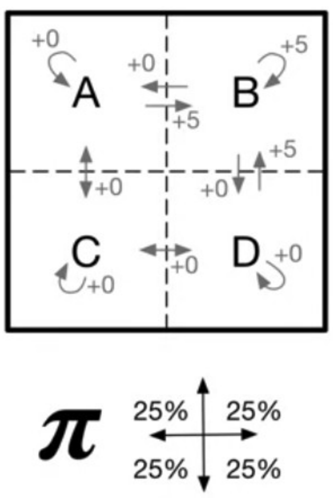
$$V_\pi(B) = \frac{1}{2}(5 + 0.7V_\pi(B)) + \frac{1}{4}0.7V_\pi(A) + \frac{1}{4}0.7V_\pi(D)$$

$$V_\pi(C) = \frac{1}{4}0.7V_\pi(A) + \frac{1}{4}0.7V_\pi(D) + \frac{1}{2}0.7V_\pi(C)$$

$$V_\pi(D) = \frac{1}{4}(5 + 0.7V_\pi(B)) + \frac{1}{4}0.7V_\pi(C) + \frac{1}{2}0.7V_\pi(D)$$

Bellman Equation

- Example
  - The unique solution is shown here.
  - Bellman equation reduced an unmanageable infinite sum over possible futures, to a simple linear algebra problem.



$$V_\pi(s) = \sum_a \pi(a \mid s) \sum_r \sum_{s'} p(s', r \mid s, a)[r + \gamma V_\pi(s')]$$

$$V_\pi(A) = 4.2$$

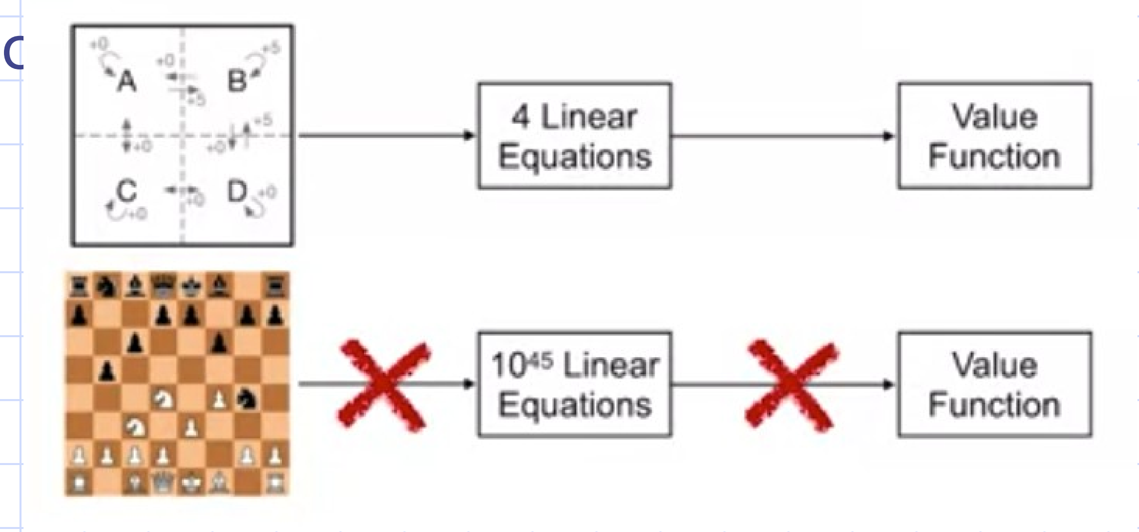
$$V_\pi(B) = 6.1$$

$$V_\pi(C) = 2.2$$

$$V_\pi(D) = 4.2$$

Bellman Equation

- Bellman equations to compute value functions
- The Bellman equation to directly write down a system of equations for the state values
- More complex problems, this won't always be prac



Summary

- Derive the Bellman equation for state value functions
- Define the Bellman equation for action value functions
- Understand how Bellman equations relate current and future values.

Q & A