Q-LEARNING

1. What is Q-Learning?

Q-Learning is a reinforcement learning algorithm that learns which action is best in a given state by using rewards from experience.

1. **Definition:** Q-Learning is a model-free, off-policy reinforcement learning algorithm that enables an agent to learn the best action in a particular state by interacting with an environment and receiving rewards or penalties. It learns through trial and error without requiring prior knowledge of the environment's transition rules.

2. **Q-Value:** The algorithm maintains a Q-value, $Q(s, a)$, for each state–action pair, which estimates the expected cumulative discounted reward obtained by taking action a in state s and following the best possible actions afterward.

3. **Learning Process:** After performing an action, the agent observes the immediate reward and the next state, then updates the Q-value using the Q-Learning update equation. The learning rate controls the influence of new experience, while the discount factor determines the importance of future rewards.

4. **Objective:** By repeatedly updating Q-values and exploring different actions, the agent learns which actions yield higher long-term rewards. After learning, it can select the action with the highest Q-value in each state to follow an effective policy.

***Q-Learning learns the best policy even when the agent explores using a different policy.*

2. Basic terms

Term	Meaning
State (S)	The current situation.
Action (A)	A choice made in a state.
Reward (R)	Feedback received after taking an action.
Q-value Q(S,A)	The estimated usefulness of taking action A in state S.
Learning rate (α)	Controls how much the new information changes the old Q-value.
Discount factor (γ)	Controls the importance of future rewards.

3. Q-Learning formula

$$\text{New Q} = \text{Old Q} + \alpha \times [R + \gamma \times \max Q(\text{next state}) - \text{Old Q}]$$

Here, R is the immediate reward. max Q(next state) is the highest Q-value available in the next state. If the next state is terminal (the task ends), use 0 for its future Q-value.

4. numerical exercise

Question: An agent has Old Q = 2. It receives reward R = 4. The highest Q-value in the next state is 6. Let α = 0.5 and γ = 0.9. Calculate the New Q-value.

Step 1: Write the formula.

$$\text{New Q} = \text{Old Q} + \alpha \times [R + \gamma \times \max Q(\text{next state}) - \text{Old Q}]$$

Step 2: Substitute the values.

$$\text{New } Q = 2 + 0.5 \times [4 + (0.9 \times 6) - 2]$$

Step 3: Calculate inside the brackets.

$$\text{New } Q = 2 + 0.5 \times [4 + 5.4 - 2] = 2 + 0.5 \times 7.4$$

Step 4: Calculate the final answer.

$$\text{New } Q = 2 + 3.7 = 5.7$$

5. Numerical exercise

An agent is in State A and can choose Action 1 or Action 2. The agent takes Action 1, moves to State B, and receives reward R = 3. The old Q-value for (A, Action 1) is 1. The Q-values for actions available in State B are shown below. Use $\alpha = 0.5$ and $\gamma = 0.8$ to update Q(A, Action 1).

State	Action	Q-value
B	Action 1	4
B	Action 2	2

Solution:

1. 1. Highest Q-value in State B = max(4, 2) = 4.
2. 2. New Q = 1 + 0.5 × [3 + (0.8 × 4) − 1].
3. 3. New Q = 1 + 0.5 × [3 + 3.2 − 1] = 1 + 0.5 × 5.2.
4. 4. New Q = 1 + 2.6 = 3.6.

Final answer: Q(A, Action 1) changes from 1 to 3.6.