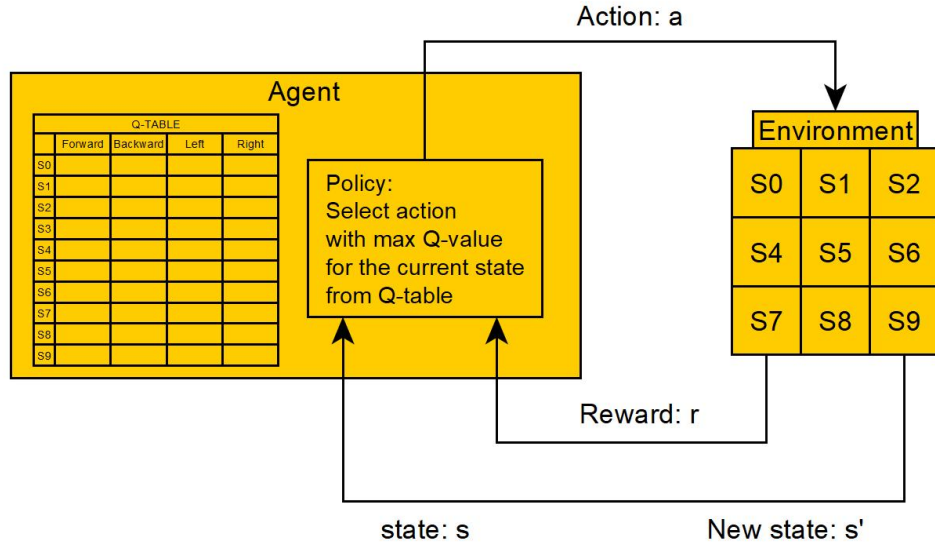


Q Reinforcement Learning with PyTorch Step by Step:



1. Observe the current state **s**.
2. Select an action **a** based on the agent's policy.
3. Perform the action **a** in the environment.
4. The state transitions to a new state **s'**.
5. Receive a reward **r**.
6. Update the Q-value for the previous state-action pair **Q(s, a)** using the Bellman equation:

$$Q(s, a) = Q(s, a) + \alpha * [r + \gamma * \max(Q(s', a)) - Q(s, a)]$$

where:

- α is the learning rate.
 - γ is the discount factor.
 - From the Q-table, take **Q(s, a)** for the previous state-action pair.
 - For the current new state **s'**, look up the Q-table and take the maximum Q-value, **max(Q(s', a))**.
7. Repeat steps 1-6 for every action taken.
 8. Over time, the agent learns an optimal policy by balancing exploration and exploitation.