SARSA

SARSA (State-Action-Reward-State-Action) is a reinforcement learning algorithm that belongs to the family of temporal difference (TD) methods. It's an on-policy algorithm, meaning it learns a policy while following that policy.

How SARSA Works:

1. **Initialize:**
    - Initialize the Q-value function `Q(s, a)` for all state-action pairs to an arbitrary value (often 0).
    - Set the learning rate `α` and discount factor `γ`.
2. **Choose Action:**
    - Given the current state `s`, choose an action `a` using an ε-greedy policy. This means that with probability ε, a random action is chosen, and with probability 1-ε, the action with the highest estimated Q-value is chosen.
3. **Take Action and Observe:**
    - Take action `a` in state `s` and observe the next state `s'` and the reward `r`.
4. **Update Q-Value:**
    - Update the Q-value function using the following equation:

      where
      `a'` is the next action chosen using the ε-greedy policy in state `s'`.

      $$Q(s,a) < -Q(s,a) + \alpha * (r + \gamma * Q(s',a') - Q(s,a))$$
5. **Repeat:**
    - Repeat steps 2-4 until convergence or a desired number of episodes.

Advantages of SARSA:

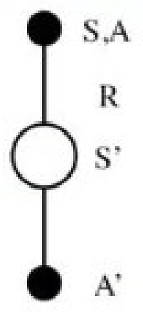
- **Simple to implement:** SARSA is relatively easy to understand and implement.
- **Online learning:** It can learn from experience as it is acquired, making it suitable for tasks with long or infinite episodes.
- **Efficient:** SARSA can be computationally efficient, making it suitable for large-scale problems.

Disadvantages of SARSA:

- **Can be slow to converge:** SARSA can be slow to converge, especially for complex environments.
- **Sensitive to hyperparameters:** The learning rate and discount factor can significantly affect the performance of SARSA.

Applications of SARSA:

- **Game playing:** SARSA has been successfully applied to various games, including tic-tac-toe and backgammon.
- **Robotics:** SARSA can be used to learn control policies for robots.
- **Natural language processing:** SARSA can be used for tasks such as machine translation and dialogue systems.



$$Q(S,A) \leftarrow Q(S,A) + \alpha\left(R + \gamma Q(S',A') - Q(S,A)\right)$$