Criterion of Optimality

In reinforcement learning, an agent interacts with an environment to learn how to make decisions that maximize a notion of cumulative reward over time. The criterion of optimality is a mathematical formulation of the goal of the agent, which is to maximize the expected cumulative reward.

The criterion of optimality is usually expressed in terms of the value function, which is a function that assigns a value to each state or state-action pair. The value function represents the expected cumulative reward starting from that state or state-action pair. The criterion of optimality states that the optimal policy is the one that maximizes the value function.

There are different algorithms for computing the value function and the optimal policy, such as dynamic programming, Monte Carlo methods, and temporal difference learning. These algorithms use different techniques to estimate the value function and update it based on the rewards and the transitions observed in the environment.

In practice, the criterion of optimality is useful in many real-life applications of reinforcement learning, such as robotics, game playing, and control systems. For example, in robotics, an agent may learn how to navigate a complex environment by maximizing the expected cumulative reward, which could be defined as the distance traveled or the number of objects collected.