Markov Decision Processes [MDP's]

Markov decision processes [MDP's] is a theoratical formalization of sequential decision making problems. It provides a framework by which we can theoratically describe sequential decision making problems.

The need of this framework arises from the limitations of the one used to describe bandit problems. The bandit problem idea cant be used if :-

① The action to be taken varies across time steps.

② The agent needs to adjust it's behavior as the state changes.

The above characteristics are very often present in real world problems and so the need of a more robust and flexible framework arises.

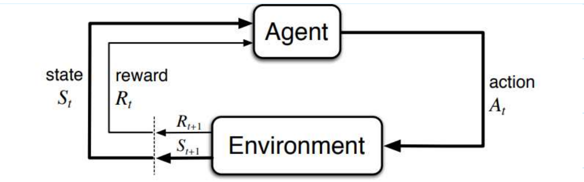
Mathmatically an MDP problem can be described using a function of the form $p: S \times R \times S \times R \rightarrow [0,1]$

or $p(s', r \mid s, a) = Pr\{S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a\}$

This can be described as follows :-

time t-1

An agent initially in state s takes an action a. It then enters the state s' and receives a reward r.



The =n above describes the probability that agent will enter a new state s' and receive a reward r given initially it's in state s and takes an action a. Since it's a probability

$$\sum_{s'\in\mathcal{S}}\sum_{r\in\mathcal{R}} p(s', r \mid s, a) = 1 \quad \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$$

In a finite MDP process, there are finite states, rewards and actions.

In a Markov process, the probabilities of each possible value of S_t and R_t depend only on the values of the immediately preceding values of states and actions. If a state or action prior to the immediately preceding state or action affects future values of states or actions, then this information must be captured in the immediately preceding state and action values.

The MDP framework is abstract and flexible and can be applied to a wide array of problems.