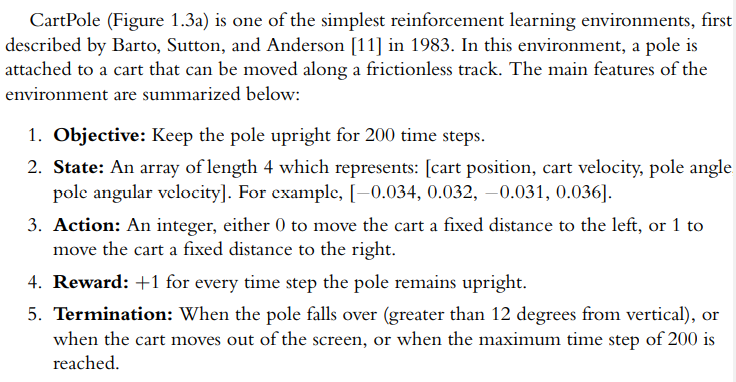
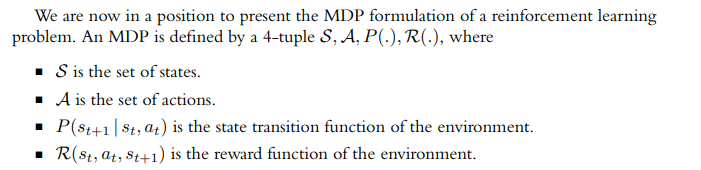
**Intro to RL**

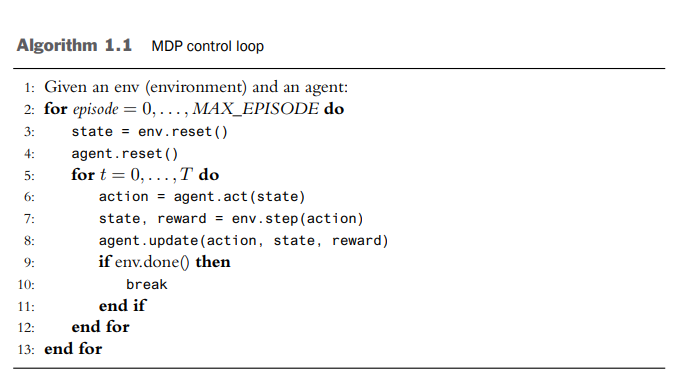
**1/ RL**

* RL is concerned with solving sequential decision-making problems
* The (st, at, rt) tuple is called an experience.
* The time horizon from t = 0 to when the environment terminates is called an episode.
* An agent needs many episolves to learn a good policy ranging from hundred to million



**2/ RL as MDP = Markov Decision Process**

* The consideration of how the environment transition from 1 state to the next using what is known as the transition function.
* In RL, transition function is formulated as MDP
* 
* Markov property implies that the current state and action at time step t contain sufficient info to fully determine the transition probability fr the next state at t+1
* 
* The agents do not have access to the transition function or reward function. The only way an agent can get info about these function is thru the state, actions, and reward is thru experiences in the environment



* This algorithm expresses the interactions between an agent and env over many episodes and time steps.
* At the beginning of each episode, the env and agent are reset.
* On reset, the environment produces an initial state. Then the agent and environment begin to interact meaning that the agent produce action (line 6) and then env produce next state and reward (line 7)

**3/ Learnable Functions In RL**

* With RL formulate as MDP, what should the agent learn?