Markov Chain

A Markov chain is a mathematical system that experiences transitions from one state (a situation or set of values) to another according to certain probabilistic rules. The basic property of a Markov chain is that the probability of the new state x_{n+1} depends only upon the current state x_n, not upon the previous states. If the chain is currently in a state $x_n = j$, then it moves to state $x_{n+1} = i$ at the next step with a probability denoted by p_{ij}, which can form transition matrix *P*. (Note: all p_{ij}>0 and each column of *P* adds to 1)

$$y_{n+1} = \begin{bmatrix} Prob\{x_{n+1} = 1\} \\ \vdots \\ Prob\{x_{n+1} = N\} \end{bmatrix} = \begin{bmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{bmatrix} \begin{bmatrix} Prob\{x_n = 1\} \\ \vdots \\ Prob\{x_n = N\} \end{bmatrix}$$

$$y_{n+1} = Py_n$$

Linear Dynamical System

The linear dynamical system is sometimes called a Markov model (after the mathematician Andrey Markov).
As we mentioned before, Markov studied systems in which the next state value depends on the current one, and not on the previous state values. The linear dynamical system is the special case of a Markov system where the next state is a linear function of the current state.
A linear dynamical system is a simple model for the sequence, in which each x_{t+1} is a linear function of x_t:

$$x_{t+1} = A_t x_{t.} \qquad \text{t} = 1, 2, \ldots \tag{1.1}$$

If we know x_t (and $A_t, A_{t+1}, \ldots$) we can determine $x_{t+1}, x_{t+2}, \ldots$ simply by iterating the dynamics equation (1.1). In other words: If we know the current value of x, we can find all future values.
Linear dynamical systems have many applications as statistical models of real-world processes such as studying the dynamics of infection and the spread of an epidemic, describing the evolution of the age distribution in some population over time, studying the dynamics of a supply chain the dynamics of a supply chain, (approximately) describing the motion of many mechanical systems and so on.

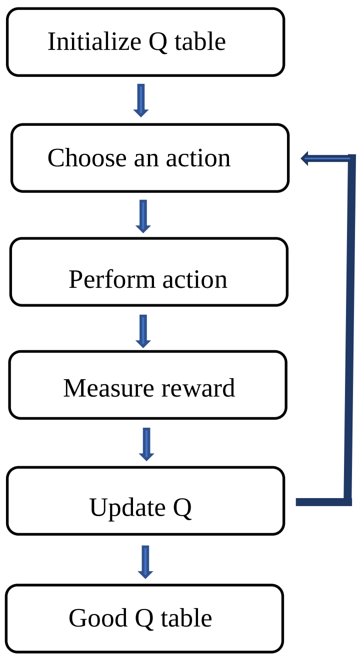
Q-Learning

Q-learning is a model-free reinforcement learning algorithm. In Q-learning, we have agents, a set of actions, a set of states, and rewards. Q-learning algorithm allows agents to learn the best action in a given state by trying every action in a state and updating the expected reward with the actual reward for that action.

$$\text{new}Q_{S,A} = (1 - \alpha)Q_{S,A} + \alpha(R_{S,A} + \gamma \times maxQ'_{S',A'})$$

Learning Rate (α): the rate of how much old Q learns from new Q. $\alpha = 0$ means that only the old knowledge is considered, while $\alpha = 1$ means new information is the only thing that is important.

Discount Factor (γ): the importance of future rewards. When $\gamma = 0$, only short-term rewards are considered. When $\gamma = 1$, long-term rewards are important.



Q-Learning Steps