Tips: From Text Book, Slides and Website.

(一) Markov Chain (马尔可夫链)

(I) Introduction:

A Markov Chain is a stochastic (随机的) model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

A countably infinite sequence, in which the chain moves state at discrete time steps, gives a discrete-time Markov Chain (DTMC).

(II) Definition:

A Markov process is a stochastic process that satisfies the Markov property[1] (sometimes characterized as "memorylessness"). In simpler terms, it is a process for which predictions can be made regarding future outcomes based solely on its present state and—most importantly—such predictions are just as good as the ones that could be made knowing the process's full history.[11] In other words, conditional on the present state of the system, its future and past states are independent.

来自 <https://en.wikipedia.org/wiki/Markov_chain#Definition>

定义: 考虑一个随机变量的序列 $X=\{X_1, X_2, \cdots, X_t, \cdots\}$, 这里 X_t 表示时刻 t 的随机变量。每个 X_t 的取值集合维度相同. 称为状态空间, 表示为 S。随机变量可以是随机的, 也可以是连续的。以上随机变量的序列构成随机过程 (stochastic process)。

假设在初始状态中, 随机变量 X_0 遵循概率分布 $P(X_0)=\pi_0$, 称为初始状态分布。

- 马尔可夫性:

如果 X_t 只依赖于 X_{t-1}, 而不依赖于过去的随机变量 $\{X_0, X_1, \cdots X_{t-2}\}$

$$P(X_t \mid X_0, X_1, X_2, \cdots, X_{t-1}) = P(X_t \mid X_{t-1}), \quad t=1,2,\cdots$$

具有马尔可夫性的随机变量序列 $\{X_0, X_1, \cdots, X_t\}$ 称为马尔可夫链或马尔可夫过程。条件概率分布 $P(X_t \mid X_{t-1})$ 称为马尔可夫链的转移概率分布。

- 时间齐次的马尔可夫链:

若转移概率分布 $P(X_t \mid X_{t-1})$ 与 t 无关, 即

$$P(X_{t+s} \mid X_{t-1+s}) = P(X_t \mid X_{t-1}), \quad t=1,2\cdots; \ s=1,2\cdots$$

- n 阶马尔可夫链:

$$P(X_t \mid X_0 X_1 \cdots X_{t-1}) = P(X_t \mid X_{t-n} \cdots X_{t-2} X_{t-1}).$$

- Type of Markov Chains:

	Countable state space	Continuous or general state space
Discrete-time	(discrete-time) Markov chain on a countable or finite state space	Markov chain on a measurable state space (for example, Harris chain)
Continuous-time	Continuous-time Markov process or Markov jump process	Any continuous stochastic process with the Markov property (for example, the Wiener process)

(III) Discrete-Time Markov Chain (离散状态马尔可夫链)

process)

(Ⅲ) Discrete-Time Markov Chain（离散状态马尔可夫链）

◦ 转移概率矩阵.

若马尔可夫链在时刻(t-1)处于状态j，在时刻t移动到i，将转移概率记作：

$$P_{ij} = (x_t = i \mid x_{t-1} = j),\ i=1,2,\cdots;\ j=1,2,\cdots$$

满足：$P_{ij} \ge 0$；$\sum_i P_{ij} = 1$

用矩阵可表示为：　　　　　　　　在t时刻的概率分布：

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{ij} \\ P_{21} & P_{22} & \cdots & P_{2j} \\ \vdots & \vdots & & \vdots \\ P_{ji} & P_{j2} & \cdots & \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \qquad \pi(t) = \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \vdots \end{bmatrix}$$

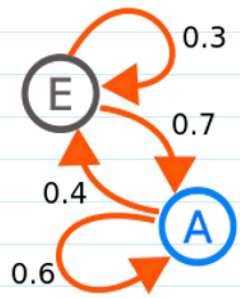
有限离散状态的马尔可夫链，可以由有向图来表示。结点表示状态；边表示状态之间的转移，边上的数值表示转移概率。

马尔可夫链X在时刻t的状态分布，可以由时刻$(t-1)$的状态分布以及转移概率分布决定，即

$$\pi(t) = P \cdot \pi(t-1)$$

证明：

$$\pi_i(t) = P(x_i = i)$$
$$= \sum_m P(x_i = i \mid x_{t-1} = m) \cdot P(x_{i-1} = m)$$
$$= \sum_m P_{im} \cdot \pi_m(t-1)$$



其中i便是状态分布中的维度。

故，从上面的式子可以推理得到：

$$\pi(t) = P\pi(t-1) = P^2\pi(t-2) = \cdots = P^t\pi(0)$$

上面的P^t表示t步转移概率：

$$P^t_{ij} = P(x_i = i \mid x_0 = j)$$

(Ⅳ) Continuous-Time Markov Chains（连续状态马尔可夫链）

连续状态马尔可夫链 $X = \{x_0, x_1, \cdots, x_t, \cdots\}$ 定义在连续空间S，转移概率分布由概率转移核或转移核来表示：

设S是连续空间，对任意的$s \in S$, $A \subset S$，转移核$P(x,A)$定义为：

$$P(x,A) = \int_A p(x,y)\,dy.$$

其中$p(x,\cdot)$是概率密度函数。满足$p(x,\cdot) \ge 0$. $p(x,S) = \int_S p(x,y)\,dy = 1$

转移核$p(x,A)$表示从$x \sim A$的转移概率.

$$P(x,A) = P(X_t = A \mid X_{t-1} = x)$$

(二) Hidden Markov Mode（隐马尔可夫模型）

Hidden Markov Model (**HMM**) is a statistical Markov model in which the system being modeled is assumed to be a Markov process – call it X – with unobservable ("*hidden*") states. HMM assumes that there is another process Y whose behavior "depends" on X. The goal is to learn about X by observing Y. HMM stipulates that, for each time instance n_{0}, the conditional probability distribution of Y_{n_0} given the history $\{X_n = x_n\}_{n \le n_0}$ must **not** depend on $\{x_n\}_{n<n_0}$.

Hidden Markov models are known for their applications to thermodynamics, statistical mechanics, physics, chemistry, economics, finance, signal processing, information theory, pattern recognition - such as speech, handwriting, gesture recognition,[1] part-of-speech tagging, musical score following,[2] partial discharges[3] and bioinformatics.[4]

隐马尔可夫模型是可以用于标注问题的统计学习模型，描述由隐藏的马尔可夫链随机生成观测序列的过程，属于生成模型。

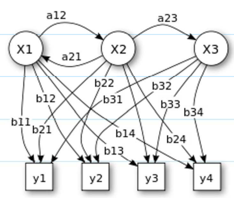
(I) Definition:(隐马尔可夫模型)

隐马尔可夫模型是关于时序的概率模型，描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列，再由各个状态生成一个观测从而产生观测随机序列的过程。隐藏的马尔可夫链随机生成的状态序列，称为状态序列(state sequence)；每个状态生成一个观测，而由此产生的观测的随机序列，称为观测序列(observation sequence)。序列的每一个位置都可以看成是一个时刻。

该模型主要由初始概率分布、状态转移概率分布以及观测概率分布确定。形式如下：

设 Q 是所有可能的状态的集合，V 是所有可能的观测的集合。

$$Q=\{q_1,q_2,\cdots,q_N\};\quad V=\{v_1,v_2,\cdots,v_M\}$$



其中，N 是可能的状态数；M 是可能的观测数。

I 是长度为 T 的状态序列；O 是对应的观测序列。

$$I=\{i_1,i_2,\cdots,i_T\};\quad O=\{o_1,o_2,\cdots,o_T\}.$$

A 是状态转移概率矩阵。

$$A=[a_{ij}]_{N\times N}$$

其中，

$$a_{ij}=P(i_{t+1}=q_j\mid i_t=q_i),\quad i=1,2,\cdots,N;\ j=1,2,\cdots,N$$

是在时刻 t 处于状态 q_i 的条件下，在时刻 $t+1$ 状态转移到 q_j 的概率。

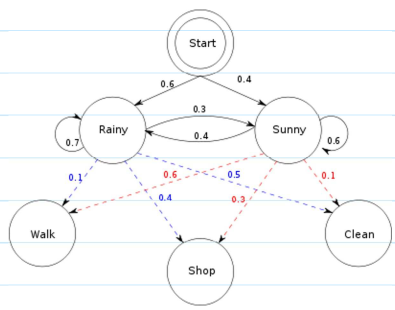
B 是观测概率矩阵：

$$B=[b_j(k)]_{N\times N}$$

其中，

$$b_j(k)=P(o_t=v_k\mid i_t=q_j),\quad k=1,2,\cdots,M,\ j=1,2,\cdots,N$$

是在时刻 t 处于状态 q_j 的条件下生成观测 v_k 的概率。



π 是初始状态概率向量：

$$\pi=(\pi_i)$$

其中，

$$\pi_i=P(i_1=q_i),\quad i=1,2,\cdots,N$$

是时刻 $t=1$ 时，处于 q_i 状态的概率。

隐马尔可夫模型由初始状态概率向量 π，状态转移概率矩阵 A 和观测概率矩阵 B 三者决定。π 和 A 决定状态序列，B 决定观测序列。因此，隐马尔可夫模型可以用三元符号表示，即：

$$\lambda=(A,B,\pi)$$，其中 A、B、π 是三要素。

A 与 π 确定了隐藏的马尔可夫链，生成不可观测的状态序列。

B 确定了从状态生成观测。