- 学习过程
  - 宏观掌握思路，清楚问题是什么，为什么要这么做。
  - 掌握技术性细节。
  - 通过编程实践，测试自己对算法的理解。
- 概率复习
- Bayesian Nets　　　　会出现动态规划
- Applications:
  - Naïve Bays
  - Markov Chain　　　　会出现矩阵的特征值
  - Hidden Markov Model
  - Restricted Boltzmann machine

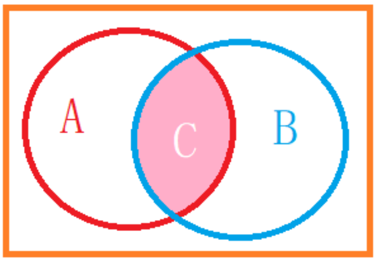
隐马尔可夫模型就是对时序数据进行建模，但时间复杂度比较高。

我们自己来发明一个隐马尔可夫模型

概率

- 3 Axioms
  - $0 \leq p(A) \leq 1$
  - $P(True) = 1$; $P(False) = 0$
  - $P(A) + P(B) - P(A \; and \; B) = P(A \; or \; B)$

三个公理，假设它是正确的　　**Sa+Sb –Sc = Sab**



P(A|B)=Sc/Sb

- $P(A|B) \stackrel{\text{def}}{=} \dfrac{P(A \; and \; B)}{P(B)}$

- $P(a,b) = p(a|b)P(b)$
- $P(a,b,c) =$

$P(a,Y) = P(a|Y)P(Y) = P(a|b,c)P(b|c)P(c)$

$= P(c)P(b|c)P(a|b,c)$

$P(X,c) = P(c|X)P(X) = P(c|a,b)P(a|b)P(b)$

$= P(b)P(a|b)P(c|a,b)$

- 独立
  - $P(A|B) = P(A)\ if\ A \perp B$
- Bayesian Nets

出现动态规划

当我们知道了一组联合分布 x1,x2,x3,x4,..xi 即 P(x1,x2,x3,..xi)

我们就知道了 P(xi), P(xi|xj) 等。所以我们要求 P(x1,x2,x3,..xi)

如果直接求/直接建模参数会是指数级的增长。我们没有办法存下来这麽多的信息。

概率图模型的动机就是通过一种有效的方式来求 P(x1,x2,x3,..xi)

也就是用一种有效的方式对联合分布进行建模。

概率图模型：动机

- 维度灾难问题
- d binary variables, 2^d-1 terms!
- 终极目标：如何对一个联合概率分布有效表示。
- 因为知道了联合分布，就知道了所有信息。

维度灾难举例：3个变量的联合分布

同时选择X1,X2,X3的联合概率是多少

X1	X2	X3	P(X1,X2,X3)
T	T	T	0.2
T	F	T	
.			
.			
.			
F	F	F	0.41

2**3

2**n

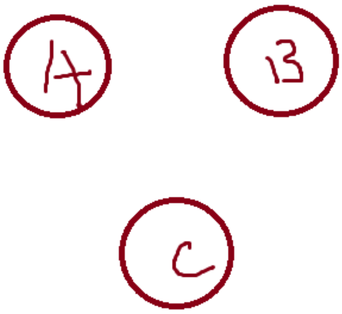
1

当变量个数==n，都进行这样类似独立的运算即求乘积，变量的运算量和储存量是指数级的增长。而概率图模型就是为了解决这个问题。概率图模型就是先假设 x1,x2,...变量之间不是独立的而是存在某种联系关系，就可以画出有向关系图进行建模。

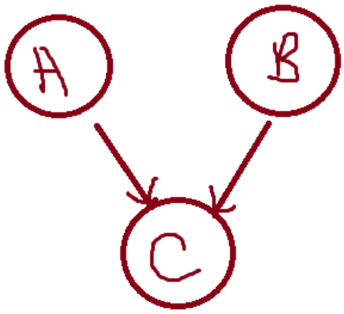
Directed Graphical Models

- 有向无环图
- 关键结论：
  - $p(x_1, x_2, ..x_n) = \prod p(x_i|parents(x_i))$
  - x is independent with its non-decendents given its parents. 给定某变量的父节点，该变量和非后代变量条件独立。

如果它们相互独立，要求他们的关系（联合概率）就要就乘积

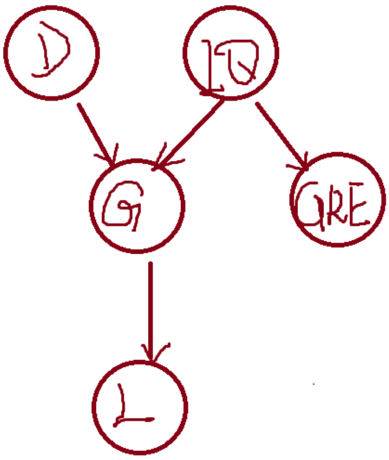


如果存在A,B决定了C，这样会不会减少储存的空间？



□ $$\boxed{\mathrm{p}(x_1,x_2,..x_n)=\prod p(x_i|parents(x_i))}$$

通过概率图进行计算减少了储存量和计算量



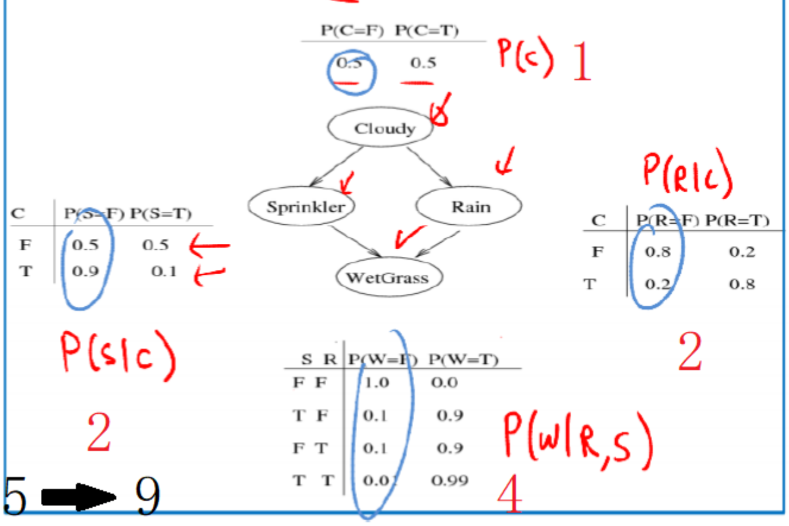
P(D,IQ,G,GRE,L)=
P(L|G)P(G|D,IQ)
P(GRE|IQ)P(D)
P(IQ)

原始联合分布 vs 分解(factorized)分布

c	s	r	w	prob
0	0	0	0	0.200
0	0	0	1	0.000
0	0	1	0	0.005
0	0	1	1	0.045
0	1	0	0	0.020
0	1	0	1	0.180
0	1	1	0	0.001
0	1	1	1	0.050
1	0	0	0	0.090
1	0	0	1	0.000
1	0	1	0	0.036
1	0	1	1	0.324
1	1	0	0	0.001
1	1	0	1	0.009
1	1	1	0	0.000
1	1	1	1	0.040

15

$$p(C,S,R,W) = p(C)p(S|C)p(R|C)p(W|S,R)$$



15 ➡ 9

- Given the graph, what is p(s=1) ?

对联合概率求积分，求S=1就求C,R,W的积分将C,R,W约掉

$$\boxed{P(s=1)=\sum_{C=0}^{1}\sum_{R=0}^{1}\sum_{W=0}^{1}P(C,R,W,s=1)}$$

$$=\sum_{C}\sum_{R}\sum_{W}P(C)P(s=1|C)P(R|C)P(W|s=1,R)$$

$$\sum_{C}\sum_{R}\sum_{W=0}^{1}P(W|s=1,R)P(s=1|C)P(R|C)P(C)$$

ab+ac

a(b+c)

$$=^{000}\ P(W=0|S=1,R=0)\times P(S=1|C=0)\times P(R=0|C=0)\times P(C=0)$$

0.1 0.5 0.8 0.5

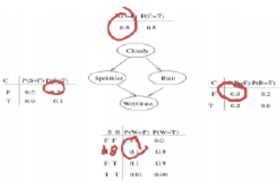
$$^{001}+P(W=1|S=1,R=0)P(S=1|C=0)P(R=0|C=0)P(C=0)$$

$$^{010}+P(W=0|S=1,R=1)P(S=1|C=0)P(R=1|C=0)P(C=0)$$

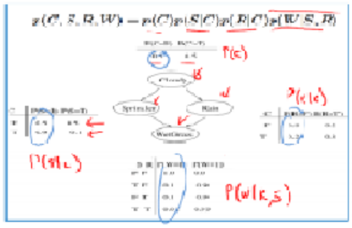
$$^{011}+\cdots$$

$\vdots$

$^{111}+$



```
For R = 0:1
    for C = 0:1
        for w = 0:1
            prob = prob +
```



Smart Approach: 变量消除 / 动态规划

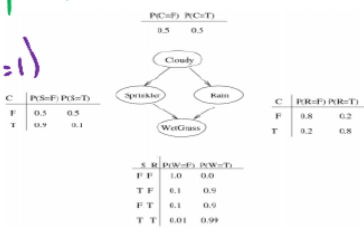
$$P(s=1) = \sum_C \sum_R \sum_W P(w|s=1,R)\, P(R|c)\, P(c)\, P(s=1|c)$$

$$= \sum_C \sum_R P(R|c)P(c)P(s=1|c) \boxed{\sum_W P(w|s=1,R)} = 1 \quad \sim \phi = 1$$

$$= \sum_C P(c)P(s=1|c) \boxed{\sum_R P(R|c)} = 1 \quad \sim \psi = 1$$

$$= P(s=1|c=0)P(c=0) + P(s=1|c=1)P(c=1)$$

$$= 0.3$$



$\psi = 0$
$\phi = 0$
$\Theta = 0$
FOR $W = 0:1:1$
  $\phi_R = \phi_R + P(W|s=1,R)$
END
FOR $R = 0:1:1$
  $\psi_c = \psi + P(R|c)\phi_R$
END
FOR $C = 0:1:1$
  $\Theta = \Theta + P(s=1|c)P(c)\psi_c$
END

$\rightarrow \Theta = 0.3$

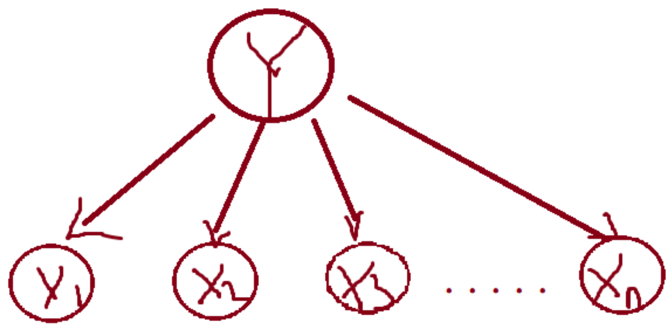
概率图应用一：朴素贝叶斯

- $P(y = 1 \mid x1, x2, x3, ..xn)$

$$= \frac{p(x_1, x_2, x_3, ..xn \mid y=1)\ p(y=1)}{p(x_1, x_2, ..xn)} \quad \text{(always true)}$$

（Naïve bayes）

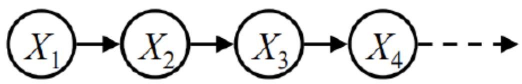
$$= \frac{\boldsymbol{P(X1|Y)P(X2|Y)...P(Xn|Y)P(Y=1)}}{\boldsymbol{P(X1,X2....Xn)}},$$



$P(Y, X1, X2, X3...XN) = P(X1|Y)P(X2|Y)....P(XN|Y)P(Y)$

这就是朴素贝叶斯，当确定了父节点，父节点与不同子节点关系相互独立，它是一种简单的概率图模型。

概率图应用二：Markov model



$P(X_1) \qquad P(X_t|X_{t-1})$

当我们变量的概率图为链式的，就是马尔可夫模型

$P(x1,x2,x3,..) = P(x1)P(x2|x1)P(x3|x2)....P(Xt|Xt-1)$

$P(x_1, x_2, x_3, X_4)$

$$X_1 \to X_2 \to X_3 \to X_4$$

Query: $P(X_4)$

- Slow answer: inference by enumeration
  - Enumerate all sequences of length t which end in s
  - Add up their probabilities

$$P(X_4) = \sum_{x_1, x_2, x_3} P(x_1, x_2, x_3, X_4)$$

$$\begin{aligned}
&= P(X_1 = +x_1)P(X_2 = +x_2|X_1 = +x_1)P(X_3 = +x_3|X_2 = +x_2)P(X_4|X_3 = +x_3) \\
&+P(X_1 = +x_1)P(X_2 = +x_2|X_1 = +x_1)P(X_3 = -x_3|X_2 = +x_2)P(X_4|X_3 = -x_3) \\
&+P(X_1 = +x_1)P(X_2 = -x_2|X_1 = +x_1)P(X_3 = +x_3|X_2 = -x_2)P(X_4|X_3 = +x_3) \\
&+P(X_1 = +x_1)P(X_2 = -x_2|X_1 = +x_1)P(X_3 = -x_3|X_2 = -x_2)P(X_4|X_3 = -x_3) \\
&+P(X_1 = -x_1)P(X_2 = +x_2|X_1 = -x_1)P(X_3 = +x_3|X_2 = +x_2)P(X_4|X_3 = +x_3) \\
&+P(X_1 = -x_1)P(X_2 = +x_2|X_1 = -x_1)P(X_3 = -x_3|X_2 = +x_2)P(X_4|X_3 = -x_3) \\
&+P(X_1 = -x_1)P(X_2 = -x_2|X_1 = -x_1)P(X_3 = -x_3|X_2 = -x_2)P(X_4|X_3 = -x_3) \\
&+P(X_1 = -x_1)P(X_2 = -x_2|X_1 = -x_1)P(X_3 = -x_3|X_2 = -x_2)P(X_4|X_3 = -x_3)
\end{aligned}$$

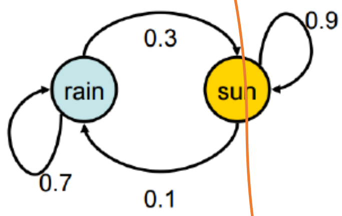
- Fast answer: variable elimination
  - Order: X_1, X_2, X_3

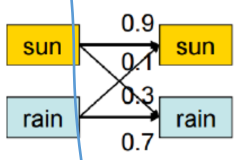
$$\begin{aligned}
P(X_4) &= \sum_{x_1, x_2, x_3} P(x_1, x_2, x_3, X_4) \\
&= \sum_{x_3} \sum_{x_2} \sum_{x_1} P(X_4|x_3)P(x_3|x_2)P(x_2|x_1)P(x_1) \\
&= \sum_{x_3} \sum_{x_2} P(X_4|x_3)P(x_3|x_2) \sum_{x_1} P(x_2|x_1)P(x_1) \\
&= \sum_{x_3} \sum_{x_2} P(X_4|x_3)P(x_3|x_2)P(x_2) \\
&= \sum_{x_3} P(X_4|x_3) \sum_{x_2} P(x_3|x_2)P(x_2) \\
&= \sum_{x_3} P(X_4|x_3)P(x_3)
\end{aligned}$$

- States: X = {rain, sun}
- CPT $P(X_t \mid X_{t-1})$:

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Two new ways of representing the same CPT, that are often used for Markov models (These are not BNs!)





问题：X1 是下雨，X2 是天晴的概率是多少

P(X1,X2) = P(X2|X1)P(X1)

P(x2) $=\sum_{x1} \boldsymbol{p(x1,x2) = p(x2|x1)p(x1=1) + p(x2|x1)p(x1=0)}$

$$\begin{matrix} \mathbf{0.9} & \mathbf{0.3} \\ \mathbf{0.1} & \mathbf{0.7} \end{matrix} \Rightarrow$$ 状态转移矩阵/马尔可夫矩阵

$$\begin{matrix} \mathbf{0.9} & \mathbf{0.3} \\ \mathbf{0.1} & \mathbf{0.7} \end{matrix} * \begin{matrix} \boldsymbol{P(x1=1)} \\ \boldsymbol{P(x1=0)} \end{matrix} = \begin{matrix} \boldsymbol{P(x2=1)} \\ \boldsymbol{P(x2=0)} \end{matrix} = \begin{matrix} \mathbf{0.9} \\ \mathbf{0.3} \end{matrix}$$

当初始值不一样的时候，P(Xn)不变，如下

- From initial observation of sun

$$\underset{P(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{P(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{P(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{P(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \Rightarrow \underset{P(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- From initial observation of rain

$$\underset{P(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{P(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle} \quad \underset{P(X_3)}{\left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle} \quad \underset{P(X_4)}{\left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle} \Rightarrow \underset{P(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

steady state

- From yet another initial distribution $P(X_1)$:

$$\underset{P(X_1)}{\left\langle \begin{matrix} p \\ 1-p \end{matrix} \right\rangle} \quad \cdots \Rightarrow \underset{P(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

说明马尔可夫链无记忆性，有它独特的特点

我们来证明一下马尔可夫链/模型 有稳定态。

设 $\boldsymbol{A}$ 是 n 阶方阵，若存在 n 维非零向量 $\boldsymbol{x}$，使

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$$

则称数 λ 为 $\boldsymbol{A}$ 的特征值，$\boldsymbol{x}$ 为 $\boldsymbol{A}$ 属于 λ 的特征向量

矩阵的平方乘以向量 等于 特征值的平方乘以向量

$$A\vec{x}=\lambda\vec{x}\ ,\ A^2\vec{x}=\lambda A\vec{x}=\lambda\lambda\vec{x}=\lambda^2\vec{x}$$

$x_0 \longrightarrow x$　　$\begin{pmatrix}1\\0\end{pmatrix}$　$\begin{pmatrix}p\\1-p\end{pmatrix}$

矩阵的幂

$$\vec{x}_0=\vec{x}_0$$

$$\vec{x}_1=A\cdot\vec{x}_0$$

$$\vec{x}_2=A\cdot\vec{x}_1=A\cdot A\cdot\vec{x}_0=A^2\vec{x}_0$$

$$\vec{x}_n=A^n\cdot\vec{x}_0$$

$$\Rightarrow\boxed{A^k\cdot\vec{x}=\lambda^k\cdot\vec{x}}$$

C 是相互独立的特征向量

A: n 个 eigen vector $\vec{c}_1,\cdots\vec{c}_n$

$$\boxed{\vec{x}_0=a_1\vec{c}_1+a_2\vec{c}_2+\cdots a_n\vec{c}_n}$$

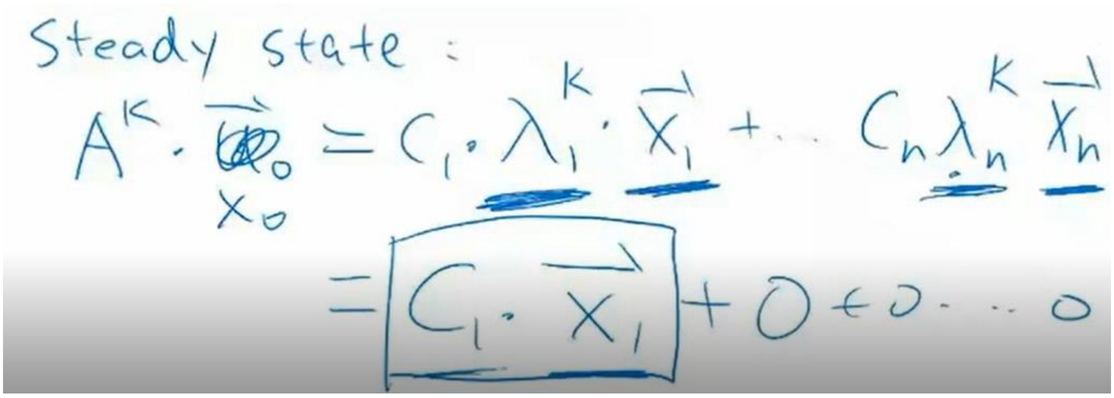
$$A^k\cdot\vec{x}_0=a_1A^k\cdot\vec{c}_1+a_2\cdot A^k\vec{c}_2\cdots a_nA^k\vec{c}_n$$

$$=a_1\cdot\lambda_1\cdot\vec{c}_1+a_2\cdot\lambda_2\vec{c}_1$$

$$=\cdots a_n\lambda_n\vec{c}_n$$

Claim：1是马科夫矩阵的一个特征值

马 尔 可 夫 矩 阵 必 然 有 一 个 特 征 值 λ =1 ， 其 他 特 征 值 小 于 1



我们最后的 steady state 就是马尔科夫矩阵的特征值为 1 所对应的特征向量的一个倍数

$$\underset{P(X_1)}{\begin{pmatrix} p \\ 1-p \end{pmatrix}} \quad \cdots \quad \longrightarrow \underset{P(X_\infty)}{\begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}}$$

0.75,0.25 是马尔科夫矩阵所对应的特征值为 1 的所对应的特征向量的一个倍数

因为他的稳定态是马尔可夫矩阵的 K 次方乘以初始值，而初始值可以分解为他的特征向量的一个线性表达，而马尔可夫的矩阵的特征值一个为 1 其他小于 0，为 1 的就是 C1X1

为什么马尔可夫矩阵的特征值一个为 1，其他小于 1?

$$A = \begin{matrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{matrix} \qquad A - 1*I = \begin{matrix} -0.1 & 0.2 \\ 0.1 & -0.2 \end{matrix}$$

$$\det(A - 1*I) = 0$$

特征值 = 1

应用3：HMM

*Assume you have a little robot that is trying to **estimate** the **posterior** probability that you are **happy** or **sad**, given that the robot has observed whether you are **watching Game of Thrones (w)**, **sleeping (s)**, **crying (c)** or **face booking (f)**.*

*Let the **unknown state** be **X=h** if you're happy and **X=s** if you're sad.*

*Let **Y** denote the **observation**, which can be **w**, **s**, **c** or **f**.*

We want to answer queries, such as:
$P(X=h|Y=f)$?
$P(X=s|Y=c)$?

*Assume that an expert has compiled the following **prior** and **likelihood** models:*

$P(x|y)$?

$P(x) =$

S	H
0.2	0.8

$P(X=s) = 0.2$

$P(y|x) =$

X \ Y	w	s	c	f
s	0.1	0.3	0.5	0.1
H	0.4	0.4	0.2	0

$\frac{.32}{.32+.02} = 0.94$

$$P(x=H|y=w) = \frac{\overset{0.4}{P(y=w|x=H)}\,\overset{0.8}{P(x=H)}}{\overset{0.4}{P(y=w|x=H)}\,\overset{0.8}{P(x=H)} + \overset{0.1}{P(y=w|x=s)}\,\overset{0.2}{P(x=s)}}$$

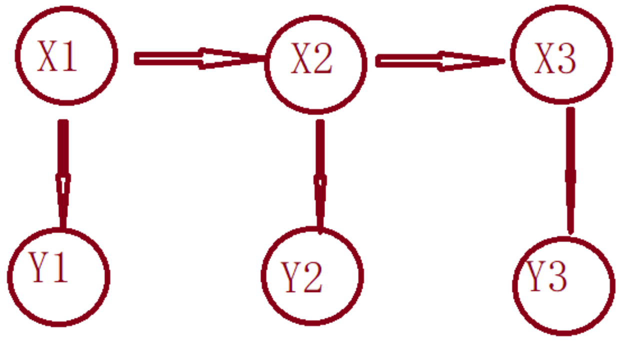
Given a history of observations, say $Y_1=w$, $Y_2=f$, $Y_3=c$, we want to compute the posterior distribution that you are happy at step 3. That is, we want to estimate:

$$\boxed{P(X_3=h|\ Y_1=w, Y_2=f, Y_3=c)}$$

*Clearly, to know if you're happy when crying it helps to know if the sequence of observations is **wcw** or **ccc**.*

*In general, we assume we have an **initial** distribution $P(X_0)$, a **transition** model $P(X_t|\ X_{t-1})$, and an **observation** model $P(Y_t|\ X_t)$.*

P(X1,X2,X3,Y1,Y2,Y3) = P(X1)P(X2|X1)P(X3|X2)P(Y1|X1)P(Y2|X2)P(Y3|X3)



HMM 当我知道 y1,x1,y2,x2 的时候求 x3 出现的概率即 p(x3|y3)

利用联合概率进行分解，利用递归的方式从 0 到 n-1 进行求解

---

*Our goal is to compute, **for all t**, the posterior (aka **filtering**) distribution:*

$$P(X_t | Y_{1:t}) = P(X_t | Y_1, Y_2, \ldots, Y_t)$$

*We derive a **recursion** to compute $P(X_t| Y_{1:t})$ assuming that we have as input $P(X_{t-1}| Y_{1:t-1})$. The recursion has two steps: **prediction** and **Bayesian update**.*

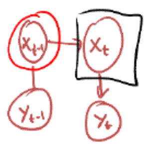
$$P(A|C) = \sum_{B} P(AB|C)$$

$$f(x) = \int_y f(y,x) \Rightarrow \text{trick}$$

$$P(AB|C) = P(A|BC)P(B|C)$$

First, we compute the state prediction: $P(X_t| Y_{1:t-1})$

$$P(x_t|Y_{1:t-1}) = \sum_{x_{t-1}\in\{H,S\}} P(x_t, x_{t-1}|Y_{1:t-1})$$



P(Xt|Y1:t)

$$= \sum_{x_{t-1}} P(x_t|x_{t-1}, \cancel{Y_{1:t-1}})\, P(x_{t-1}|Y_{1:t-1})$$

$$= \sum_{x_{t-1}} P(x_t|x_{t-1})\, P(x_{t-1}|Y_{1:t-1})$$

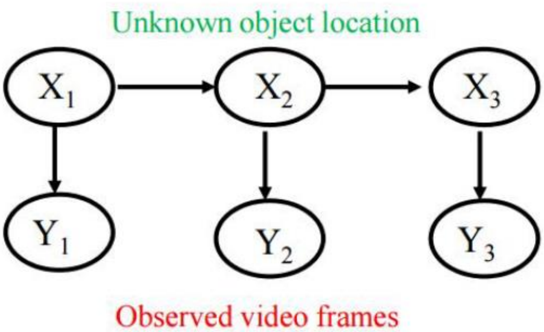
知道Xt-1,Yt-1就是相互独立的。HMM假设

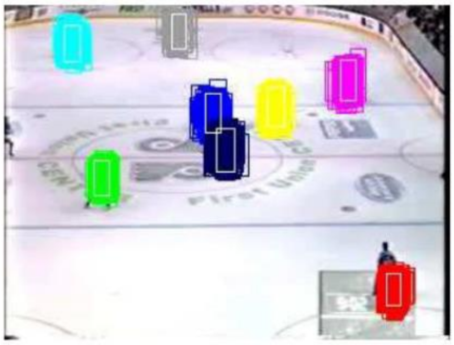
Second, we use Bayes rule to obtain ***$P(X_t|\ Y_{1:t})$***

$$P(x_t | y_{1:t}) = P(\overset{A}{x_t} | \overset{B}{y_t}, \overset{C}{y_{1:t-1}})$$

$$= \frac{P(y_t | x_t, y_{1:t-1})\, P(x_t | y_{1:t-1})}{\sum_{x_t} P(y_t | x_t, y_{1:t-1})\, P(x_t | y_{1:t-1})}$$

$$= \frac{P(y_t | x_t)\, P(x_t | y_{1:t-1})}{\sum_{x_t} P(y_t | x_t)\, P(x_t | y_{1:t-1})}$$

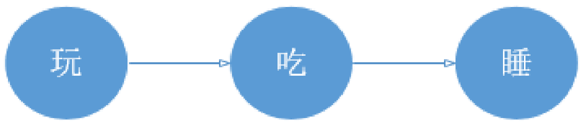


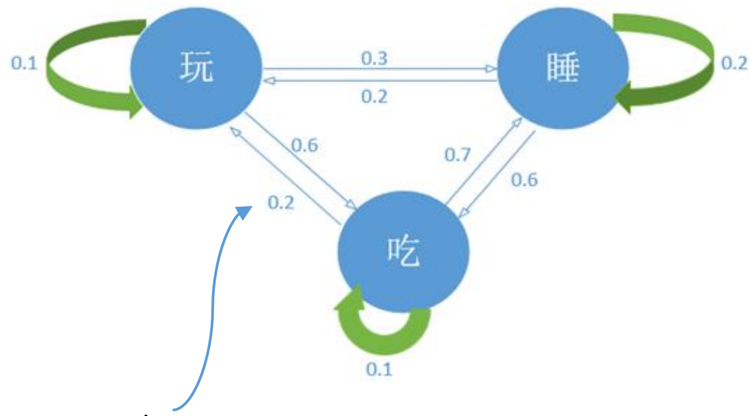


隐马尔可夫模型（Hidden Markov Model，HMM）

1，马尔可夫过程

先来看一个例子。假设几个月大的宝宝每天做三件事：玩（兴奋状态）、吃（饥饿状态）、睡（困倦状态），这三件事按下图所示的方向转移：

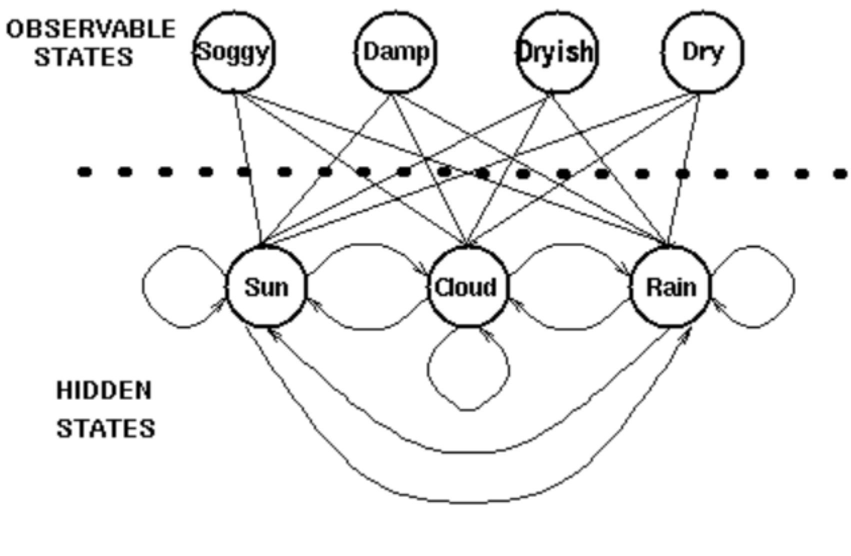




概率状态转移矩阵

2，隐马尔可夫模型　　是一种结构最简单的动态贝叶斯网络 一种有向图

我们知道‘晴天’，‘阴天’，‘雨天’的概率转换矩阵，就可以通过今天的晴天推断明天的天气，后天天气也通过明天天气来判断，这类问题可以用马尔可夫模型来描述。我们不能用前一天的天气去推断第二，三天的天气，而是要通过第二，三天的水藻的状态推测第二，三天天气情况。我们想推算的东西不能直接推算出来。



五大条件来描述一个 HMM

观测序列 O　　水藻的序列

状态序列 I　　天气的序列

初始状态概率向量　pi

状态转移概率矩阵　A

观测概率矩阵　　　B

HMM 模型的三要素　描述马尔可夫模型

初始状态概率向量　pi

状态转移概率矩阵　A

观测概率矩阵　　　B

HMM 的两个假设　隐马尔可夫模型只有在这两个假设的条件下才成立

齐次马尔可夫性假设

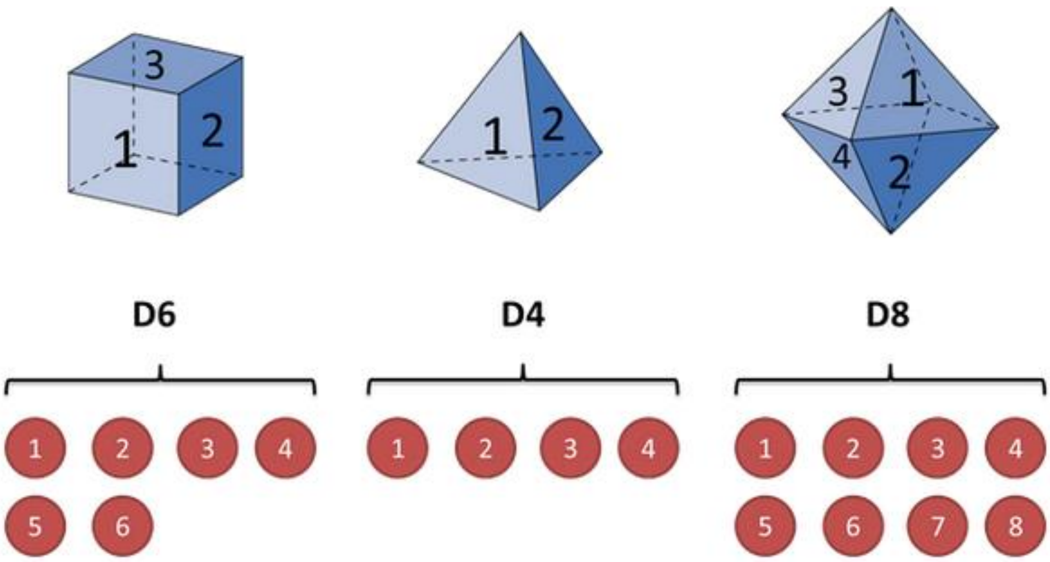
马尔可夫链的状态只跟它上一个时刻的状态有关

$$P(i_t|i_{t-1},o_{t-1},\dots,i_1,o_1)=P(i_t|i_{t-1}),t=1,2,\dots,T$$

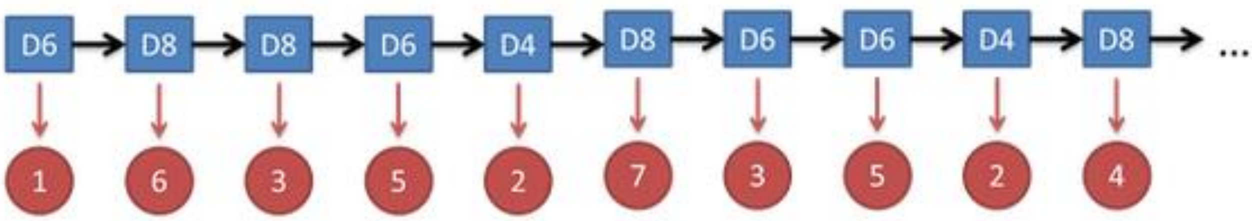
观测独立性假设

任一时刻的观测值只依赖于它在这个时刻的状态值,与以前时刻的状态值和观测值没有关系

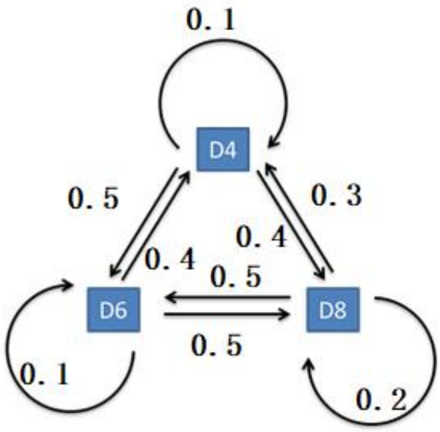
$$P(o_t|i_T,o_T,i_{T-1},o_{T-1},\dots i_{t+1},o_{t+1},i_t,i_{t-1},o_{t-1},\dots,i_1,o_1)=P(o_t|i_t)$$



隐马尔可夫模型示意图



隐含状态转换关系示意图（概率状态转移矩阵）



状态集合：Q={D4，D6，D8}, N=4

观测集合：V={1,2,3,4,5,6,7,8},M=8

观测序列：O=(1,6,3,5,2,7,3,5,2,4)

初始概率分布：$\pi=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^T$

拿完 D4 拿 D6 的概率是 0.5

状态转移概率分布：$A=\begin{bmatrix}0.1 & 0.5 & 0.4\\ 0.4 & 0.1 & 0.5\\ 0.3 & 0.5 & 0.2\end{bmatrix}$

观测概率分布：$B=\begin{bmatrix}0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0\\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0\\ 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125\end{bmatrix}$

隐马尔可夫模型解决的三类问题

概率计算问题 给定模型 model = (pi,A,B)和状态序列 I，求观测序列 O 出现的概率 P(O|model)。forward-backward 前向后向算法

解码问题 给定 模型 model=(pi,A,B)和观测序列 O(求对给定观测序列条件概率 p(i|o)最大的状态序列，即给定观测序列)，求最有可能的状态序列也是概率，使用 viterbi 算法(应用最广泛的动态规划算法)

学习问题 给定观测序列 o 求模型的参数 model=(pi,A,B)使得在这个模型下面观测概率 p(o|model)最大（极大似然）

概率计算问题：

直接计算法

给定模型 model = (A, B,π)和 n 个长度为 T 的状态序列 I = (i1,i2, ..., iT)，观测序列 O = (o1,o2, ..., oT)计算测序列 O 出现的概率 P(O|λ)。

因为存在所有长度 T 的状态序列 I,求在 n 个长度为 T 的状态序列下与观测序列的联合概率 P(O,I|model)再求和得 SUM(P(O,I|model)) = P(O|model)

通过列举 所有可能的长度为 T 的状态序列 I = (i1,i2, ..., iT)，求各个状态序列 I 与观测序列 O = (o1, o2, ..., oT)的联合概率 P(O,I|λ)，然后对所有可能的状态序列求和，得到 P(O|λ)。

1，**状态序列 I = (i1,i2, ..., iT)的概率（状态 1 的概率*状态 1 转到状态 2 的概率*...*状态 T-1 转到状态 T 的概率）模型固定，各个状态对应的概率是相互独立的，所以它们同时发生的概率是相乘关系。**

$$P(I\mid\lambda)=\pi_{i_1}a_{i_1i_2}a_{i_2i_3}\cdots a_{i_{T-1}i_T}$$

上式中的π对应初始概率分布，aij 对应概率状态转移矩阵

2，对固定的状态序列 I =(i1, i2, ..., iT)，观测序列 O = (o1, o2, ..., oT)的概率(第一个观测的概率*第二个观测的概率*...*第 T 个观测的概率)

$$P(O \mid I,\lambda) = b_{i_1}(o_1) b_{i_2}(o_2) \cdots b_{i_T}(o_T)$$

，b(a) = 观测概率矩阵的元素即概率*观测值

3，O 和 I 同时出现的概率（利用了条件概率公式）

$$\begin{aligned} P(O,I \mid \lambda) &= P(O \mid I,\lambda) P(I \mid \lambda) \\ &= \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{T-1} i_T} b_{i_T}(o_T) \end{aligned}$$

4，对所有的状态序列 I 求和，得到观测序列 O 的概率 P(O|λ) 是 O(TNT)阶

$$\begin{aligned} P(O \mid \lambda) &= \sum_{I} P(O \mid I,\lambda) P(I \mid \lambda) \\ &= \sum_{i_1, i_2, \cdots, i_T} \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{T-1} i_T} b_{i_T}(o_T) \end{aligned}$$

前向传播

给定隐马尔可夫模型 lambda = (pi,A,B)，定义时刻 t 的观测序列为 O1...Ot,对应的状态为 qi 的概率，这个概率为前向概率

$$\alpha_t(i) = P(o_1, o_2, \cdots, o_t, i_t = q_i \mid \lambda)$$

就是给定模型参数，计算一个观测序列和第 t 时刻为状态 qi 的联合概率即前向概率，观测序列 O 出现的概率是多少 它是一个递推的过程，一轮一轮的去计算。

▶ 可以递推求得前向概率$\alpha_t(i)$和观测序列概率P(O | λ)

输入：隐马尔科夫模型 λ，观测序列 O;
输出：观测序列概率 P(O|λ)
过程：

（1）初值

$$\alpha_1(i) = \pi_i b_i(o_1), \qquad i = 1,2,\cdots,N$$

（2）递推　对 $t = 1,2,\cdots,T-1$，

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji} \right] b_i(o_{t+1}), \qquad i = 1,2,\cdots,N$$

（3）终止

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

第一步计算初始值：初始值就是在第一个时间点，各个隐含状态的值是多少，右边表示在初始时刻第 i 个状态出现的概率 乘以 在这个状态下某个观测状态的概率
第二步递推计算 中括号就是当前层所有 N 个隐含状态，与下一层的第 i 个状态的链接
其中 阿尔法 是到时刻 t 部分观测序列为 o1 到 ot 且在 t 时刻下处于状态 j 的概率，而 a 是 t 时刻第 j 个状态到 t+1 时刻的第 i 个状态的转移概率得到 a，
b 是当前状态下对应观测情况下发生的概率 就像水藻和天气的关系

第三步 求解 在时刻 t 所有状态的概率求和