Probabilistic Graphical Model

数据嗨客

基本概念

独立 | 给定变量x,y，若其联合概率分布$p(x,y)$ 可以写为$p(x,y)=p(x)p(y)$. 那么，变量x和变量y独立.

条件独立 | 给定变量x,y,z，若其条件分布$p(x,y|z)$可以写为

$$p(x,y|z)=p(x||z)p(y|z).$$

那么，变量x和y在z的条件下独立，记为$x\perp\!\!\!\perp y|z$.

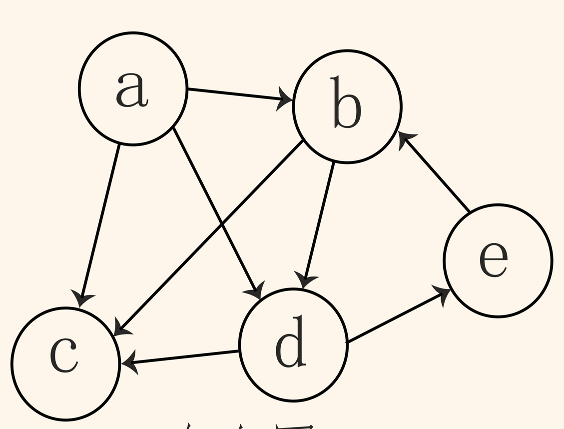
似然函数 | 给定模型参数θ，观测数据$\mathcal{D}$，模型的似然函数为

$$p(\theta|\mathcal{D})=\frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}=\frac{p(\mathcal{D}|\theta)p(\theta)}{\int_\theta p(\mathcal{D}|\theta)p(\theta)d\theta}$$

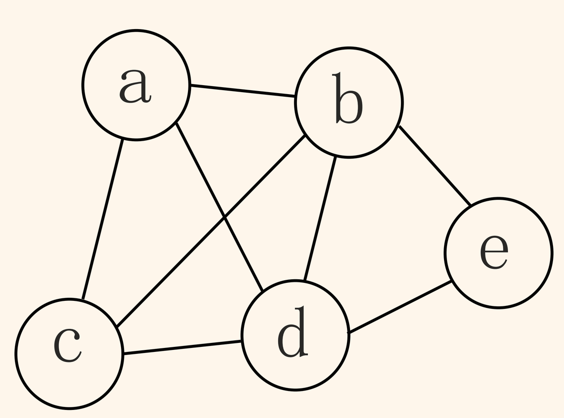
其中$p(\mathcal{D}|\theta)p(\theta)$为后验概率分布，$p(\theta)$为先验概率分布. 如果先验概率分布为常数，那么**最大后验估计**(MAP)等价于**最大似然估计**(MLE).

无向图 | 无向边和节点组成的图.

有向图 | 有向边和节点组成的图.



有向图

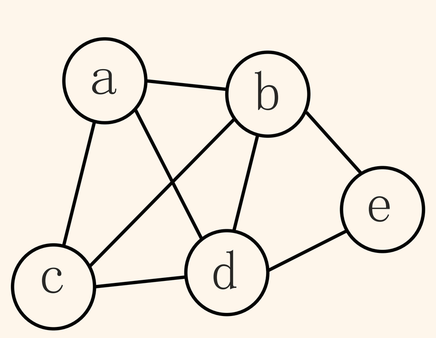


无向图

有向无环图 | 任意一条边有方向，且没有环的图.

团 | 无向图中两两连接的顶点的子集.

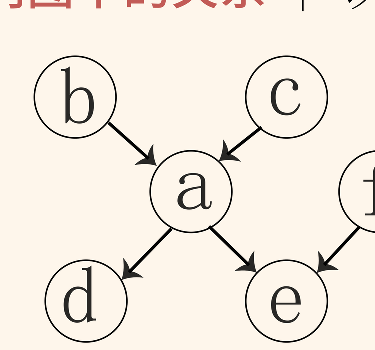
极大团 | 一个团，且不存在点与团顶点之间有边.



团 {a,b,c}

最大团 {a,b,c,d} {b,c,e}

无向图中的关系 | 以下图的节点a为例



a的父(parent)节点
pa(a)={b, c}

a的子(children)节点
ch(a)={d, e}

a的马尔可夫毯(markov blanket)
{b, c, d, e, f}

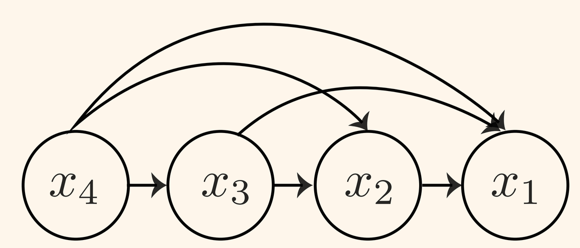
有向图模型

信念网络 | 指的是形如下式的概率分布

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^n p(x_i|\text{pa}(x_i))$$

信任网络可以表示为一个有向图，图中的有向边从父节点指向子节点，图中的节点表示因子$p(x_i|\text{pa}(x_i))$. 根据贝叶斯法则，概率分布写为

$$\begin{aligned}p(x_1,\cdots,x_n)&=p(x_1|x_2,\cdots,x_n)p(x_2,\cdots,x_n)\\&=p(x_1|x_2,\cdots,x_n)p(x_2|x_3,\cdots,x_n)p(x_3,\cdots,x_n)\\&=p(x_n)\prod_{i=1}^{n-1}p(x_i|x_{i+1},\cdots,x_n)\end{aligned}$$



根据上图，$p(x_1,x_2,x_3,x_4)$为

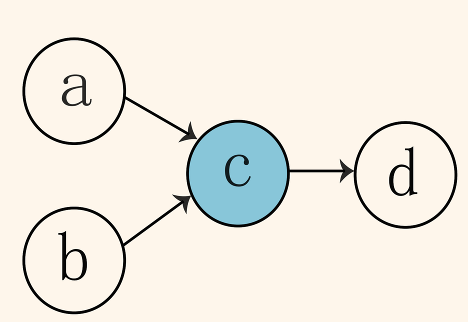
$$p(x_1|x_2,x_3,x_4)p(x_2|x_3,x_4)p(x_3|x_4)p(x_4)$$

阻塞 | 给定B,C,D为有向图中任意不相交的节点集. 若D为条件集合（D的节点被观测到），从B中节点到C中节点的所有可能路径中，某条路径被堵塞，如果该路径包含节点n满足以下任一条件：

✓在节点n处，路径上的有向边呈现“头对尾”或者“尾对尾”的形态，且$n\in D$;

✓在节点n处，有向边呈现“投对头”的形态，且$n,\text{ch}(n)\notin D$.

有向分离 | 如果从B到C的所有路径都被堵塞，那么B和C被D有向分离. 因此，有向图所表示的有关这些节点的联合概率分布满足$B\perp\!\!\!\perp C|D$.



在上图中，从a到b的路径没有被c阻塞.
因为在c处，有向边呈“头对头”的形态，但是节点c在条件节点集合中，即c被观测到，用阴影表示.

无向图模型

势函数 | 变量x的一个非负函数，即$\phi(x)\geq 0$. 变量的分布是一个特例，视为一个归一化的势函数，$\sum_x\phi(x)=1$. 分布$p(a,b,c)$还可以分解为

$$p(a,b,c)=\frac{1}{Z}\phi(a,b)\phi(b,c),Z=\sum_{a,b,c}\phi(a,b)\phi(b,c)$$

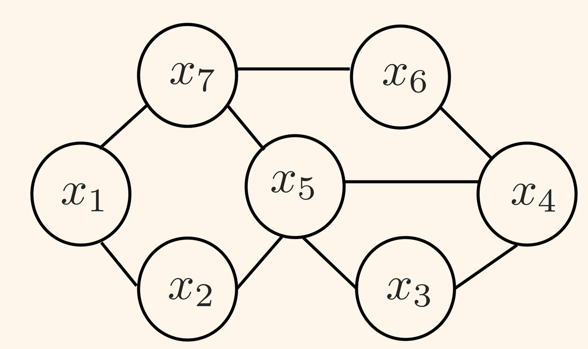
马尔可夫网络 | 给定$\mathcal{X}=\{x_1,\cdots,x_n\}$，马尔可夫网络是变量集合$X_c\in\mathcal{X}$定义的势函数的乘积.

$$p(x_1,\cdots,x_n)=\frac{1}{Z}\prod_{c=1}^C\phi_c(X_c)$$

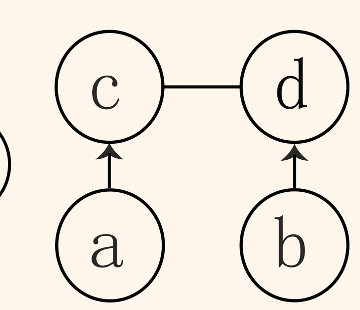
其中X_c为网络对应的无向图中的极大团.

分离 | 给定无向图G中三个不相交的节点子集B,C,D，如果从B中节点到C中节点的每一条路径都需要通过D，那么B和C被D分离.

全局马尔可夫性 | 给定无向图的三个不相交的节点子集B,C,D，且子集D将子集B和子集C分离，则有$B\perp\!\!\!\perp C|D$.



马尔可夫网络



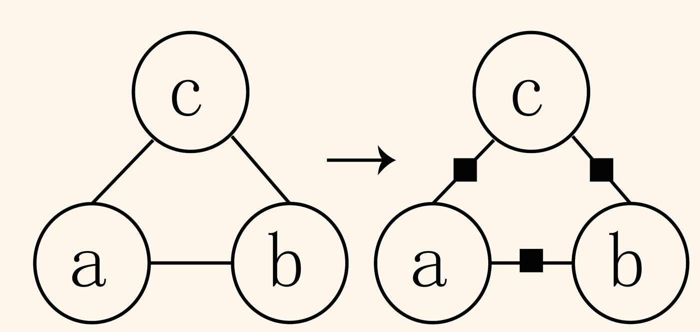
链式图

由上图的马尔可夫网络可知$x_1\perp\!\!\!\perp x_7|\{x_3,x_4,x_5\}$.

链式图模型 | 同时包含有向边和无向边的图模型.

链式成分 | 链式图去掉有向边，剩余的相互连接的部分. 上图的链式成分有$(a),(b),(c,d)$.

因子图模型 | 给定$p(x_1,\cdots,x_n)=\prod_i\psi_i(X_i)$，因子图模型中的方块对应函数中的因子$\psi_i$，节点表示变量$x_j$，对于$x_j\in X_i$，存在一条无向边连接$\psi_i$和$x_j$. 例如，$p(a,b,c)=\phi(a,b)\phi(b,c)\phi(a,c)$.



马尔可夫网络　　因子图

independence map | 给定概率分布P以及对应的图G，如果从G中推断出的条件独立在分布P中真实存在，那么G为分布P的I-map.

dependence map | 给定概率分布P以及对应的图G，如果分布P中表示的条件独立在G中真实存在，那么G为分布P的D-map.

图模型的推断

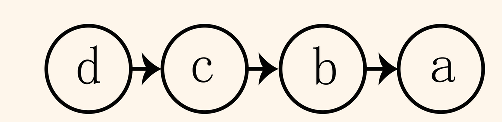
边缘推断 | 若$p(x_1,\cdots,x_4),x_1=b$，边缘推断可为

$$p(x_4|x_1=b)\propto\sum_{x_2,x_3}p(x_1=b,x_2,x_3,x_4)$$

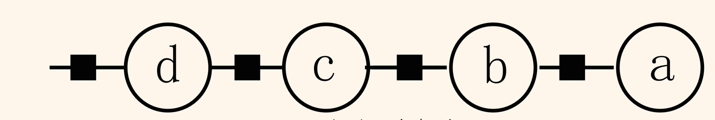
变量消除 | 逐次消除分布的变量，计算边缘分布.

$$p(a)=\sum_{b,c,d}p(a,b,c,d)=\sum_{b,c,d}p(a|b)p(b|c)p(c|d)p(d)$$

$$=\sum_b p(a|b)\overbrace{\sum_c p(b|c)\underbrace{\sum_d p(c|d)p(d)}_{\gamma_d(c)}}^{\gamma_c(b)}$$



马尔可夫链



因子图

sum-product算法 | 在因子图上计算边缘分布.

$$p(a,b,c)=\sum_{b,c,d}\phi_1(a,b)\phi_2(b,c)\phi_3(c,d)\phi_4(d)$$

$$=\phi_1(a,b)\phi_2(b,c)\underbrace{\sum_d p(a|b)\phi_3(c,d)\phi_4(d)}_{\mu_{d\to c}(c)}$$

其中$\mu_{d\to c}(c)$表示从d到c传递的信息.

Junction Tree算法 | 变量消除等方法解决单链接图的推断问题，如果概率图模型属于多连接图，推断需使用Junction Tree算法：

✓ 对原始的图进行**moralisation**，融合父节点.
✓ 对图进行**triangulation**构造**chordal graph**.
✓ 寻找graph中的极大团，构造Junction Tree.
✓ 在树结构上运行广义的sum-product算法.