机器学习

Machine learning

第九章 概率图模型

Probabilistic Graphical Model

授课人：周晓飞
zhouxiaofei@iie.ac.cn
2021-12-10

第九章 概率图模型

9.1 有向图模型：贝叶斯网络

9.2 无向图模型：马尔可夫随机场

9.3 学习与推断

9.4 近似推断

9.5 实例模型

第九章 概率图模型

9.1 有向图模型：贝叶斯网络

9.2 无向图模型：马尔可夫随机场

9.3 学习与推断

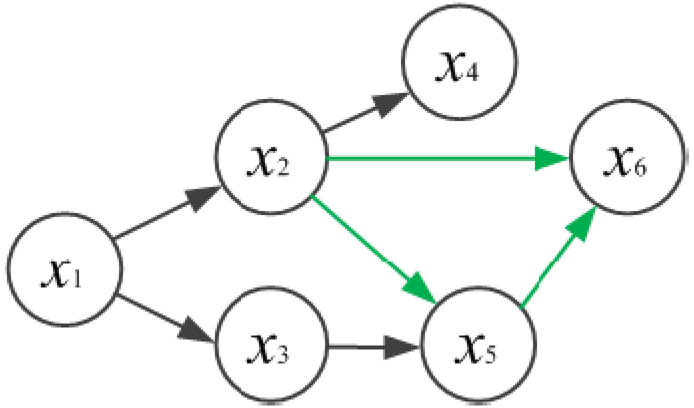
9.4 近似推断

9.5 实例模型

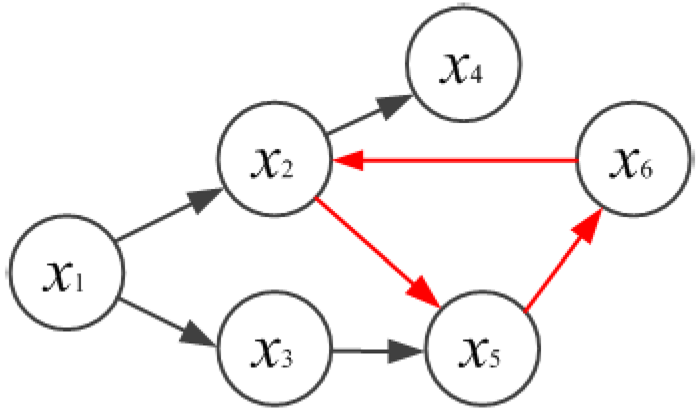
图结构：有向无环图（DAG）

结点：一个或一组随机变量。

边：随机变量之间的单向、直接影响（加班 →生病）。



YES



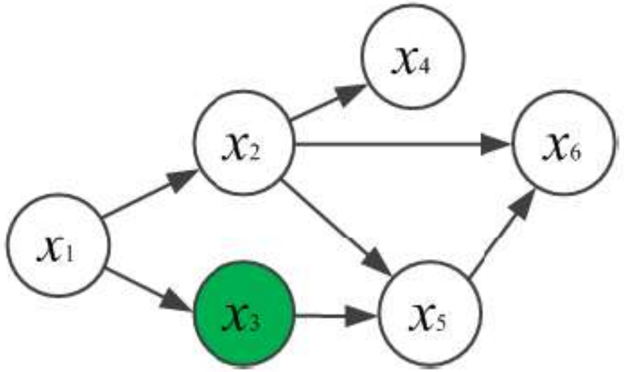
NO

有向图模型：贝叶斯网络

图结构：有向无环图（DAG）

结点：一个或一组随机变量。

边：随机变量之间的单向、直接影响（加班 → 生病）。



- 当前结点：x_3
- 父结点：$\{x_1\}$
- 子结点：$\{x_5\}$
- 祖先结点：$\{x_1\}$
- 后代结点：$\{x_5, x_6\}$
- 马尔可夫毯：$\{x_1, x_2, x_5\}$

祖先：所有长辈节点

后代：所有后继结点

贝叶斯网 MB：子结点、父结点、子结点的父节点

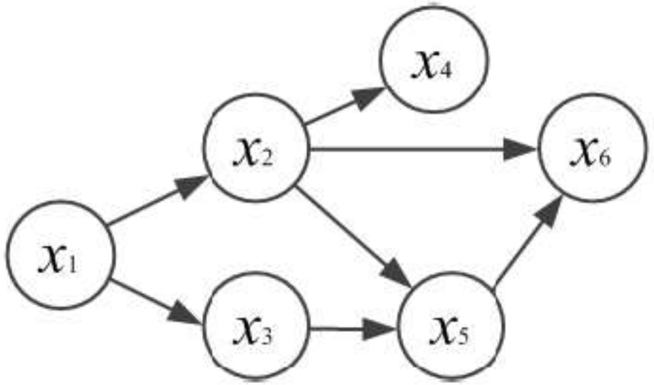
有向图模型：贝叶斯网络

联合概率分布

分解形式：

$$P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i \mid \mathbf{x}_{\pi i})$$

其中，$\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ ；$\mathbf{x}_{\pi i}$ 为 x_i 所有父结点构成的集合。

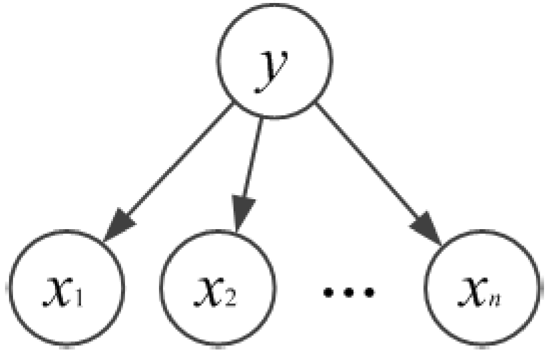


$$\begin{aligned} & P(x_1, x_2, x_3, x_4, x_5, x_6) \\ = & P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2) \\ & P(x_5|x_2, x_3)P(x_6|x_2, x_5) \end{aligned}$$

有向图模型：贝叶斯网络

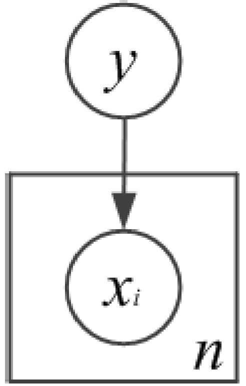
示例：朴素贝叶斯

$\mathbf{x}=(x_1,x_2,...,x_n)$ 为特征向量，y 是类别标签。



普通变量关系图

=

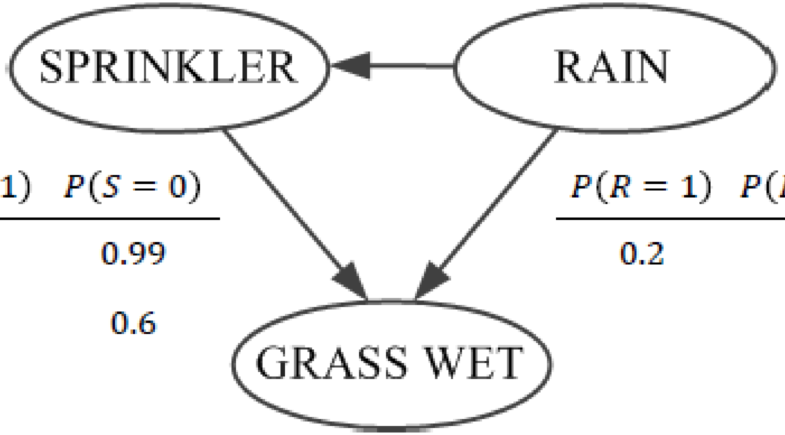


盘式记法

$$P(\mathbf{x},y)=P(y)\prod_{i=1}^{n}P(x_i|y)$$

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

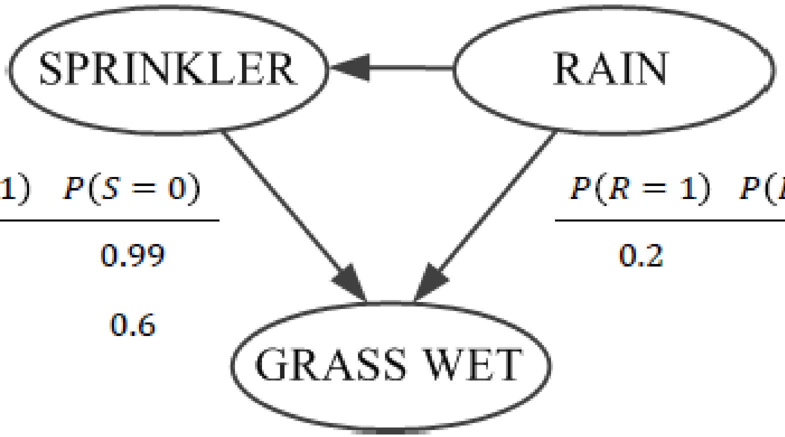
S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

- 同时观测到下雨、给草坪浇水、草坪湿的概率有多大？
- 当已知不下雨时，观测到草坪湿的概率有多大？
- 当观测到草坪湿以后，推测下雨的概率有多大？
- 当观测到草坪湿以后，推测给草坪浇过水的概率有多大？

。。。 。。。

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

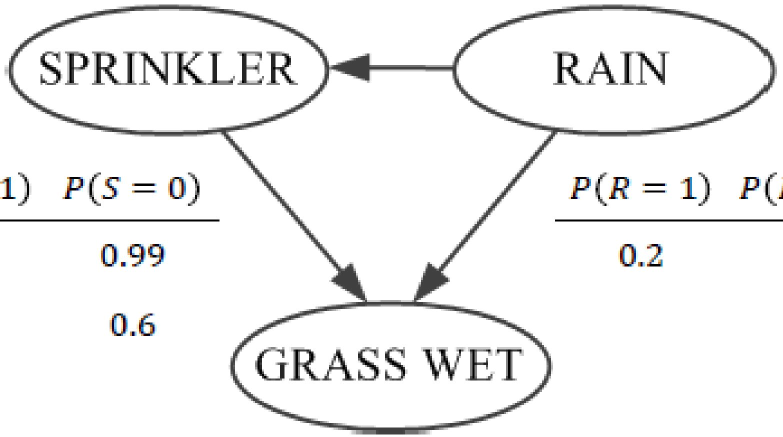
、

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

$$P(G,S,R)=P(G\mid S,R)P(S\mid R)P(R)$$

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

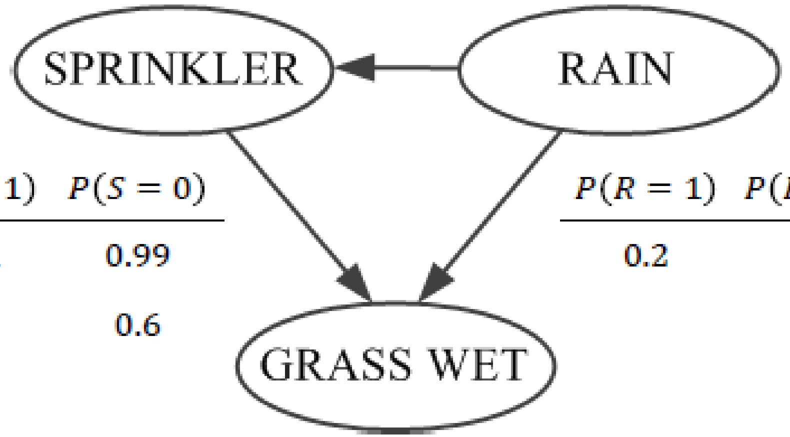
$$P(G,S,R)=P(G\mid S,R)P(S\mid R)P(R)$$

同时观测到下雨、给草坪浇水、草坪湿的概率有多大?

$$\begin{aligned}
&P(G=1,S=1,R=1)\\
&=P(G=1|S=1,R=1)P(S=1|R=1)P(R=1)\\
&=0.99\times 0.01\times 0.2=0.00198
\end{aligned}$$

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

$$P(G,S,R)=P(G\mid S,R)P(S\mid R)P(R)$$

当已知不下雨时，观测到草坪湿的概率有多大？

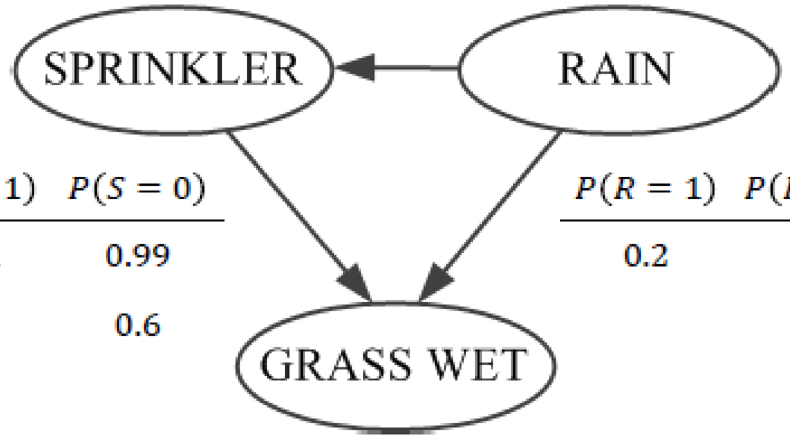
$$P(G=1|R=0)$$

$$=\frac{P(G=1,R=0)}{P(R=0)}=\frac{\sum_{S\in\{1,0\}}P(G=1,S,R=0)}{P(R=0)}$$

$$=\frac{0.288+0}{0.8}=0.36$$

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

$$P(G,S,R)=P(G\,|\,S,R)P(S\,|\,R)P(R)$$

当观测到草坪湿以后，推测下雨的概率有多大?

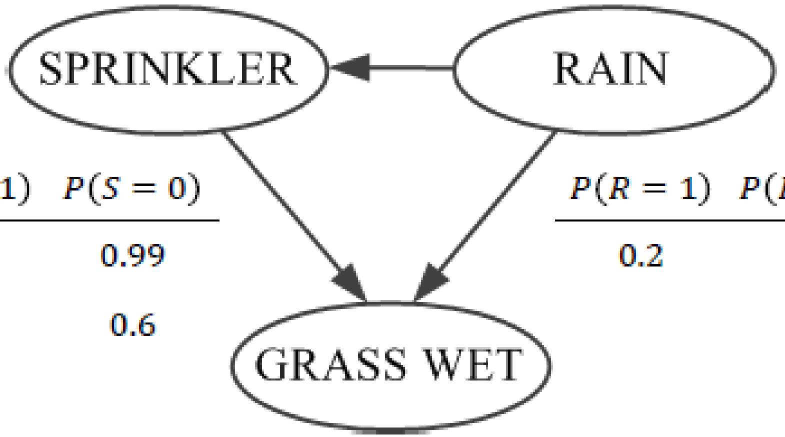
$$P(R=1|G=1)$$

$$=\frac{P(G=1,R=1)}{P(G=1)}=\frac{\sum_{S\in\{1,0\}}P(G=1,S,R=1)}{\sum_{S,R\in\{1,0\}}P(G=1,S,R)}$$

$$=\frac{0.00198+0.1584}{0.00198+0.288+0.1584+0.0}\approx 0.3577$$

有向图模型：贝叶斯网络

示例：草坪问题



R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

$P(R=1)$	$P(R=0)$
0.2	0.8

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

$$P(G,S,R)=P(G\mid S,R)P(S\mid R)P(R)$$

当观测到草坪湿以后，推测给草坪浇过水的概率有多大?

$$P(S=1|G=1)$$

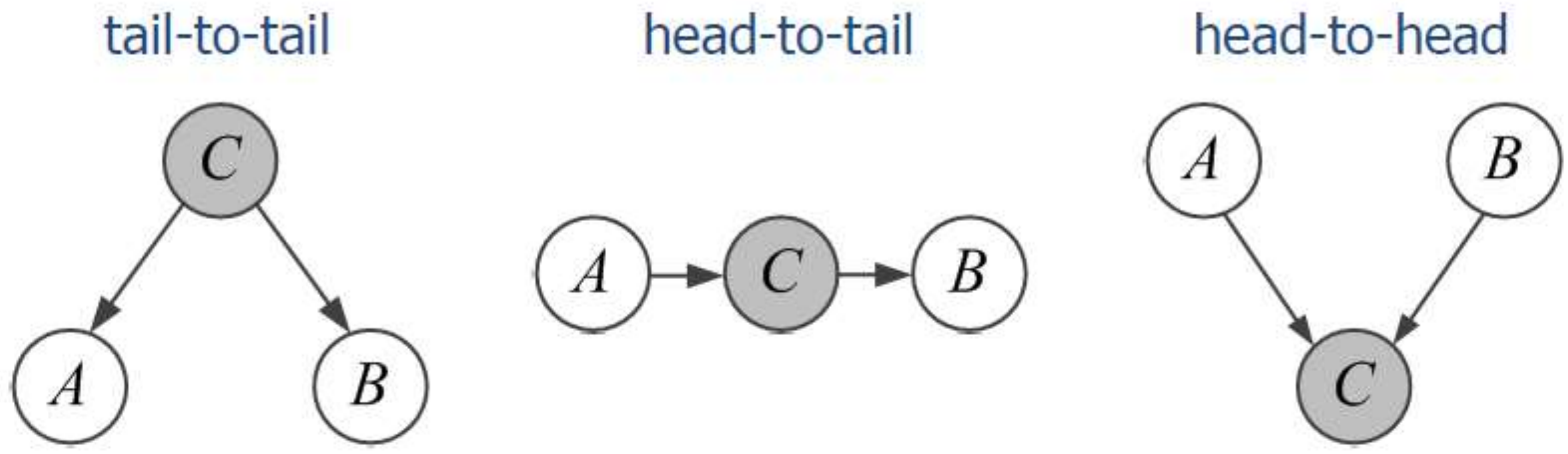
$$=\frac{P(G=1,S=1)}{P(G=1)}=\frac{\sum_{R\in\{1,0\}}P(G=1,S=1,R)}{\sum_{S,R\in\{1,0\}}P(G=1,S,R)}$$

$$=\frac{0.00198+0.288}{0.00198+0.288+0.1584+0.0}\approx 0.6467$$

有向图模型：贝叶斯网络

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

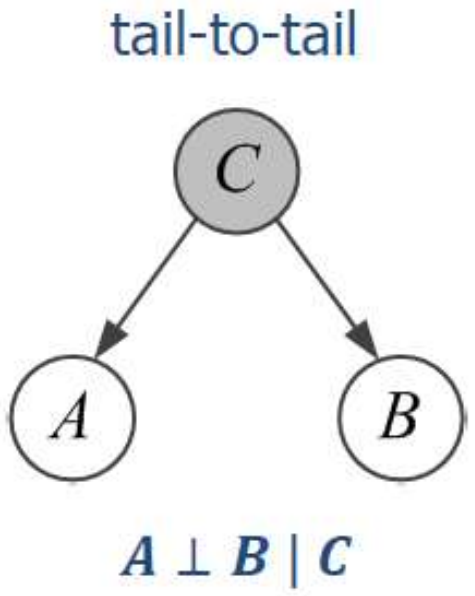


$A \perp B \mid C$ $A \perp B \mid C$ $A \not\perp B \mid C$ 或其任意后代

有向图模型：贝叶斯网络

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。



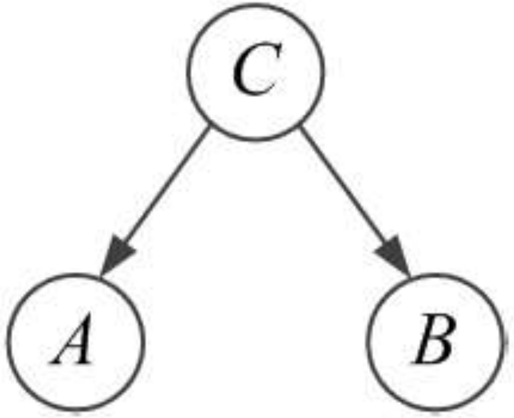
$$P(A,B|C)=\frac{P(A,B,C)}{P(C)}$$
$$=\frac{P(C)P(A|C)P(B|C)}{P(C)}$$
$$=P(A|C)P(B|C)$$

有向图模型：贝叶斯网络

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

tail-to-tail



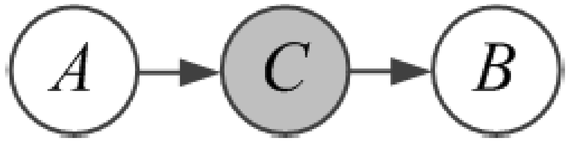
$P(A, B) \neq P(A)P(B)$

$A \not\perp B$

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

head-to-tail



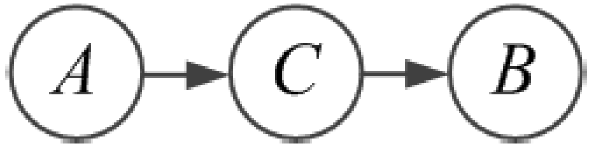
$$A \perp B \mid C$$

$$
\begin{aligned}
P(A,B|C) &= \frac{P(A,B,C)}{P(C)} \\
&= \frac{P(A)P(C|A)P(B|C)}{P(C)} \\
&= \frac{P(A,C)P(B|C)}{P(C)} \\
&= P(A|C)P(B|C)
\end{aligned}
$$

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

head-to-tail



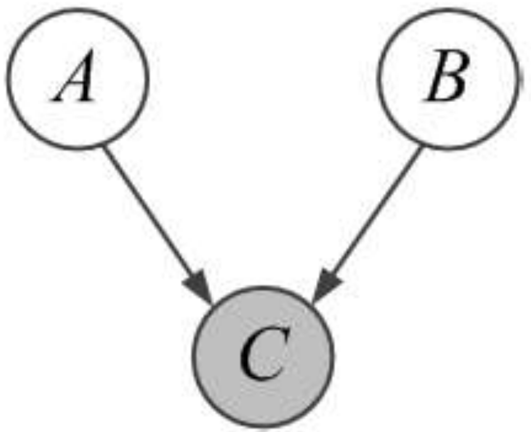
$$P(A,B) \neq P(A)P(B)$$

$$A \not\perp B$$

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

head-to-head



$$P(A,B|C) \neq P(A|C)P(B|C)$$

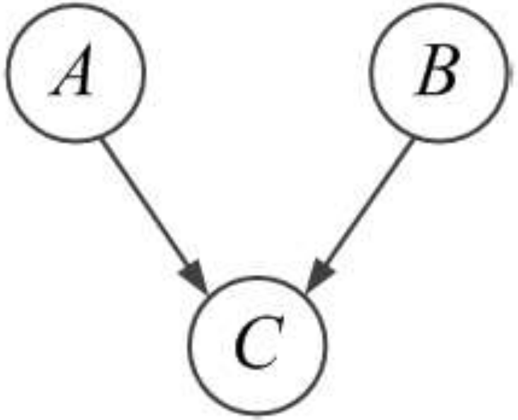
$A \not\perp B \mid C$ 或其任意后代

有向图模型：贝叶斯网络

条件独立性

D-分离准则（D-separation criterion）：判断贝叶斯网络结点之间的条件独立性。

head-to-head



$A \perp B$

$$
\begin{aligned}
P(A,B) &= \sum_C P(A,B,C) \\
&= \sum_C P(A)P(B)P(C|A,B) \\
&= P(A)P(B)\sum_C P(C|A,B) \\
&= P(A)P(B)
\end{aligned}
$$

有向图模型：贝叶斯网络

条件独立性

贝叶斯网络的全局马尔科夫性：给定结点集合 A，B，C，若 A 到 B 中结点的所有无向路径都是被 C 阻塞的（blocked），则称 A 和 B 被 C ***D*-分离（D-separated）**，即 A 和 B 关于 C 条件独立。

若一条无向路径包含结点 x 满足以下条件之一，则称其是阻塞的：

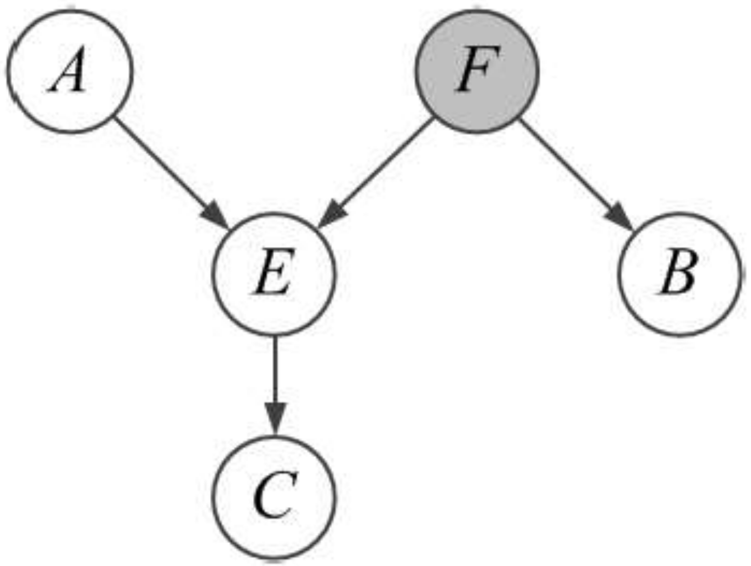
（1）x 是 tail-to-tail 或 head-to-tail 结点，并且 x 包含在 C 中。

（2）x 是 head-to-head 结点，并且 x(及 x 的任意后代均)不包含在 C 中。

有向图模型：贝叶斯网络

条件独立性

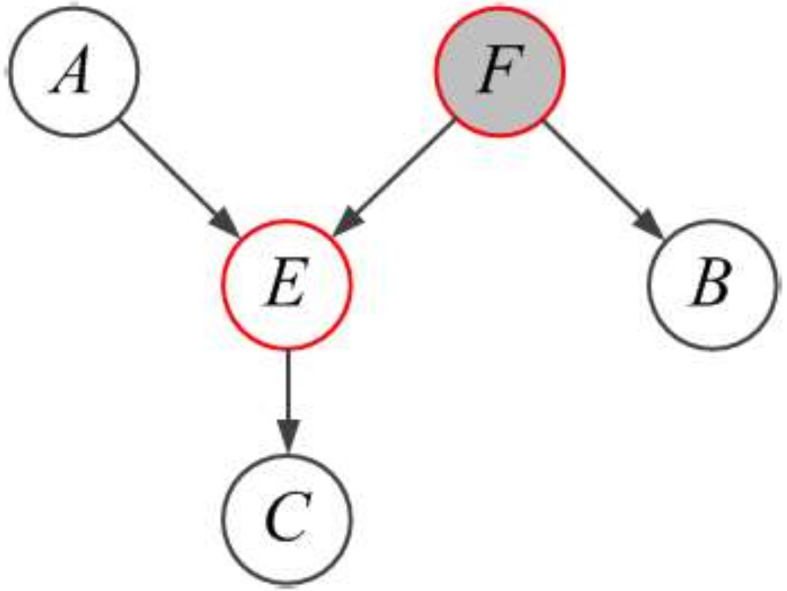
例子：A 和 B 是否关于 F 条件独立？



有向图模型：贝叶斯网络

条件独立性

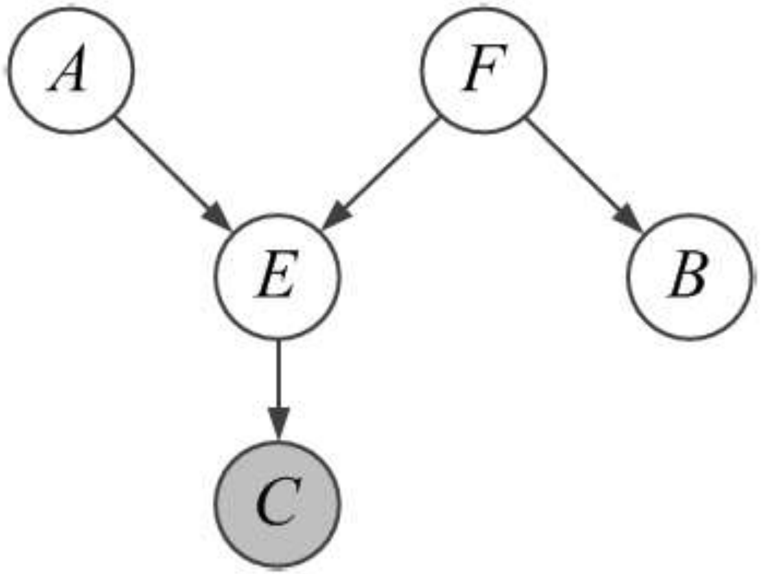
例子：A 和 B 是否关于 F 条件独立？



$A \perp B \mid F$

条件独立性

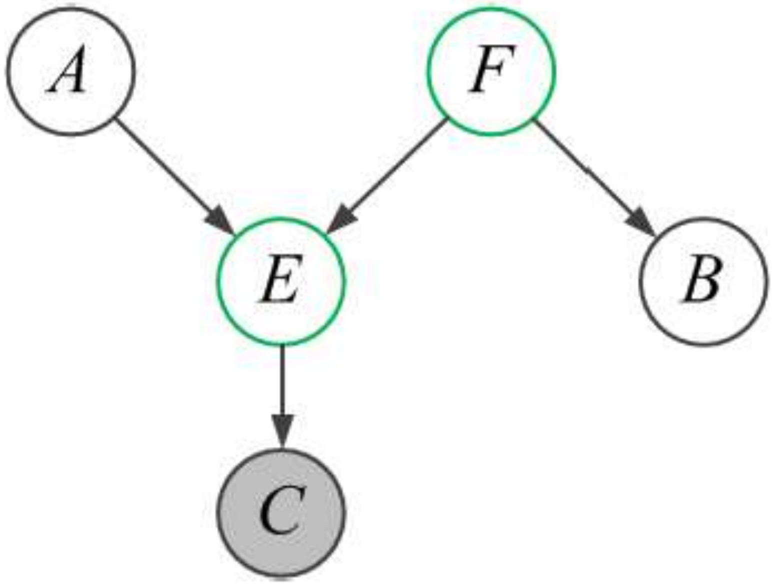
例子：A 和 B 是否关于 C 条件独立？



有向图模型：贝叶斯网络

条件独立性

例子：A 和 B 是否关于 C 条件独立？



$A \not\perp B \mid C$

有向图模型：贝叶斯网络

条件独立性

贝叶斯网络的局部马尔科夫性：

（1）给定某变量的父结点，则该变量条件独立于所有其他非其后代结点。

$$x_v \perp x_{V\backslash\{v\}\backslash PA(v)\backslash DE(v)} \mid x_{PA(v)}$$

（2）给定某变量的马尔可夫毯（父结点，子结点，子结点的父结点），则该变量条件独立于其他变量。

$$x_v \perp x_{V\backslash\{v\}\backslash MB(v)} \mid x_{MB(v)}$$

第九章 概率图模型

9.1 有向图模型：贝叶斯网络

9.2 无向图模型：马尔可夫随机场

9.3 学习与推断

9.4 近似推断

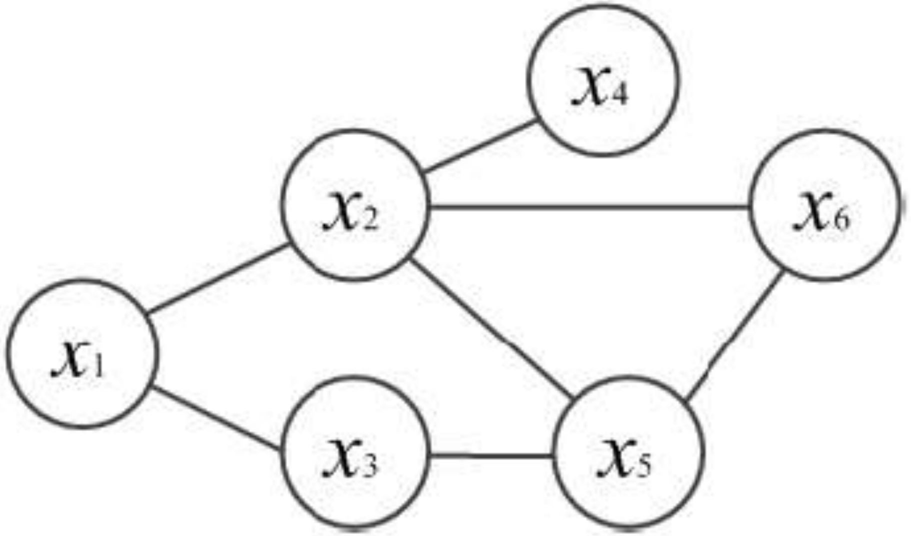
9.5 实例模型

无向图模型：马尔可夫随机场

图结构：无向图

结点：一个或一组随机变量。

边：随机变量之间的相互依赖（非“因果关系”）。

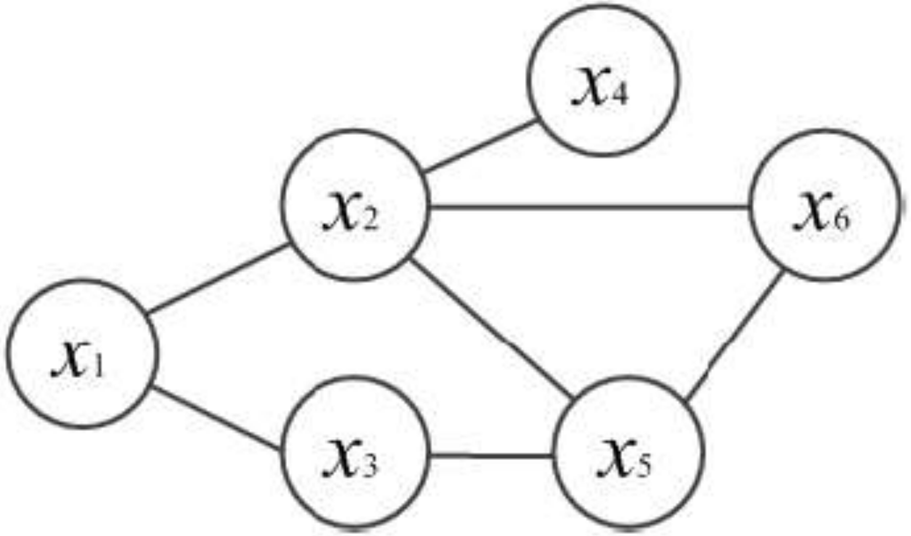


无向图模型：马尔可夫随机场

图结构：无向图

结点：一个或一组随机变量。

边：随机变量之间的相互依赖（非"因果关系"）。



团：对于图中的结点子集，若其中任意两个节点之间都有连边，则称该结点子集为一个团（clique）。

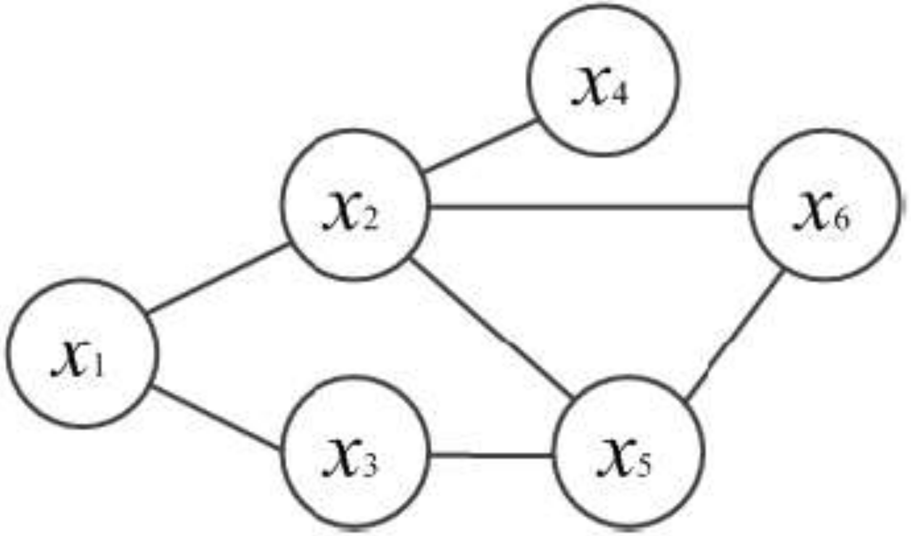
极大团：若在团中加入其他任意一个结点都不再形成团，则称该团为极大团（maximal clique）。

无向图模型：马尔可夫随机场

图结构：无向图

结点：一个或一组随机变量。

边：随机变量之间的相互依赖（非“因果关系”）。



极大团：$\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_4\}, \{x_3, x_5\}, \{x_2, x_5, x_6\}$

无向图模型：马尔可夫随机场

联合概率分布

分解形式：

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{C\in\mathbb{C}}\psi_C(\mathbf{x}_C)$$

其中，$\mathbf{x}=\{x_1,x_2,...,x_n\}$；$\mathbb{C}$ 为团集合，$\mathbf{x}_C$ 为团 C 对应的变量集合。

ψ_C 为定义在团 C 上的非负势函数（potential function）。

Z 是归一化因子：

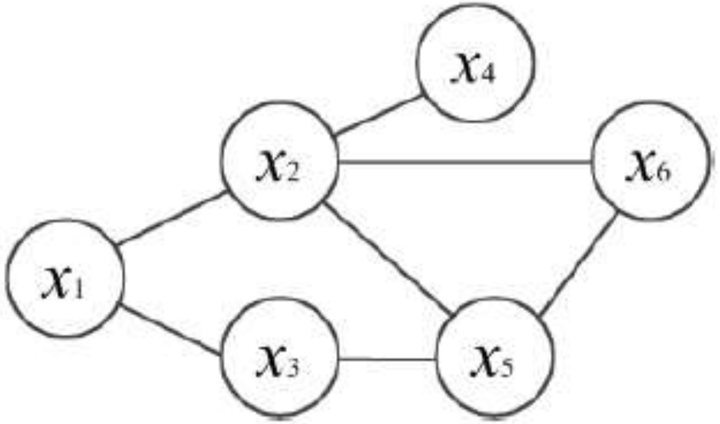
$$Z = \sum_{\mathbf{x}}\prod_{C\in\mathbb{C}}\psi_C(\mathbf{x}_C)$$

无向图模型：马尔可夫随机场

联合概率分布

分解形式：

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{C\in\mathbb{C}}\psi_C(\mathbf{x}_C)$$



$$P(x_1,x_2,x_3,x_4,x_5,x_6)$$
$$=\frac{1}{Z}\psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)\psi_{24}(x_2,x_4)$$
$$\psi_{35}(x_3,x_5)\psi_{256}(x_2,x_5,x_6)$$

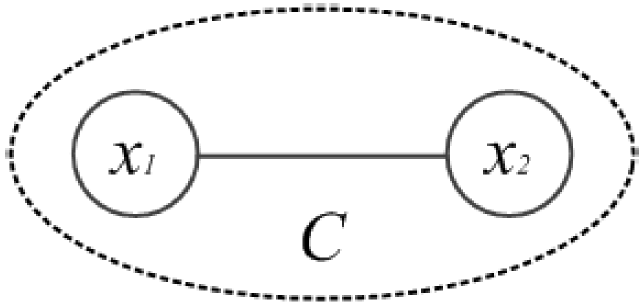
无向图模型：马尔可夫随机场

示例：最简单的马尔科夫随机场

$x_1, x_2 \in \{-1,1\}$ ；$C = \{x_1, x_2\}$ 为团；$\psi_C(x_1, x_2) = e^{a x_1 x_2}$

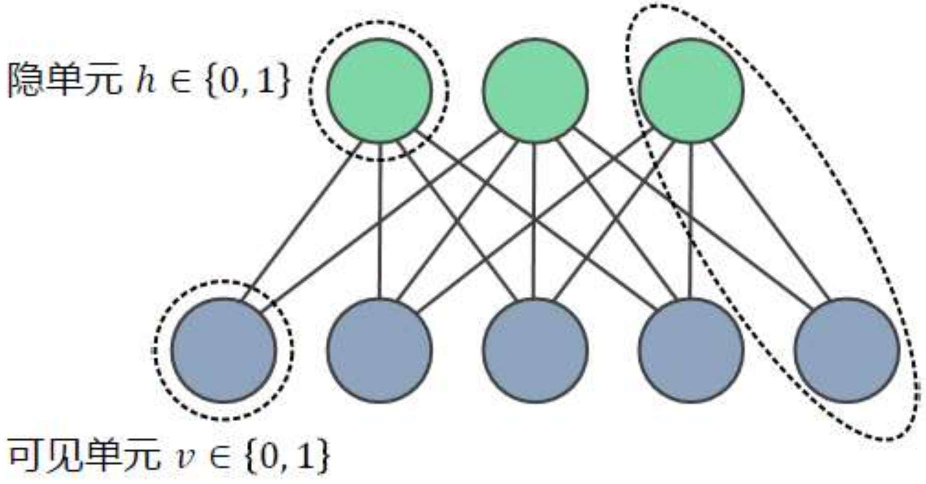
规一化因子 $Z = \sum_{x_1, x_2} \psi_C(x_1, x_2) = 2e^{a} + 2e^{-a}$

联合概率分布 $P(x_1, x_2) = \frac{1}{Z}\psi_C(x_1, x_2) = e^{a x_1 x_2} / (2e^{a} + 2e^{-a})$



无向图模型：马尔可夫随机场

示例：受限玻尔兹曼机

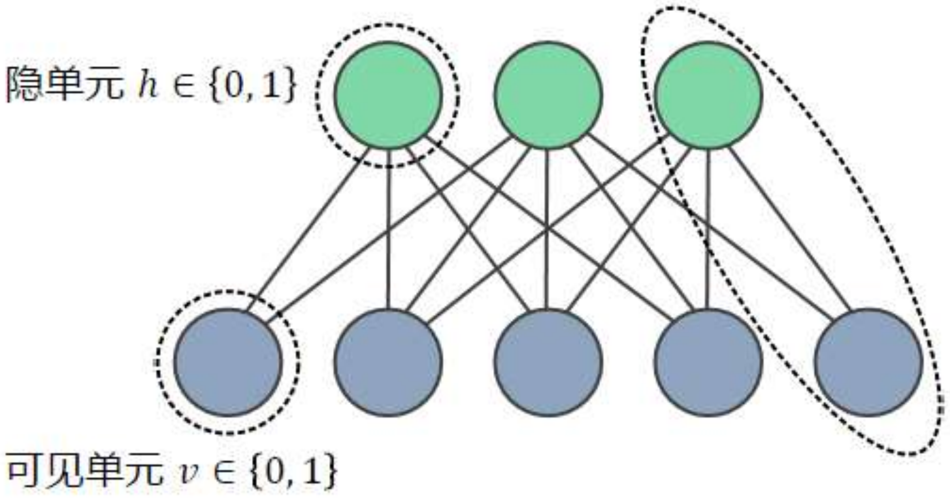


团集合：$\{\{v_i\},\{h_j\},\{v_i,h_j\}\}$

势函数： $\psi_V(v_i) = e^{a_i v_i}$ $\psi_H(h_j) = e^{b_j h_j}$ $\psi_{VH}(v_i, h_j) = e^{w_{ij} v_i h_j}$

无向图模型：马尔可夫随机场

示例：受限玻尔兹曼机



联合概率分布：$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z}\exp\left(\sum_i a_i v_i + \sum_j b_j h_j + \sum_{i,j} w_{ij} v_i h_j\right)$$

无向图模型：马尔可夫随机场

示例：图像去噪

团集合：$\{\{x_i, y_i\}, \{x_i, x_j\}\}$

势函数： $\psi_{XY}(x_i, y_i) = e^{\eta_i x_i y_i}$ $\psi_{XX}(x_i, x_j) = e^{\mu_{ij} x_i x_j}$

无向图模型：马尔可夫随机场

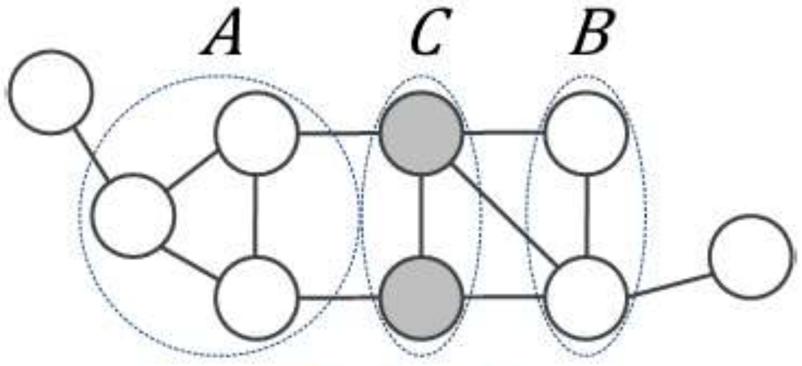
示例：图像去噪

联合概率分布：$$P(\mathbf{x},\mathbf{y})=\frac{1}{Z}\exp\left(\sum_{i\in V}\eta_i x_i y_i+\sum_{ij\in E}\mu_{ij}x_i x_j\right)$$

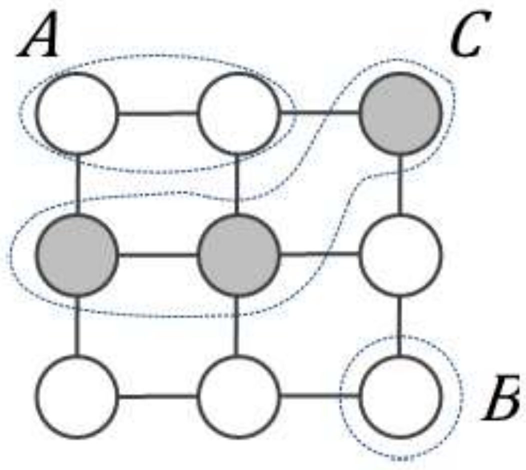
无向图模型：马尔可夫随机场

条件独立性

马尔可夫随机场的全局马尔科夫性：给定结点集合 A，B，C，若从 A 中的结点到 B 中结点必须经过 C 中的结点，则称 A 和 B 被 C 分离，即 A 和 B 关于 C 条件独立。



$A \perp B \mid C$



$A \perp B \mid C$

无向图模型：马尔可夫随机场

条件独立性

局部马尔科夫性：给定某变量的马尔可夫毯（邻接变量），则该变量条件独立于其他变量。

$$x_v \perp x_{V\backslash\{v\}\backslash MB(v)} \mid x_{MB(v)}$$

成对马尔科夫性：给定其他所有变量，两个非相邻变量条件独立。

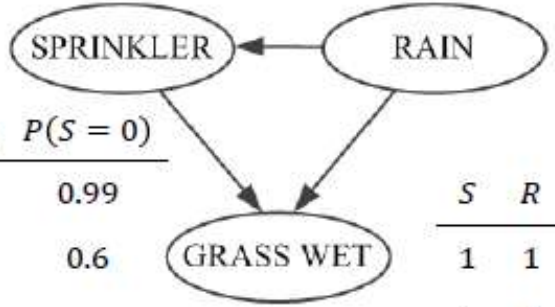
$$x_u \perp x_v \mid x_{V\backslash\{u,v\}} \quad if\,(u,v) \notin E$$

第九章 概率图模型

学习与推断

基本定义

推断：已知联合概率分布 $P(x_1,x_2,\dots,x_n)$，估计 $P(\mathbf{x}_Q|\mathbf{x}_E)$，其中 $\mathbf{x}_Q \cap \mathbf{x}_E=\emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E$ 是集合 $\{x_1,x_2,\dots,x_n\}$的子集。$\mathbf{x}_Q$ 是问题变量，$\mathbf{x}_E$ 是证据变量。



$P(R=1)$	$P(R=0)$
0.2	0.8

R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

$$P(S=1|G=1)=?$$

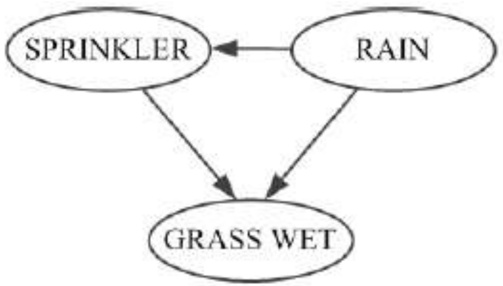
$$P(R=1|G=1)=?$$

学习与推断

基本定义

学习：从观测数据 $x^{(1)}, x^{(2)},\ldots,x^{(m)}$中学习联合概率分布 $P(x_1,x_2,\ldots,x_n)$，寻找最符合观测数据的概率图模型。

$$(G,S,R) = (1,1,0)$$
$$(G,S,R) = (1,1,1)$$
$$\vdots$$
$$(G,S,R) = (1,0,1)$$
$$(G,S,R) = (0,0,0)$$



结构学习 (structure learning)

$P(R=1)$	$P(R=0)$
0.2	0.8

R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

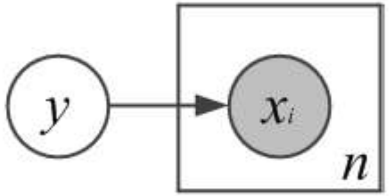
参数学习 (parameter learning)

推断

已知联合概率分布 $P(x_1,x_2,\ldots,x_n)$，估计

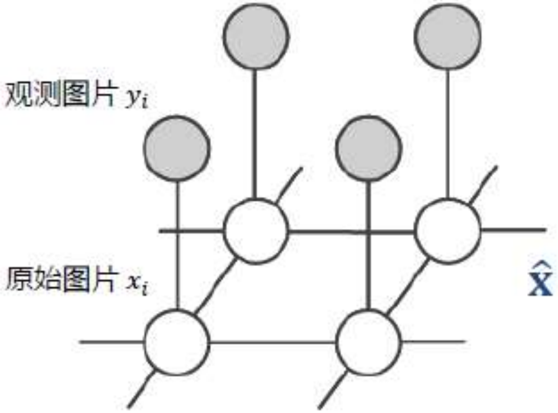
$$P(\mathbf{x}_Q \mid \mathbf{x}_E) = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{P(\mathbf{x}_E)} = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{\sum_{\mathbf{x}_Q} P(\mathbf{x}_Q,\mathbf{x}_E)}$$

其中 $\mathbf{x}_Q \cap \mathbf{x}_E = \emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E = \{x_1,x_2,\ldots,x_n\}$。（考虑 $\mathbf{x}_Q \cup \mathbf{x}_E \neq \{x_1,x_2,\ldots,x_n\}$？）



朴素贝叶斯

$\hat{y} = \operatorname{argmax} P(y|\mathbf{x})$



图像去噪

$\hat{\mathbf{x}} = \operatorname{argmax} P(\mathbf{x}|\mathbf{y})$

学习与推断

推断

已知联合概率分布 $P(x_1,x_2,\ldots,x_n)$，估计

$$P(\mathbf{x}_Q \mid \mathbf{x}_E) = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{P(\mathbf{x}_E)} = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{\boxed{\sum_{\mathbf{x}_Q} P(\mathbf{x}_Q,\mathbf{x}_E)}}$$

核心问题

其中 $\mathbf{x}_Q \cap \mathbf{x}_E = \emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E = \{x_1,x_2,\ldots,x_n\}$。（考虑 $\mathbf{x}_Q \cup \mathbf{x}_E \neq \{x_1,x_2,\ldots,x_n\}$？）

枚举：假设 $\mathbf{x}_Q$ 有 k 个变量，每个变量的取值个数的期望是 r，则时间复杂度为 r^k。

推断的核心问题：如何高效地计算边际分布 $P(\mathbf{x}_E)=\sum_{x_Q} P(x_Q|x_E)$。

学习与推断

推断方法

精确推断：计算 $P(x_E)$或 $P(x_Q|x_E)$的精确值。

（1）变量消去（variable elimination）。

（2）信念传播（belief propagation）。

近似推断：在较低的时间复杂度下获得原问题的近似解。

（1）前向采样（forward sampling）。

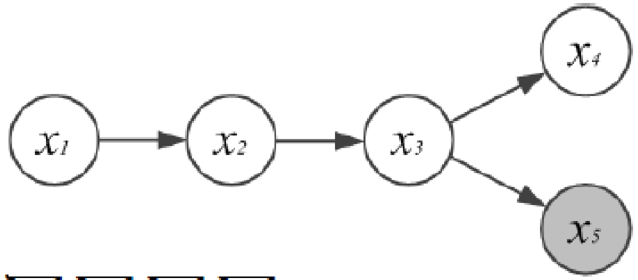
（2）吉布斯采样（Gibbs sampling）。

学习与推断

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



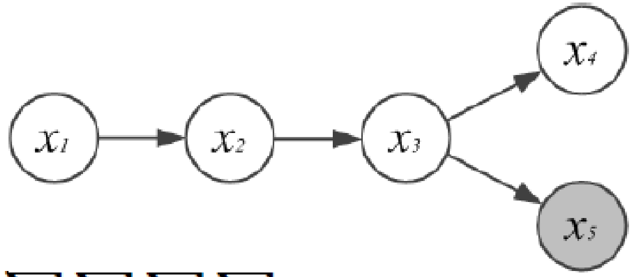
推断 $P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1, x_2, x_3, x_4, x_5)$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



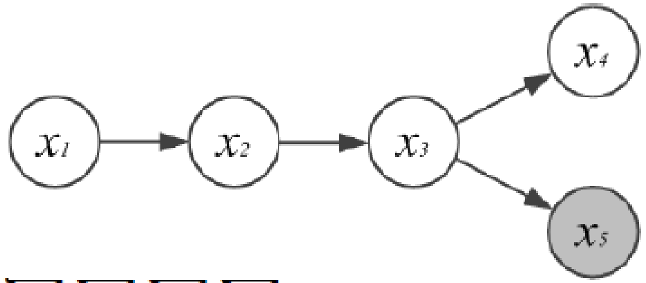
$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1)$$

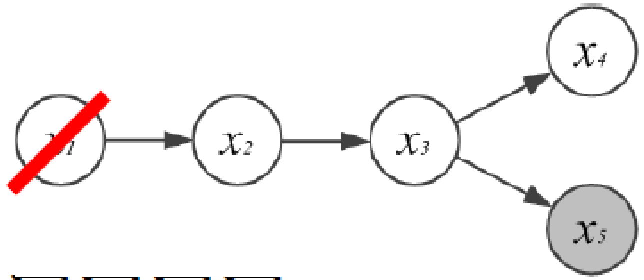
$$= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

学习与推断

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5)=\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

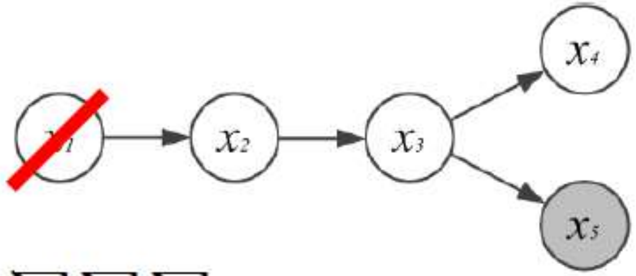
$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1}P(x_1)P(x_2|x_1)$$

$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

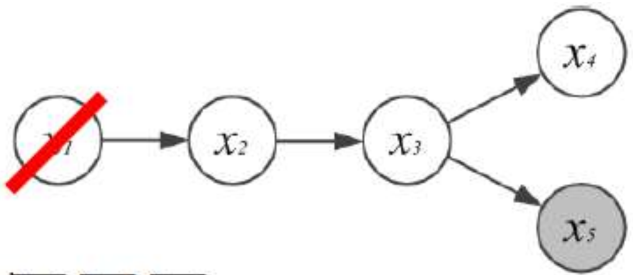


$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$
$$= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)\sum_{x_2} P(x_3|x_2)m_{12}(x_2)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

$$= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)\sum_{x_2} P(x_3|x_2)m_{12}(x_2)$$

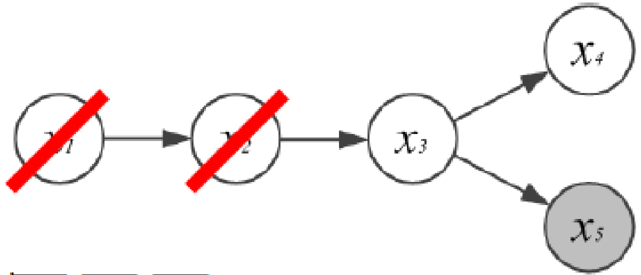
$$= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$

学习与推断

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

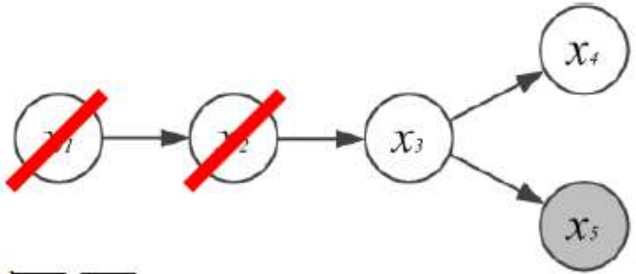
$$= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)\sum_{x_2} P(x_3|x_2)m_{12}(x_2)$$

$$= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

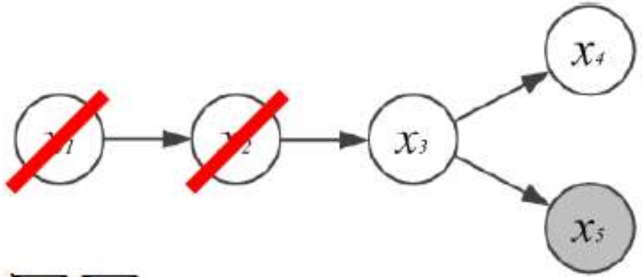


$$P(x_5) = \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$
$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

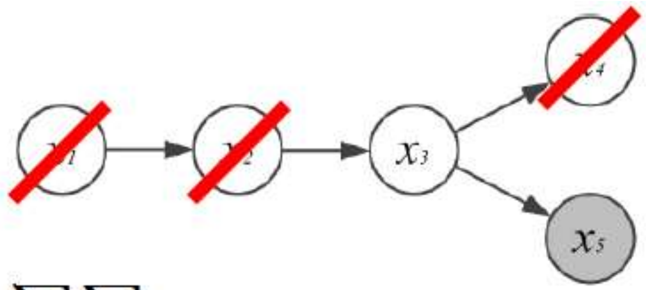


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$

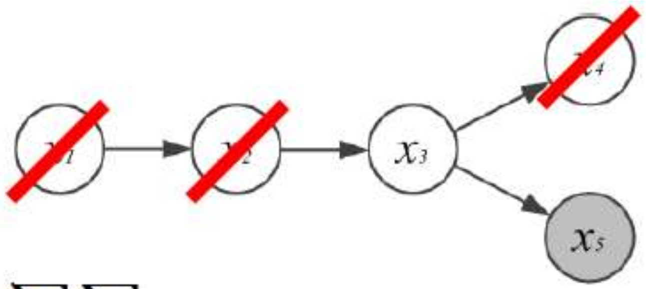
$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3)$$

$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

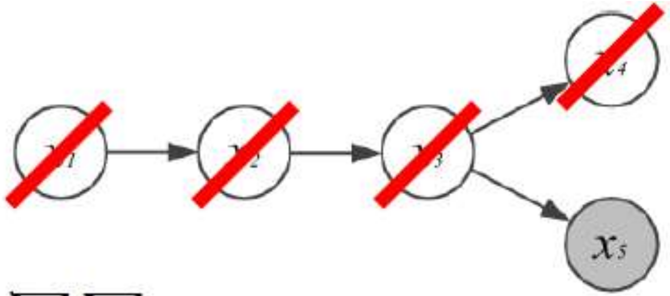


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3) = m_{35}(x_5)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$

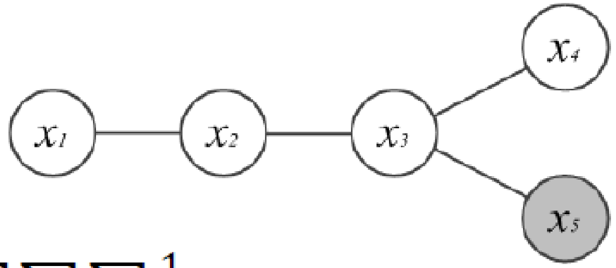
$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3)$$

$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3) = m_{35}(x_5)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

无向图



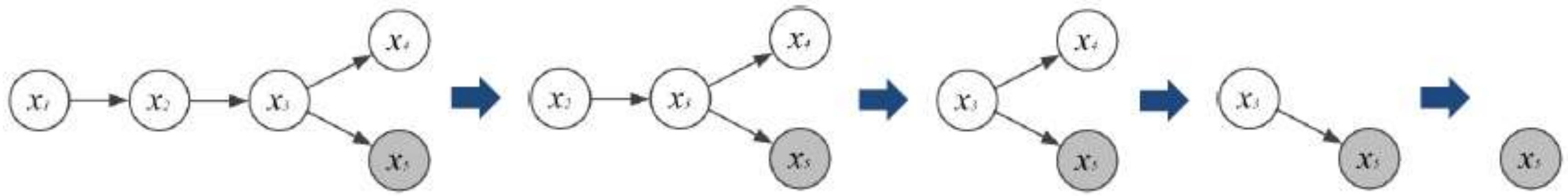
$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}\frac{1}{Z}\psi_{12}(x_1,x_2)\,\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5)$$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2}\frac{1}{Z}\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5)\sum_{x_1}\psi_{12}(x_1,x_2)$$

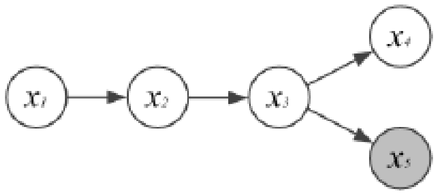
$$= \sum_{x_4}\sum_{x_3}\sum_{x_2}\frac{1}{Z}\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5)m_{12}(x_2) = \cdots$$

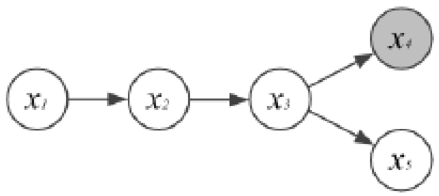
变量消去

优点：简单直观，代数上的消去对应图中结点的消去。



缺点：针对不同证据变量会造成大量冗余计算。



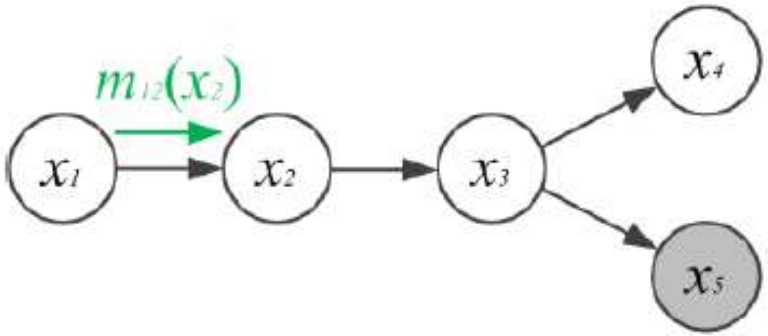


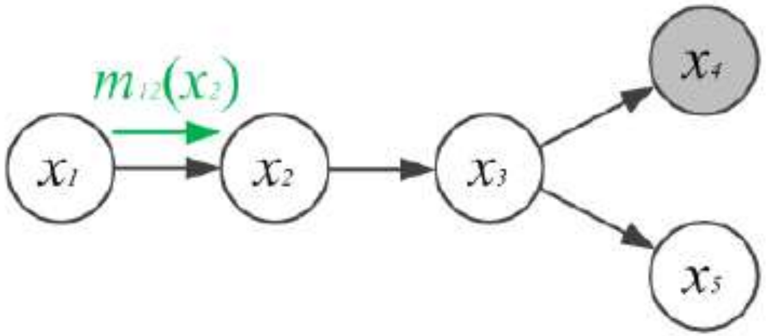
$$P(x_5)=\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$
$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1}P(x_1)P(x_2|x_1)$$
$$=\cdots$$

$$P(x_4)=\sum_{x_5}\sum_{x_3}\sum_{x_2}\sum_{x_1}P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$
$$=\sum_{x_5}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1}P(x_1)P(x_2|x_1)$$
$$=\cdots$$

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



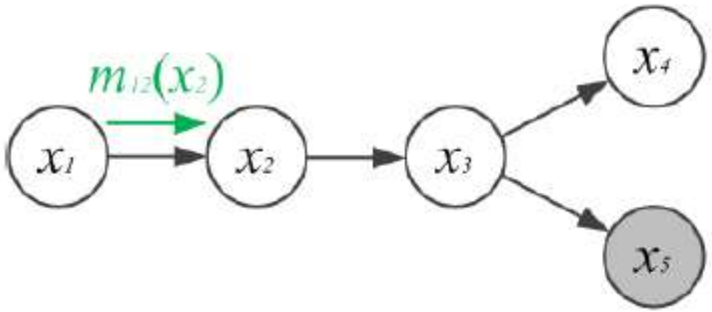


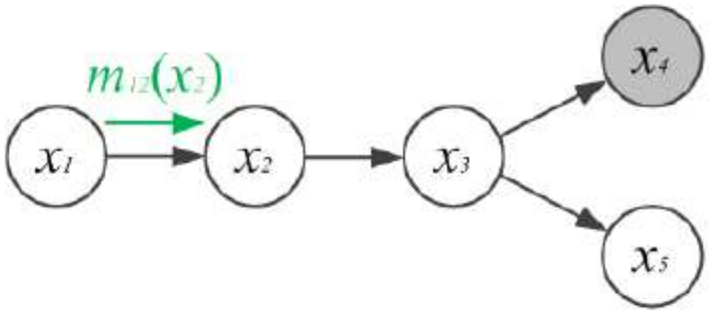
$$\begin{aligned} P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1) \\ &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\ &= \cdots \end{aligned}$$

$$\begin{aligned} P(x_4) &= \sum_{x_5}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1) \\ &= \sum_{x_5}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\ &= \cdots \end{aligned}$$

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。





$m_{12}(x_2)$是 x_1 从向 x_2 传递的一条消息。

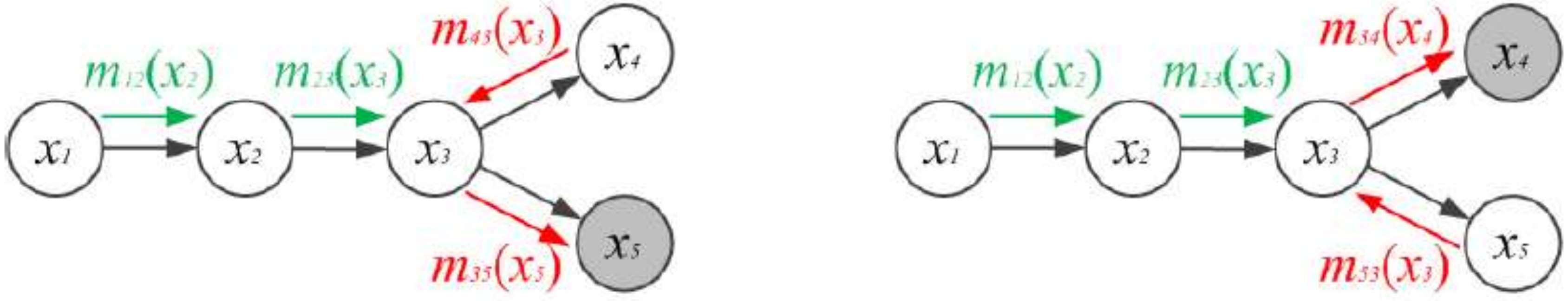
$m_{12}(x_2)$对 x_1 进行了求和，是关于 x_2 的函数。

$m_{12}(x_2)$仅与图的拓扑结构有关，与证据变量的选取无关（可复用）。

学习与推断

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。

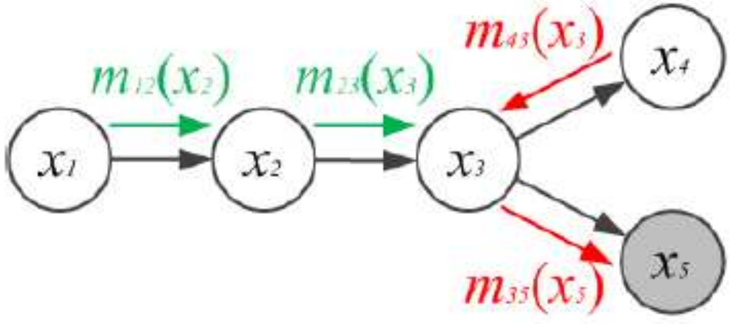


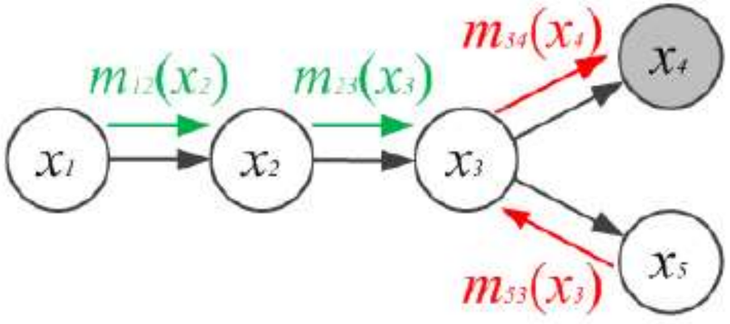
$m_{12}(x_2)$，$m_{23}(x_3)$可重复使用。

$m_{43}(x_3)$，$m_{35}(x_5)$，$m_{34}(x_4)$，$m_{53}(x_3)$需重新计算。

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



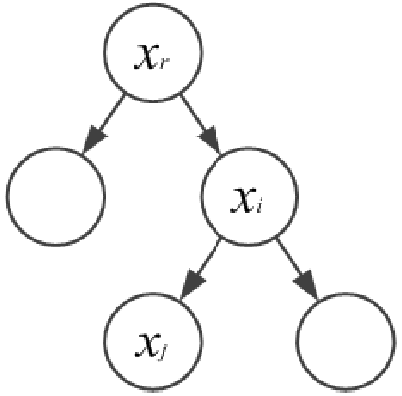


消息传递仅在邻接变量之间发生。

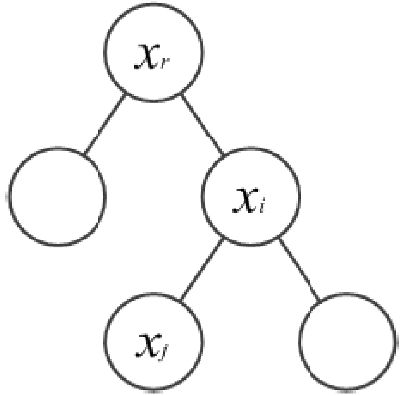
消息传递与边的方向性无关。

信念传播

树结构：有向树=无向树



$$P(\mathbf{x}) = P(x_r) \prod_{(i,j)\in E} P(x_j|x_i)$$

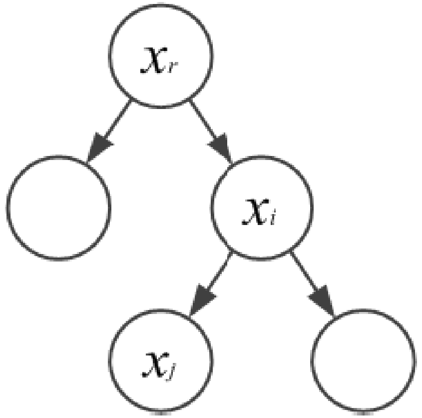


$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i\in V} \psi(x_i) \prod_{(i,j)\in E} \psi(x_i, x_j)$$

学习与推断

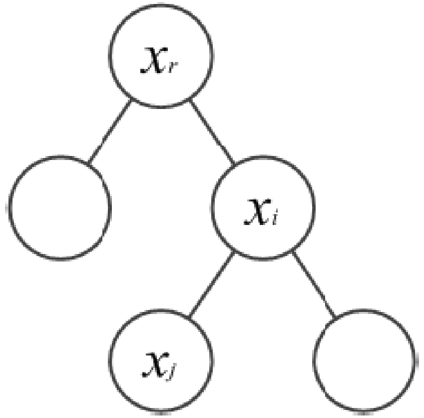
信念传播

树结构：有向树=无向树



$$P(\mathbf{x}) = P(x_r)\prod_{(i,j)\in E} P(x_j|x_i)$$

=

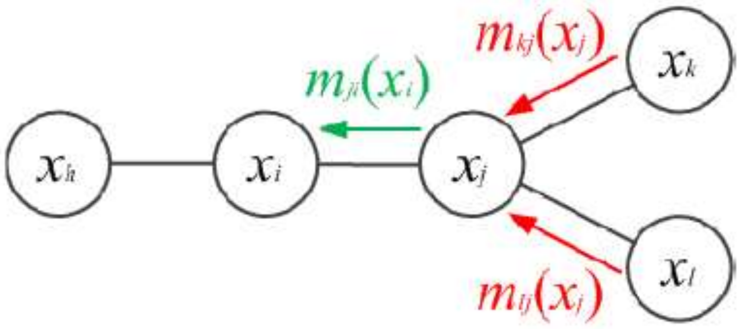


$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i\in V}\psi(x_i)\prod_{(i,j)\in E}\psi(x_i,x_j)$$

$$\boxed{Z = 1;\ \psi(x_r) = P(x_r);\ \psi(x_{\neg r}) = 1;\ \psi(x_i,x_j) = P(x_j|x_i)}$$

信念传播

树结构上的消息传递：



消息计算公式： （N(j)表示 j 的邻域节点）

$$m_{ji}(x_i) = \sum_{x_j} \psi(x_i, x_j) \prod_{k \in N(j) \setminus i} m_{kj}(x_j)$$

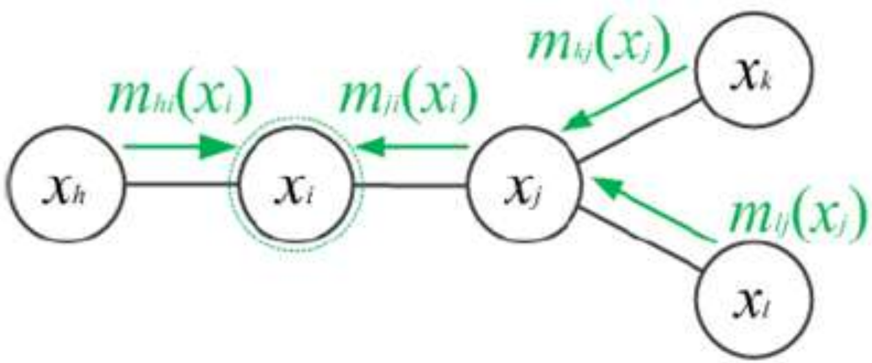
学习与推断

信念传播

边际分布：

$$P(x_i) \propto \prod_{k \in N(i)} m_{ki}(x_i)$$

结点k对结点i的信念

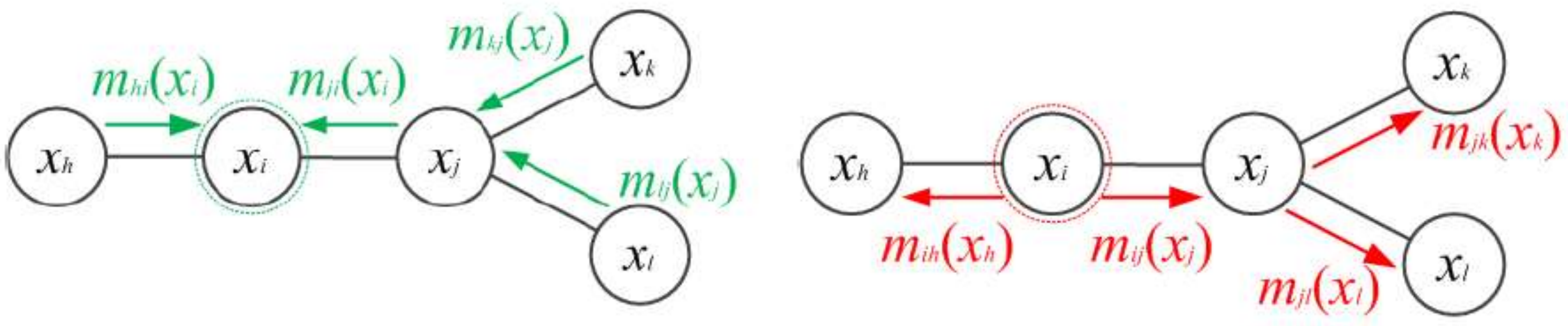


信念传播

二次扫描算法：

（1） 指定一个根结点，从所有叶结点开始向根节点传递消息，直到根结点收到所有邻接结点的消息。

（2） 从根结点开始向叶结点传递消息，直到所有叶结点均收到消息。



本讲参考文献

1. 《统计机器学习--第九章：概率图模型》课件，王泉，国科大网络安全学院，2017。

感谢王泉提供了《概率图模型》课件供本章教学参考！