马尔可夫随机场（MRF）

作者：Calvin
QQ: 179209347
Mail: 179209347@qq.com

介绍

笔记简介：

- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：

https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

*** 版权声明：**

- 仅限用于个人学习
- 禁止用于任何商业用途

条件随机场

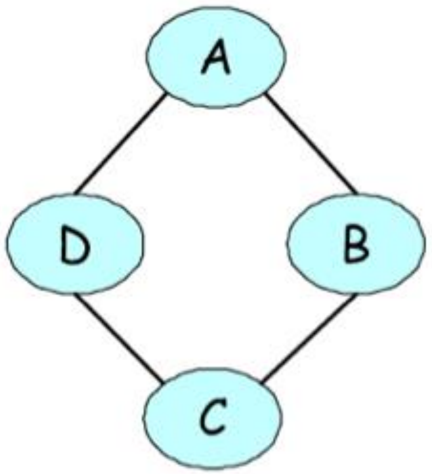
举一个有启发性的例子，假设我们正在对 A,B,C,D 个人的投票偏好进行建模。假设 (A,B) 、(B,C) 、(C,D) 和 (D,A) 是朋友，朋友往往有相似的投票偏好。这些影响可以自然地用无向图表示。

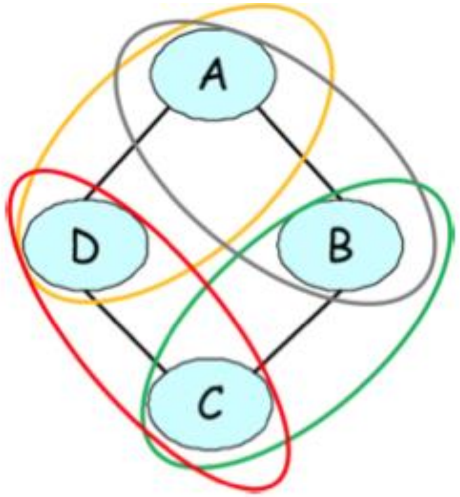
定义联合投票决策 A,B,C,D 的概率的一种方法是为这些变量的每个分配分数，然后将概率定义为归一化分数。分数可以是任何函数，但在我们的例子中，我们将其定义为以下形式：

$$\tilde{p}(A,B,C,D) = \phi(A,B)\phi(B,C)\phi(C,D)\phi(D,A)$$

其中 $\phi(X,Y)$ 是为朋友 X,Y 之间的一致投票分配更多权重的因子，例如：

$$\phi(X,Y) = \begin{cases} 10 & \text{if } X = Y = 1 \\ 5 & \text{if } X = Y = 0 \\ 1 & \text{otherwise.} \end{cases}$$





最后的概率定义为：

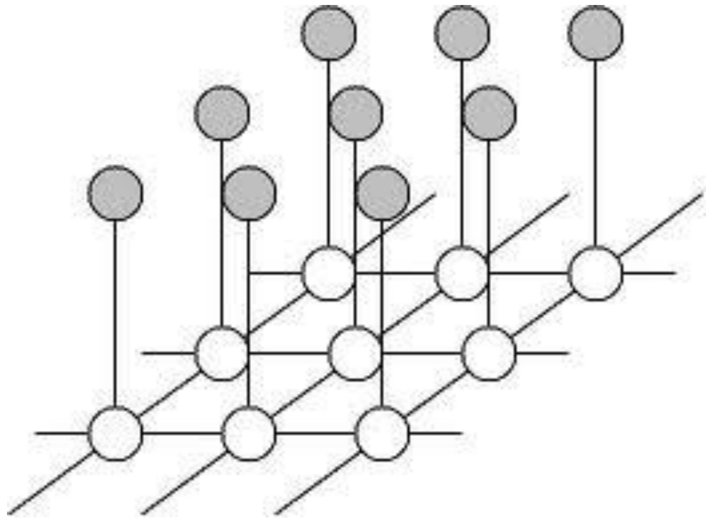
$$p(A,B,C,D) = \frac{1}{Z}\tilde{p}(A,B,C,D)$$

其中 $Z = \sum_{A,B,C,D} \tilde{p}(A,B,C,D)$ 是确保分布总和为1的归一化常数。

条件随机场

条件随机场(Conditional Random Fields，CRF)为给定一组输入随机变量条件下另外 一组输出随机变量的条件概率分布模型。可以表示为:

$$p_\Phi(Y \mid X) = \frac{1}{Z_\Phi} \tilde{p}_\Phi(X, Y)$$



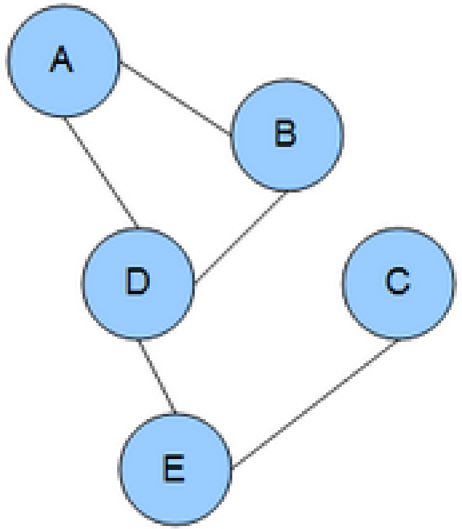
其中 X 为已知输入随机变量，Y 为输出随机变量。

$$\tilde{p}_\Phi(X, Y) = \prod_{i=1} \phi_i(D_i)$$

$$Z_\Phi = \sum_Y \tilde{p}_\Phi(X, Y)$$

马尔可夫随机场

马尔可夫随机场（通常缩写为MRF）、马尔可夫网络或无向图模型是一组具有由无向图描述的马尔可夫性质的随机变量。
马尔可夫随机场在表示依赖关系方面与贝叶斯网络相似；
不同之处在于贝叶斯网络是有向且无环的，而马尔可夫网络是无向的并且可能是有环的。

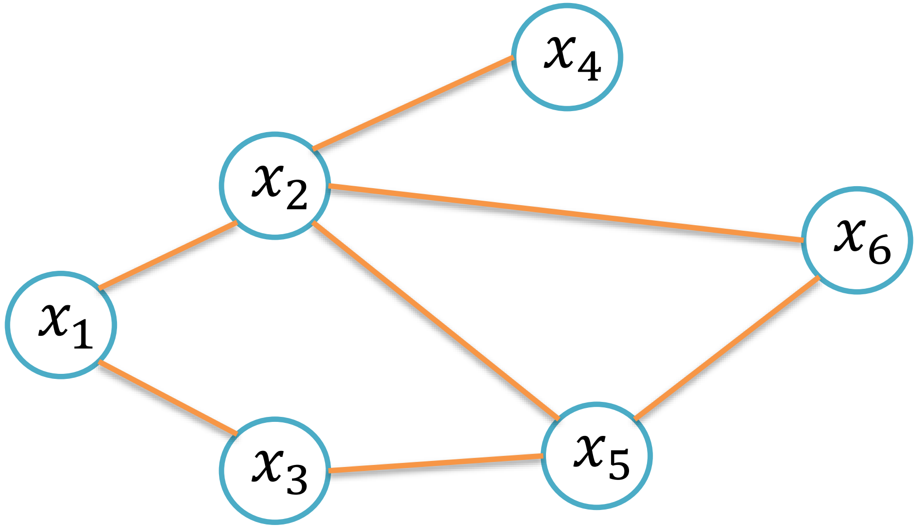


若这五个位置中的值具有马尔科夫性质，则ABCDE为一个马尔可夫随机场。每条边代表依赖性。
在这个例子中：
A依赖于B和D。B依赖于A和D。D取决于A、B和E。E取决于D和C。C取决于E。

团（Clique）

团（Clique）：一个全连通子图，即团内的所有节点之间都连边。
若在一个图中加入任何一个结点均不再构成团，则称该团为"极大团"。

对于图中的一个简单的马尔可夫随机场：
图中结点的一个子集，若其中任意两点都有边连接，则称该结点子集为“团”。



图中，团有{ x_1，x_2}、{ x_1，x_3}、{ x_2 x_4}、{ x_2，x_5}、{ x_2，x_6}、{ x_3，x_5}、{ x_5，x_6} 和 { x_2，x_5，x_6}，其中除了{ x_2，x_5}、{ x_2，x_6}、{ x_5，x_6} 之外都是极大团。

马尔可夫随机场

在马尔可夫随机场中，多个变量之间的联合概率分布能基于团分解成多个因子的乘积，每个因子仅与一个团相关。具体来说，对于n个变量 x={ x_1，x_2 ,,, x_n}，联合概率分布 P(x) 定义为：

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{Q\in\mathcal{C}} \psi_Q(\mathbf{x}_Q)$$

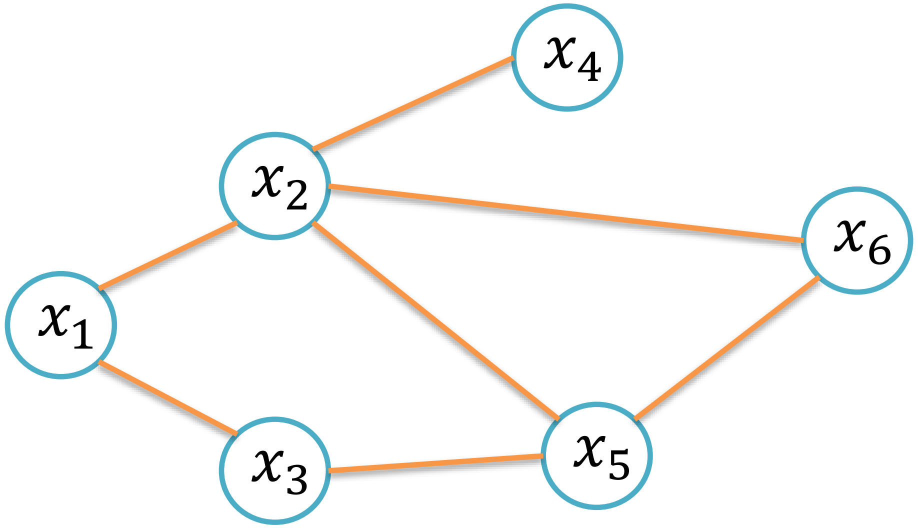
$P(\mathbf{x})$ 的含义就是所有团的势函数的乘积

- C：所有团构成的集合
- Q：Q 也是一个集合，表示一个团，是 C 集合的元素
- $\mathbf{x}_Q$：集合 Q 的所有元素结点。
- $\psi_Q(\mathbf{x}_Q)$：团 Q 的势函数
- $\boldsymbol{Z} = \sum_{\mathbf{x}} \Pi_{Q\in\mathcal{C}}\psi_Q(\mathbf{x}_Q)$：规范化因子，含义就是针对每一个结点 x，所有包含该结点的团的势函数的乘积，然后对图中所有的结点 x 进行求和。以确保 P(x) 是被正确定义的概率。

马尔可夫随机场

若Q不是极大团，则它必被一个极大团 Q* 所包含。于是联合概率 P(x) 可基于极大团来定义。假定所有极大团构成的集合为 C*，则有：

$$P(\mathbf{x}) = \frac{1}{Z^*} \prod_{Q \in \mathcal{C}^*} \psi_Q(\mathbf{x}_Q)$$



图中，极大团有 { x_1， x_2}、{ x_1， x_3}、{ x_2 x_4}、 { x_3， x_5}、和 { x_2， x_5， x_6}。

由于极大团 { x_2， x_5， x_6}，的存在，使得我们不再需要团 { x_2， x_5}、{ x_2， x_6}、{ x_5， x_6} 。

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

马尔科夫随机场

马尔科夫随机场(Markov Random Fields，MRF)联合概率分布可以表示以下分解形式:

$$p(X_1,\ldots,X_n)=\frac{1}{Z_\Phi}\prod_{i=1}\phi_i(D_i)$$

Z_Φ为联合概率分布的归一化因子，通常称为配分函数(partition function)。
D_i 是随机变量的集合
因子$\phi_i(D_i)$ 是从随机变量集合到实数域的一个映射，称为势函数。

$$\Phi=\{\phi_i(D_i),\ldots,\phi_n(D_n)\}$$

$$\tilde{p}(X_1,\ldots,X_n)=\prod_{i=1}\phi_i(D_i)$$

$$Z_\Phi=\sum_{X_1,\ldots,X_n}\tilde{p}(X_1,\ldots,X_n)$$

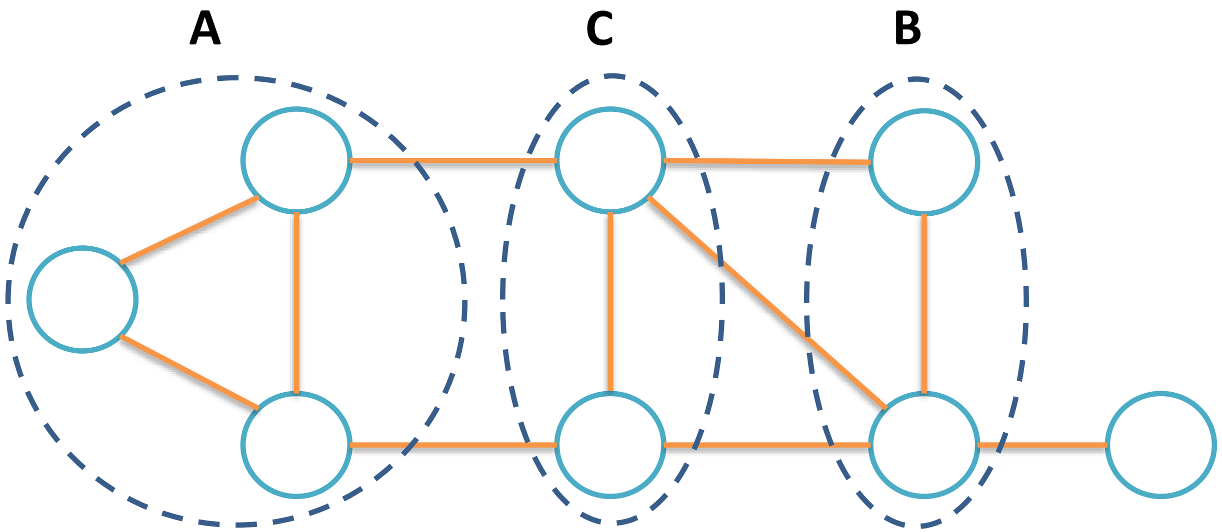
马尔可夫随机场的条件独立

马尔可夫随机场有三个性质，分别是：

- 全局马尔可夫性
- 局部马尔可夫性
- 成对马尔可夫性

后两个性质是第一个性质的推论。

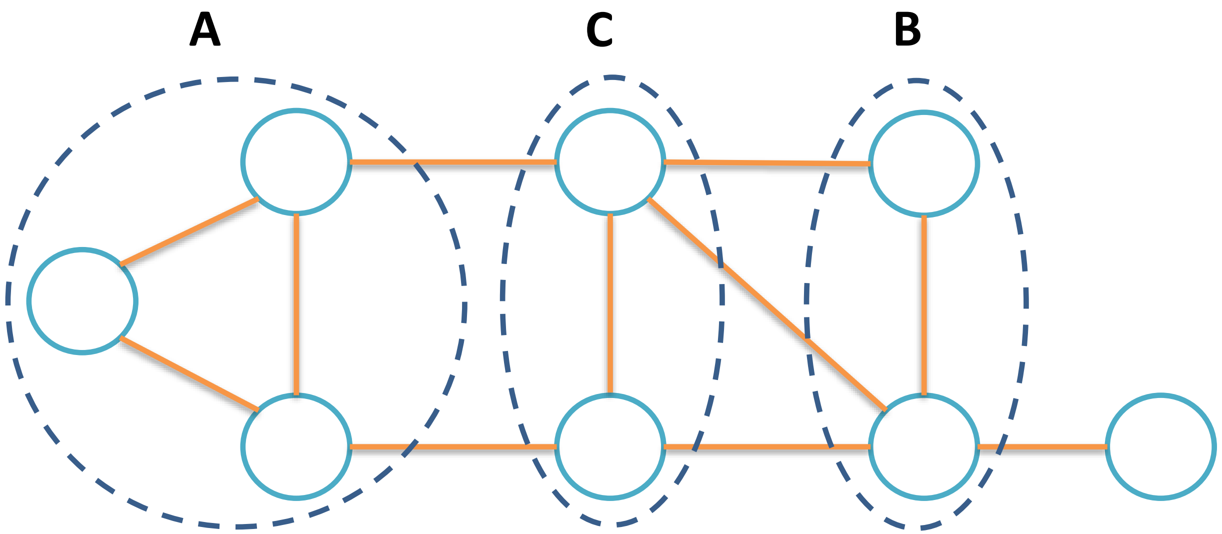
在马尔可夫随机场中，得到“条件独立”可以借助“分离”的概念。在下图所示中，若从结点集A中的结点到B中的结点都必须经过结点集 C 中的结点，则称结点集 A 和 B 被结点集 C 分离，C 称为分离集。



马尔可夫随机场的条件独立 - 全局马尔可夫性

全局马尔可夫性：指给定两个变量子集的分离集，则这两个变量子集条件独立。以图为例， x_A 、 x_B 和 x_C ，则 x_A 和 x_B 在给定 x_C 的条件下独立。

$$P(x_A, x_B \mid x_C) = P(x_A \mid x_C)P(x_B \mid x_C)$$



联合概率：

$$P(x_A, x_B, x_C) = \frac{1}{Z}\psi_{AC}(x_A, x_C)\psi_{BC}(x_B, x_C)$$

马尔可夫随机场的条件独立 - 局部马尔可夫性，成对马尔可夫性

- 局部马尔可夫性

由全局马尔可夫性可以得到推论：局部马尔可夫性，给定某个变量的相邻变量，则该变量条件独立于其他变量。形式化说，令 V 为图的结点集，n(v) 为结点 v 在图上的相邻结点，n*(v)=n(v)U{v}，有：

$$\mathbf{x}_v \perp \mathbf{x}_{V\setminus n^*(v)} \mid \mathbf{x}_{n(v)}$$

即结点与不相邻的结点独立

- 成对马尔可夫性

由以上可以得到另一个推论：成对马尔可夫性，给定所有其他变量，两个非邻接变量条件独立。形式话说，令图的结点集和边集分别为 V 和 E，对图中两个结点 u 和 v，若<u,v> 不属于 E，则有

$$\mathbf{x}_u \perp \mathbf{x}_v \mid \mathbf{x}_{V\setminus\langle u,v\rangle}$$

即两个结点之间无边，则两个结点独立

马尔可夫随机场 - 势函数

势函数的作用是刻画变量集xQ中变量之间的相关关系，它应该是非负数，且在所偏好的变量取值上有较大函数值。假定在图中的变量均为二值变量，若势函数为：

$$\psi_{AC}(x_A, x_C) = \begin{cases} 1.5, & \text{if } x_A = x_C; \\ 0.1, & \text{otherwise,} \end{cases}$$

$$\psi_{BC}(x_B, x_C) = \begin{cases} 0.2, & \text{if } x_B = x_C; \\ 1.3, & \text{otherwise,} \end{cases}$$

则说明该模型偏好变量 x_A 与 x_C 拥有相同的取值， x_B 与 x_C 拥有不同的取值。结合式子可知，令 x_A 与 x_C相同且 x_B 与 x_C 不同可以得到一个较高的联合概率。
为了满足非负性，势函数的定义往往使用指数函数，即：

$$\psi_Q(\mathbf{x}_Q) = e^{-H_Q(\mathbf{x}_Q)}$$

$H_Q(\mathbf{x}_Q)$ 是一个定义在变量 $\mathbf{x}_Q$ 上的实值函数，常见形式为：

$$H_Q(\mathbf{x}_Q) = \sum_{u,v\in Q,u\neq v} \alpha_{uv} x_u x_v + \sum_{v\in Q} \beta_v x_v$$

其中α和β为参数，第一项表示两个结点直接的关系而第二项仅考虑单结点。 α和β理解为两个部分的权重。


Thank
You

Hammersley-Clifford定理

- 无向图的联合概率可以分解为一系列定义在最大团上的非负函数的乘积形式。

定理 11.1 – Hammersley-Clifford 定理： 如果一个分布 $p(\boldsymbol{x}) > 0$ 满足无向图 G 中的局部马尔可夫性质，当且仅当 $p(\boldsymbol{x})$ 可以表示为一系列定义在最大团上的非负函数的乘积形式，即

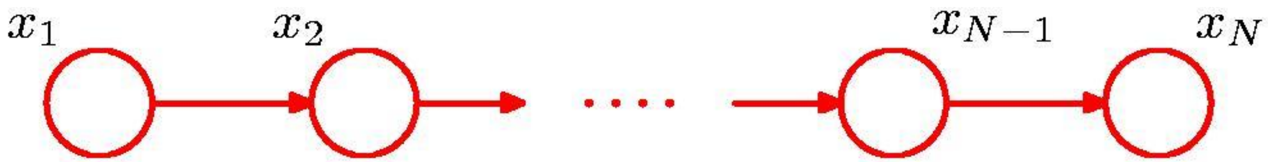
$$p(\boldsymbol{x}) = \frac{1}{Z}\prod_{c\in\mathcal{C}}\phi_c(\boldsymbol{x}_c), \tag{11.16}$$

其中 $\mathcal{C}$ 为 G 中的最大团集合，$\phi_c(\boldsymbol{x}_c) \geq 0$ 是定义在团 c 上的势能函数（Potential Function），Z 是配分函数（Partition Function），用来将乘积归一化为概率形式：

$$Z = \sum_{\boldsymbol{x}\in\mathcal{X}}\prod_{c\in\mathcal{C}}\phi_c(\boldsymbol{x}_c), \tag{11.17}$$

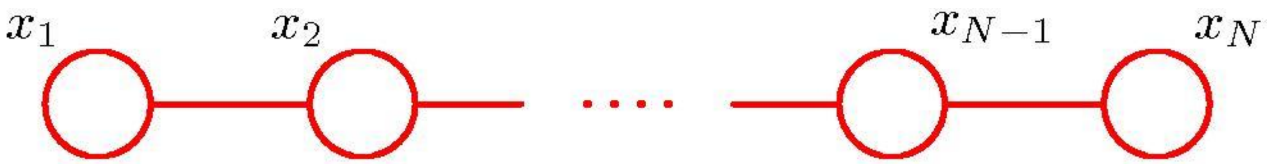
其中 $\mathcal{X}$ 为随机向量 X 的取值空间.

有向图和无向图的转换

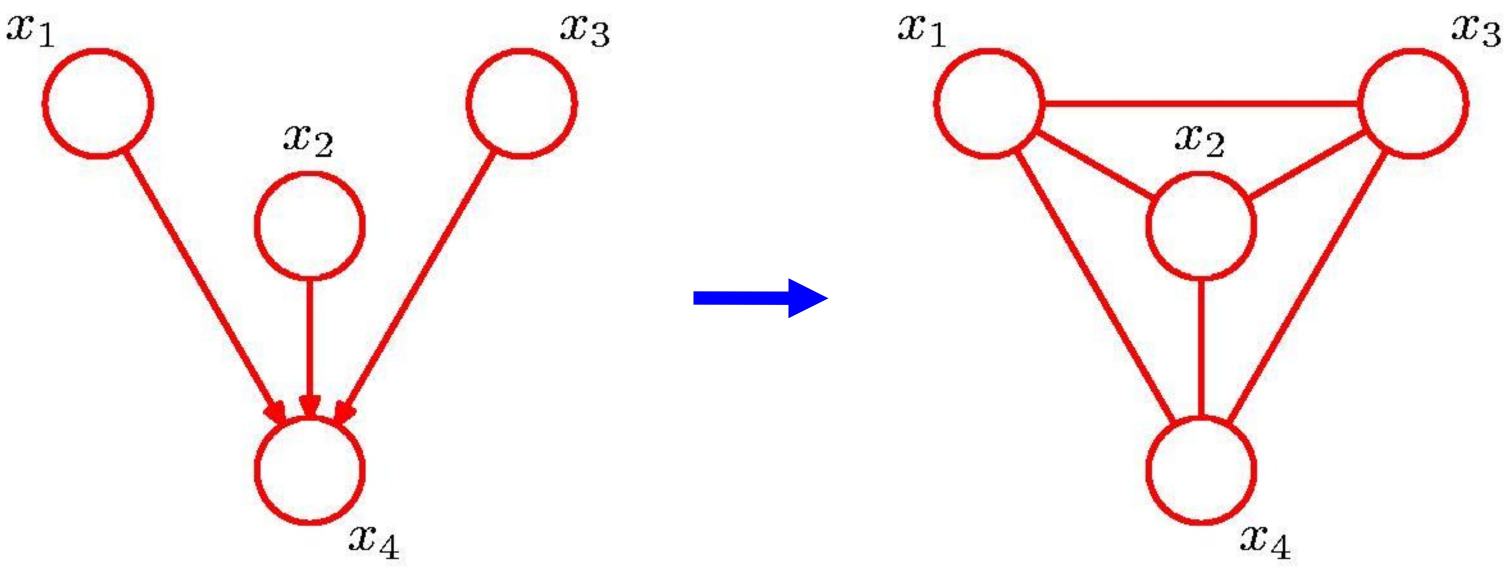


$$p(\mathbf{x}) = \underbrace{p(x_1)p(x_2|x_1)}\, p(x_3|x_2) \cdots p(x_N|x_{N-1})$$

$$p(\mathbf{x}) = \frac{1}{Z}\, \psi_{1,2}(x_1, x_2)\, \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$



有向图和无向图的转换



$$
\begin{aligned}
p(\mathbf{x}) &= p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3) \\
&= \frac{1}{Z}\psi_A(x_1,x_2,x_3)\psi_B(x_2,x_3,x_4)\psi_C(x_1,x_2,x_4)
\end{aligned}
$$