DATE

概率图1—背景介绍

1. 高维随机变量计算量大 ⟶ 简化 $\xrightarrow{\text{相互独立}}$ $P(x_1,\cdots,x_p)=\prod_{j=1}^{p}P(x_j)$

Naive Bayes: $P(x|y)=\prod_{j=1}^{p}P(x_j|y)$

$\xrightarrow{\text{Markov Property}}$ $x_j \perp x_{i+1} \mid x_i,\ j<i$

HMM: 齐次Markov假设

$\xrightarrow{\text{条件独立性}}$ $x_A \perp x_B \mid x_C$，x_A, x_B, x_C是集合且不相交（是马尔可夫性质的推广）

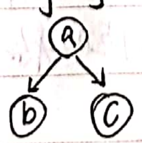
2. 概率图
- representation
  - 有向图 Bayesian Network
  - 无向图 Markov Network
- inference 推断
  - 精确推断
  - 近似推断
    - 确定性近似（变分推断）
    - 随机近似（MCMC）
- learning 学习
  - 结构学习
  - 参数学习
    - 完备数据
      - 有向图
      - 无向图
    - 隐变量: EM
- decision 决策

概率图2 — 贝叶斯网络

1. chain rule: $P(x_1,\cdots,x_p)=P(x_1)\prod_{i=2}^{p}P(x_i|x_{1:i-1})$
2. 因子分解 $P(x_{1:p})=\prod_{i=1}^{p}P(x_i|x_{pa(i)})$

$x_{pa(i)}$是x_i的父节点集合

3. 拓扑排序: $i<j \longrightarrow [\cdots,x_i,\cdots,x_j\cdots]$
4. 假定已构建了一个有向图模型:



$P(a,b,c)=\underline{P(a)P(b|a)P(c|a)}$ 因子分解

$P(a,b,c)=\underline{P(a)P(b|a)P(c|a,b)}$ 链式法则

所以 $P(c|a)=P(c|a,b) \Rightarrow$ ~~b,c独立~~ $b\perp c|a$

证明: $P(c|a)P(b|a)=P(c|a,b)P(b|a)$

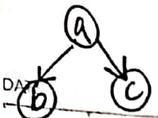
$=P(c,b|a)$

说明图与集合都能体现独立性.

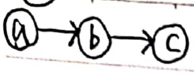
$c\perp b|a \Leftrightarrow$ a若被观测,则路径阻塞

(无向)路径: $b \leftarrow a \rightarrow c$ (tail to tail) tail: a是头b是尾

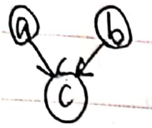
有向图三种情况



tail to tail　$b \perp c \mid a$



head to tail　$a \perp c \mid b$ $\Longleftrightarrow$ 若b被观测则路径被阻塞



head to head　默认情况下 $a \perp b$，路径堵塞

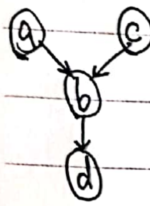
☆ 若c被观测，a和b不独立，路径是通的　类比：两个子结婚生娃C

证明：$P(a,b,c) = P(a)\,P(b)\,P(c \mid a,b)$

$= P(a)\,P(b \mid a)\,P(c \mid a,b)$

所以 $P(b) = P(b \mid a) \Rightarrow a \perp b$

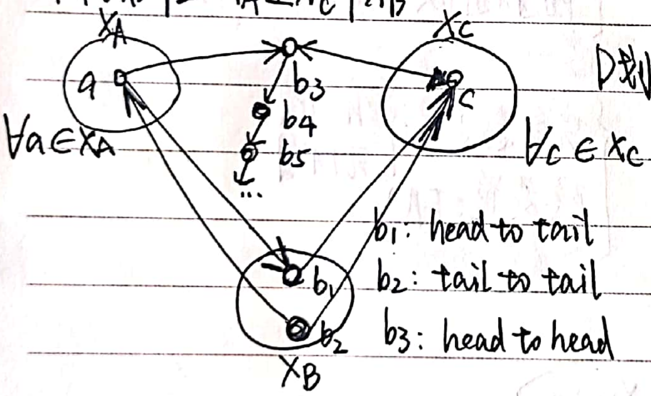
$\boxed{\text{概率图3——D划分}}$ D-Seperation



head to head的拓展

若d被观测，a—b—c也是连通的

条件独立性　$X_A \perp X_C \mid X_B$



D划分（全局马尔可夫性）

b_1, b_2 一定在 X_B 中

$b_3, b_4, b_5, \cdots$ 一定不在 X_B 中

b_1: head to tail

b_2: tail to tail

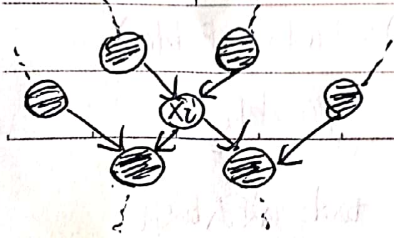
b_3: head to head

$$P(X_i \mid X_{-i}) = \frac{P(X_i, X_{-i})}{P(X_{-i})} = \frac{P(X)}{\underbrace{\int_{X_i} P(X)\,dX_i}_{\text{边缘分布转联合分布}}} = \frac{\prod_{j=1}^{p} P(X_j \mid X_{pa(j)})}{\int_{X_i} \prod_{j=1}^{p} P(X_j \mid X_{pa(j)})\,dX_i}$$

表示集合 $\{X_1 \ldots i-1, i+1, \cdots, p\}$

$= f(\triangle)$，用△表示与 X_i 相关的点的集合（未给出证明）

i.e. X_i 与其它所有节点的关系可以简化为 X_i 与"和 X_i 连通的点集"的关系。



$\underbrace{\text{和}X_i\text{连通的点集}}_{\text{相关}}$

Markov Blanket 马尔可夫毯

概率图4 — representation example

贝叶斯网络:

Naive Bayes: $P(X|y)=\prod_{i=1}^{p} p(x_i|y=1)$

GMM: $z \to x$

z is discrete

i.e. $z=1,2,\dots,K$

$X|Z \sim N(\mu,\Sigma)$

带时间: Markov Chain 随机过程中的一种

Gaussian Process 无限维高斯分布

连续: Gaussian Network / Gaussian Bayesian Network

动态系统

e.g. HMM

离散隐状态序列马氏链

e.g. LDS

卡尔曼滤波

线性,连续(高斯)

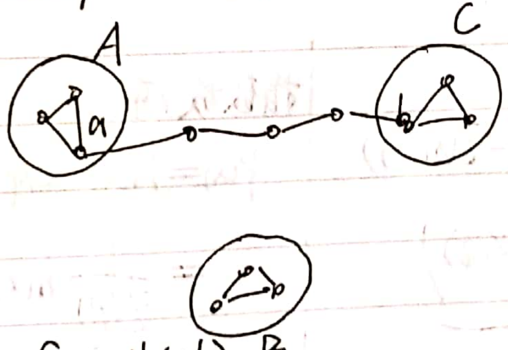
e.g. Particle Filter

非高斯,非线性

概率图5 — Markov Network

Markov无向图. Bayesian有向图

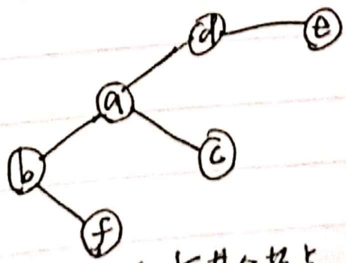
1. Global Markov



$\forall a \in A, \forall b \in C$. path(a,b)

$X_A \perp X_C | X_B$: 从a到c的路径上至少有一个节点$\in X_B$

2. local Markov



$a \perp \{e,f\} | \{b,c,d\}$

a在给定邻节点$\{b,c,d\}$之后,与其余节点$\{f,e\}$均独立.

$a \perp \{全集 - a - 邻居\} | 邻居$

DATE

3、成对 Markov

$X_i \perp X_j \mid X_{-ij} \quad (i \neq j)$

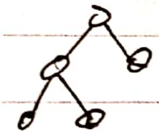
总结：Markov Network 的条件独立体现在三方面：①全局 ②局部 ③成对

有数学证明是等价的

Markov Network 的因子分解

有向图网络的因子分解：拓扑排序

团：一个关于结点的集合，结点之间相互连通；最大团 $C_i \ (i=1,2,\cdots,k)$



$$P(x) = \frac{1}{Z} \prod_{i=1}^{k} \phi(X_{C_i})$$

$$Z = \sum_{x_1, x_2, \cdots x_p} \prod_{i=1}^{k} \phi(X_{C_i}) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_p} \prod_{i=1}^{k} \phi(X_{C_i})$$

Z 归一化常数

ϕ：势函数，必须为正

Hammesley-Clifford 定理：MRF 与吉布斯分布等价

$\phi(X_{C_i}) = \exp(-E(X_{C_i}))$，取指数势能函数的 $P(x)$：Gibbs Distribution

Boltzmann Distribution

$$P(X) = \frac{1}{Z} \prod_{i=1}^{k} \phi(X_{C_i})$$

$$= \frac{1}{Z} \prod_{i=1}^{k} \exp(-E(X_{C_i}))$$

$$= \underbrace{\frac{1}{Z} \exp\left(-\sum_{i=1}^{k} E(X_{C_i})\right)}_{\text{指数族分布}}$$

指数族分布

$$p(x) = h(x) \cdot \exp\{\eta^T \phi(x) - A(\eta)\}$$

$$= \frac{1}{Z(\eta)} h(x) \cdot \exp\{\eta^T \phi(x)\}$$

最大熵原理 $\Rightarrow$ 指数族分布

Markov Random Field $\Longleftrightarrow$ Gibbs Distribution $\Longleftarrow$ 最大熵原理

3. 成对 Markov

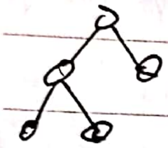
$x_i \perp x_j \mid x_{-i-j}$ $(i\neq j)$ [redacted].

总结: Markov Network的条件独立体现在三方面: ①全局 ②局部 ③成对
有数学证明是等价的

Markov Network的因子分解

有向图网络的因子分解: 拓扑排序

团: 一个关于结点的集合, 结点之间相互连通; 最大团 $C_i \ (i=1,2,\cdots,K)$



$$P(x) = \frac{1}{Z}\prod_{i=1}^{K}\phi(x_{C_i})$$

$$\underset{\text{归一化常数}}{Z} = \sum_{x_1,x_2,\cdots,x_p}\prod_{i=1}^{K}\phi(x_{C_i}) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_p}\prod_{i=1}^{K}\phi(x_{C_i})$$

ϕ: 势函数, 必须为正

Hammesley-Clifford定理: MRF与吉布斯分布等价

$\phi(x_{C_i}) = \exp(-E(x_{C_i}))$. 取指数势能函数的 $p(x)$: Gibbs Distribution
Boltzmann Distribution

$$\begin{aligned} P(x) &= \frac{1}{Z}\prod_{i=1}^{K}\phi(x_{C_i}) \\ &= \frac{1}{Z}\prod_{i=1}^{K}\text{[redacted]}\exp(-E(x_{C_i})) \\ &= \underbrace{\frac{1}{Z}\exp\left(-\sum_{i=1}^{K}E(x_{C_i})\right)}_{\text{指数族分布}} \end{aligned}$$

指数族分布

$$\begin{aligned} p(x) &= h(x)\cdot\exp\{\eta^T\phi(x) - A(\eta)\} \\ &= \frac{1}{Z(\eta)}h(x)\cdot\exp\{\eta^T\phi(x)\} \end{aligned}$$

最大熵原理 $\Rightarrow$ 指数族分布

Markov Random Field $\Longleftrightarrow$ Gibbs Distribution $\Longleftarrow$ 最大熵原理