概率图

模型

一些要体会条件独立性.

$(X_A \perp X_B \mid X_C)$

A、B、C为三个不相交的集合

关注对象

高维随机变量

$P(x_1,\cdots,x_p)$

概率图

- 边缘概率 $P(x_i)$
- 条件概率 $P(x_j|x_i)$

basic rules

Sum rule: $P(x_1)=\int P(x_1,x_2)\,dx_2$

Product rule: $P(x_1)\cdot P(x_2|x_1)=P(x_2)\cdot P(x_1|x_2)$

Chain rule: $P(x_1,x_2,\cdots,x_p)=\left[\prod_{i=2}^{p}P(x_i|x_{i-1},\cdots,x_1)\right]\cdot P(x_1)$

Bayesian Rule: $P(x_2|x_1)=\dfrac{P(x_1,x_2)}{P(x_1)}=\dfrac{P(x_1,x_2)}{\int P(x_1,x_2)\,dx_2}$

$=\dfrac{P(x_2)\cdot P(x_1|x_2)}{\int P(x_1,x_2)\,dx_2}$

困境：维度高，计算复杂。$P(x_1,x_2\cdots x_p)$计算量太大。

↓

简化

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

维度之间相互独立 → 假设太强

$P(x_1,\dots,x_p)=\prod_{i=1}^{p}P(x_i)$ → Markov property $x_j\perp x_{i+1}|x_i$, $j<i$

HMM

(齐次Markov假设)

(只与前一项有关)

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。)

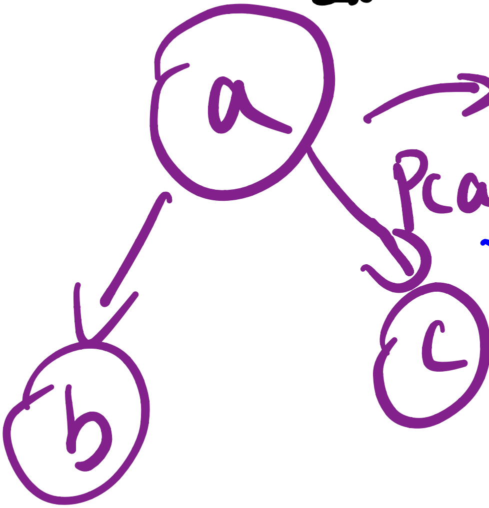
核心性质

$x_1\leftarrow\cdots\leftarrow x_p$ → 一阶Markov假设

依赖关系太过单调

马氏性质的推广

通过拓扑图 → 因子分解



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

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

chain rule

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

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

$\Rightarrow c\perp b \mid a$

因为给定a，不管b发没发生，C不受影响。

概率图

- Representation 表示
  - 有向图 Bayesian Network
  - 高斯图(连续)
    - Gaussian BN
    - Gaussian MN
  - 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
  - 参数学习
    - 完备数据
    - 隐变量
  - 结构学习

直接通过随机变量通过拓扑排序之后得到的图的分解式。

Bayesian Network → 有向图

$$P(x_1,\cdots,x_p)=P(x_1)\prod_{i=2}^{p}P(x_i|x_1,\cdots,x_{i-1}) \Rightarrow \text{chain rule}$$

条件独立性: $X_A \perp X_C \mid X_B$ (此处 X_A, X_B, X_C 均为 collection)

因子分解: $P(x_1,\cdots,x_p)=\prod_{i=1}^{p}P(x_i|x_{pa(i)})$ → 为 x_i 的父亲集合.

tail → head.

case 1: (head to tail) $a \to b \to c$

若b被观测,路径堵塞, a,c独立。 → $a \perp c \mid b$

case 2: (tail to tail) $a \leftarrow c \rightarrow b$

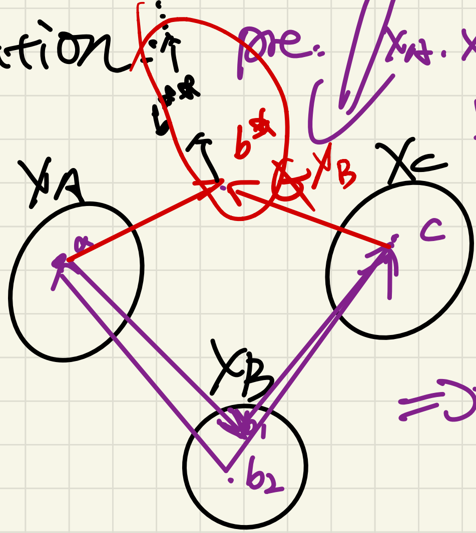
若c被观测,路径被堵塞, a-b独立。 $a \perp b \mid c$

case 3: (head to head) $a \rightarrow c \leftarrow b$

默认情况下: a,b独立, $a \perp b$ // c被观测后, a,b不独立。

D-seperation

pre: X_A, X_B, X_C 都为随机变量集合 且相互间不相交。



$\Rightarrow$ 全局 markov property

$$P(x_i \mid x_{-i}) = \frac{P(x_i, x_{-i})}{P(x_{-i})} = \frac{P(x)}{\int_{x_i} P(x)\, dx_i}$$

x_i与其他所有随机变量的关系。

边缘概率

因子分解

$$= \frac{\prod_{j=1}^{p} P(x_j \mid x_{p(x_j)})}{\int_{x_i} \prod_{j=1}^{p} P(x_j \mid x_{p(x_j)})\, dx_i}$$

$p(x_j)$: 父亲结点集合

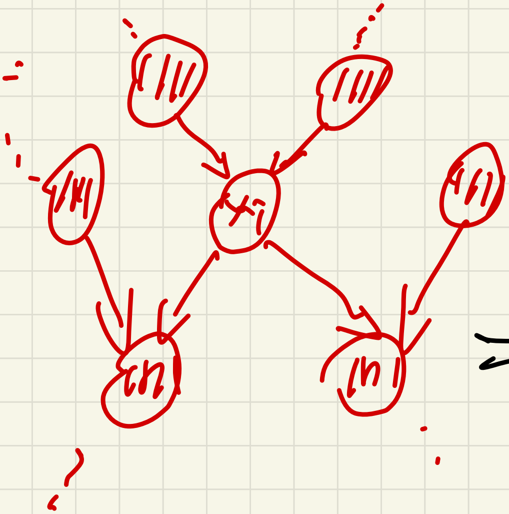
Δ: 与 x_i 有关

$\bar{\Delta}$: 与 x_i 无关

$$= \frac{f(\Delta) \cdot f(\bar{\Delta})}{f(\bar{\Delta}) \cdot \int_{x_i} f(\Delta)\, dx_i} = g(\Delta)$$

说明 x_i 与其他随机变量的关系 与 自己有关系的有关。

图示:

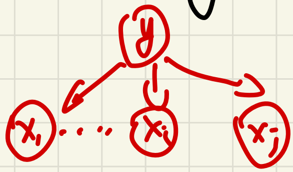


$$\begin{cases} P(x_i \mid x_{pa(i)}) \\ P(x_{child(i)} \mid x_i, x_{parent(child(i))}) \end{cases}$$

$\Rightarrow$ Markov Blanket -

Bayesian Network

单一: Naive Bayes → $P(x|y) = \prod_{i=1}^{p} P(x_i|y=1)$



混合: GMM (高斯混合模型)

时间: Markov chain
Gaussian process (无限维高斯分布)

连续: Gaussian Bayesian Network

从单一到混合

从有限到无限

空间 时间

动态模型

HMM (离散的)

LDS (Kalman Filter) 连续, 线性 → Gaussian

Particle Filter (非高斯, 非线性)

Background

Detailed models