模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 2 (Part 3): Bayesian Decision Theory

(Sections 2-9,2-11)

本讲内容

1. 离散特征的贝叶斯决策理论

2. 贝叶斯置信网

离散特征

- 在某些情况下，特征矢量不是在一个连续的 d 维实数空间中取值，其取值可以被理解为 m 个离散值 $v_1, v_2, \ldots, v_m$ 中的一个。
- 特征矢量 x 的每个分量的取值为离散值时，称为**离散特征**
  - 例如：颜色、纹理、类型等

- 在离散特征的条件下，**概率密度**代之以**概率分布**，概率也由积分形式变为求和形式：

$$\int p(x|\omega_j)\,dx \longrightarrow \sum_x p(x|\omega_j)$$

在离散特征下，在前面连续情况中得出的**贝叶斯规则**和**条件风险**的**基本结论不变**。

离散特征的贝叶斯决策理论

- 下面考虑特征矢量的分量相互独立，分量均取值为二值的两类问题的分类。即
  - 特征矢量为 $x = [x_1, x_2, \ldots, x_d]^t$，这里 x_i 取值为 0 或 1, 其概率分布为:

$$p_i = P(x_i = 1 \mid \omega_1)$$

$$q_i = P(x_i = 1 \mid \omega_2)$$

分量为二值离散的特征

只考虑两类的情况: ω_1 和 ω_2.

并且其特征取值为二值:

$x = (x_1, x_2, \cdots, x_d)^T$ 对每个 x_i 取值为 1 或 0.

对每一个特征分量，定义两个概率:

$$p_i = P(x_i = 1|\omega_1) \qquad q_i = P(x_i = 1|\omega_2)$$

在贝叶斯公式中，后验概率的计算通常非常复杂：

$$P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)}$$

分量独立的二值特征

- 假设二值特征分量是相互独立的，则概率就比较容易求得。在这种情况下，条件概率可以定义为下面简单的积形式:

$$P(x|\omega_1)=\prod_{i=1}^{d} p_i^{x_i}(1-p_i)^{1-x_i} \qquad P(x|\omega_2)=\prod_{i=1}^{d} q_i^{x_i}(1-q_i)^{1-x_i}$$

其中　$x_i \in \{1, 0\}$

而似然率为：

$$\frac{P(x|\omega_1)}{P(x|\omega_2)}=\prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}$$

分量独立的二值特征的 Bayes 决策

回顾前面的贝叶斯规则。若下式成立时，选择 ω_1

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \quad 或 \quad \ln\left[\frac{P(x|\omega_1)}{P(x|\omega_2)}\right] - \ln\left[\frac{P(\omega_2)}{P(\omega_1)}\right] > 0$$

此时判别函数为：

$$g(x) = \sum_{i=1}^{d}\left[x_i \ln\frac{p_i}{q_i} + (1-x_i)\ln\left(\frac{1-p_i}{1-q_i}\right)\right] + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

也可以表示为

$$g(x) = \sum_{i=1}^{d}\omega_i x_i + \omega_0$$

(特征分量的线性组合)

$$\omega_i = \ln\frac{p_i(1-q_i)}{q_i(1-p_i)} \qquad \omega_0 = \sum_{i=1}^{d}\ln\left(\frac{1-p_i}{1-q_i}\right) + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

本讲内容

1. 离散特征的贝叶斯决策理论

2. 贝叶斯置信网

贝叶斯置信网：变量统计独立

- 两个随机变量统计独立的含义如下：
  - 考虑一个多维分布 $p(x)$，若对其两个特征，有

$$p\left(x_i, x_j\right) = p\left(x_i\right) p\left(x_j\right)$$

则称这两个特征是统计独立的。

- 若已知了那一些特征是独立的，那一些不是独立的，则其联合概率密度的计算就比较简单

贝叶斯置信网

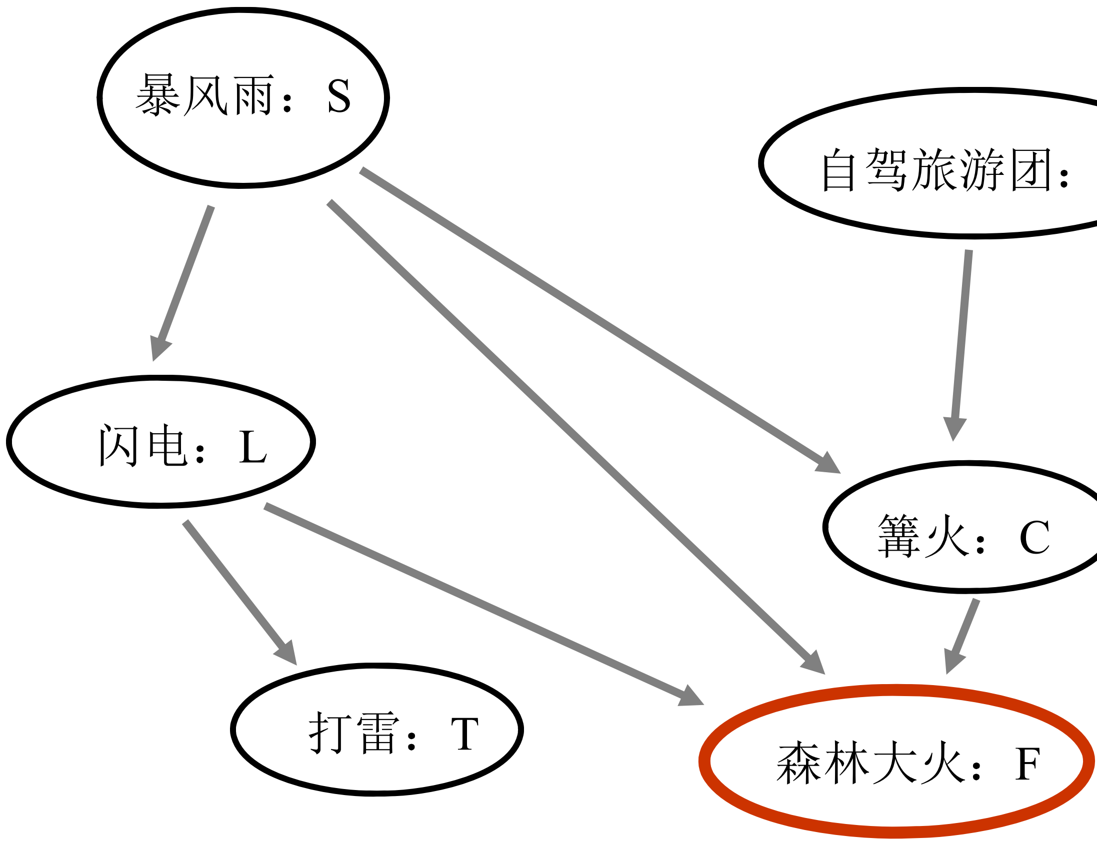
- 贝叶斯置信网（Bayesian Belief Network，BBN）是描述变量集合的联合概率分布的一种方法。
  - BBN也称为因果网（Casual Network，CN）
  - 或信任网（Belief Net，BN）。

令 $x = x_1, x_2, \ldots, x_n$ 是一个特征变量的集合

则利用贝叶斯置信网 BBN 可以计算 $x = x_1, x_2, \ldots, x_n$ 中的任意分量组合的联合概率

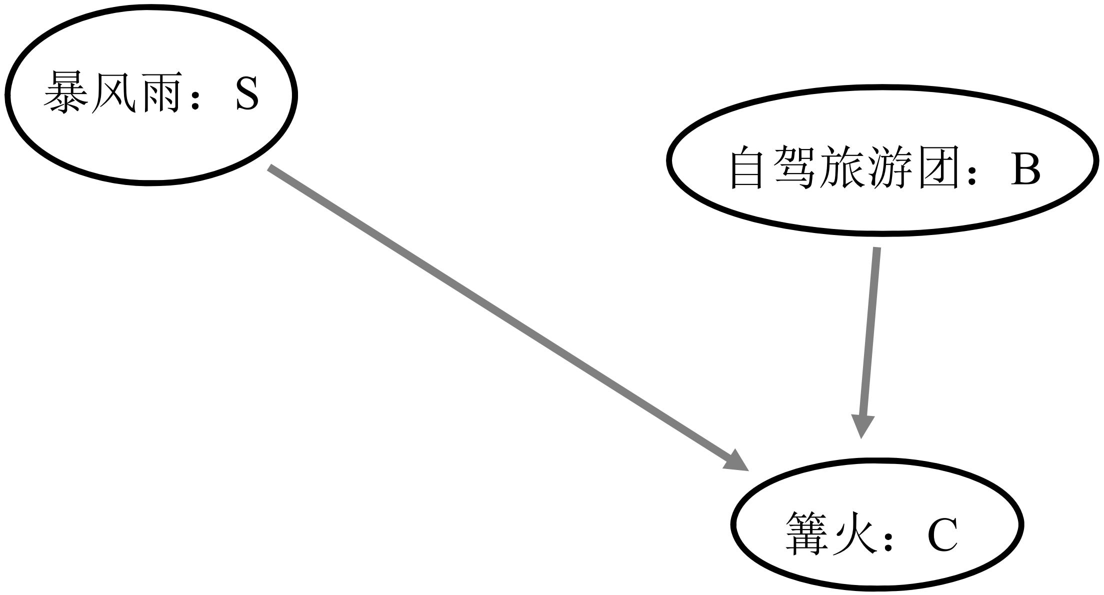
贝叶斯置信网：一个例子

- 二值变量和它们的关系集合



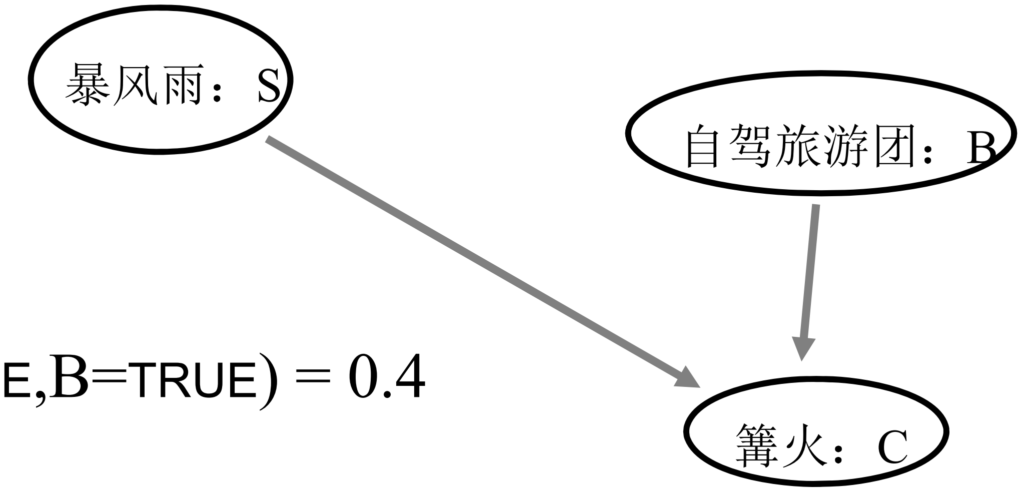
贝叶斯置信网：条件概率

	S,B	S,~B	~S,B	~S,~B
C	0.4	0.1	0.8	0.2
~C	0.6	0.9	0.2	0.8



条件概率：一个例子

	S,B	S,~B	~S,B	~S,~B
C	0.4	0.1	0.8	0.2
~C	0.6	0.9	0.2	0.8



$$P(C=\text{TRUE}|S=\text{TRUE},B=\text{TRUE}) = 0.4$$

贝叶斯置信网：条件独立

- 若 x_1 的概率独立于给定 x_3 时 x_2 的概率，则称 x_1 条件独立于给定 x_3 时的 x_2:

$$P\left(x_1 \middle| x_2, x_3\right) = P\left(x_1 \middle| x_3\right)$$

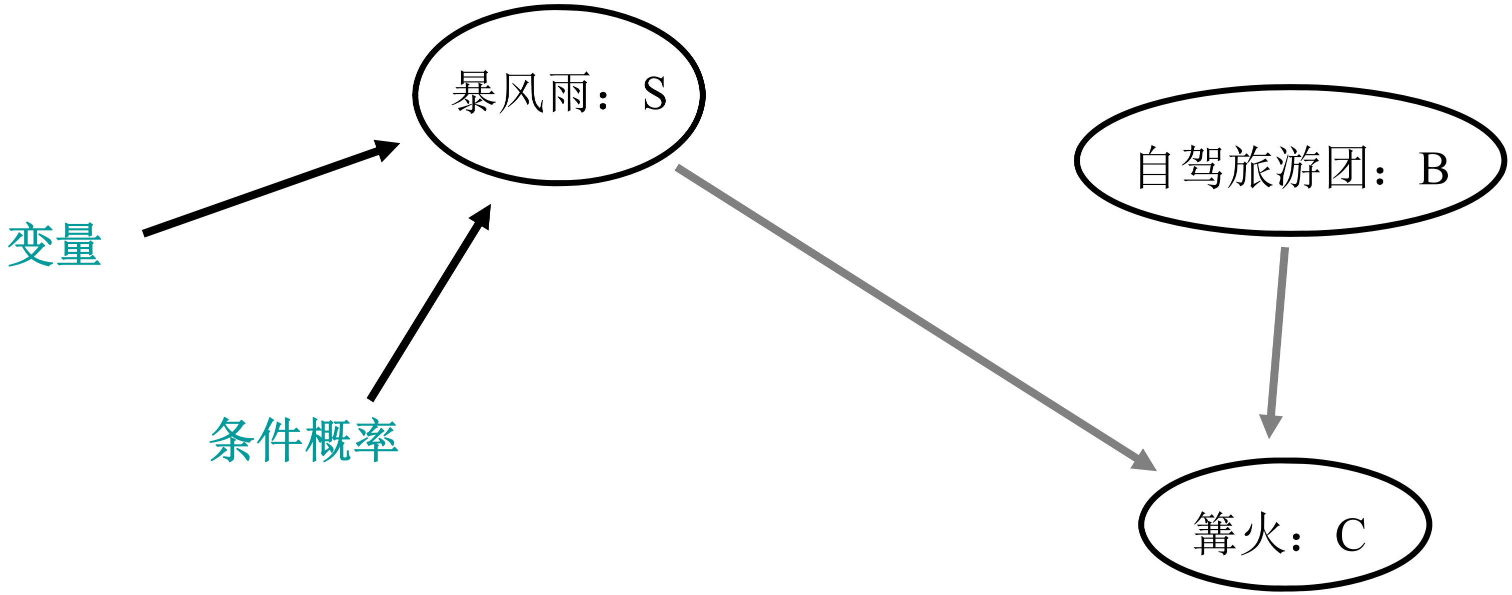
- 对于随机变量集合也具有相同的结论。即

x_1, x_2, x_3 条件独立于给定 z_1, z_2, z_3 时的 y_1, y_2, y_3 是指：

$$P\left(x_1, x_2, x_3 \middle| y_1, y_2, y_3, z_1, z_2, z_3\right) = P\left(x_1, x_2, x_3 \middle| z_1, z_2, z_3\right)$$

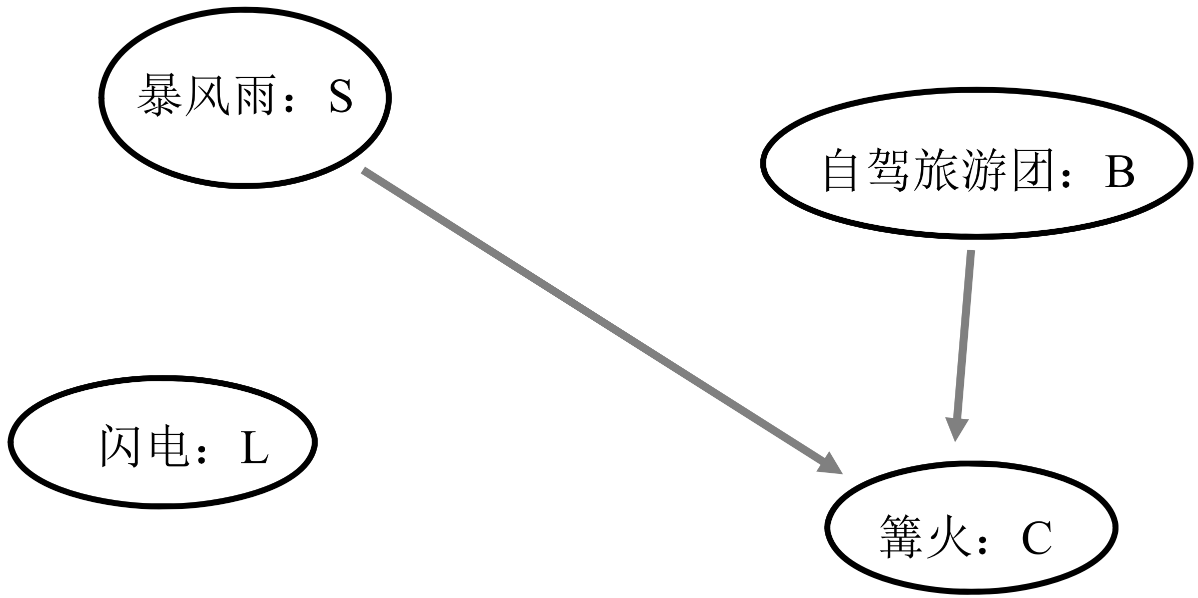
贝叶斯置信网：描述

- BBN 明确而直观地描述了**条件独立假设下**，随机变量集合的**联合概率分布**。具体要用到：
  - 有向无环图（directed acyclic graph）
  - 局部条件概率

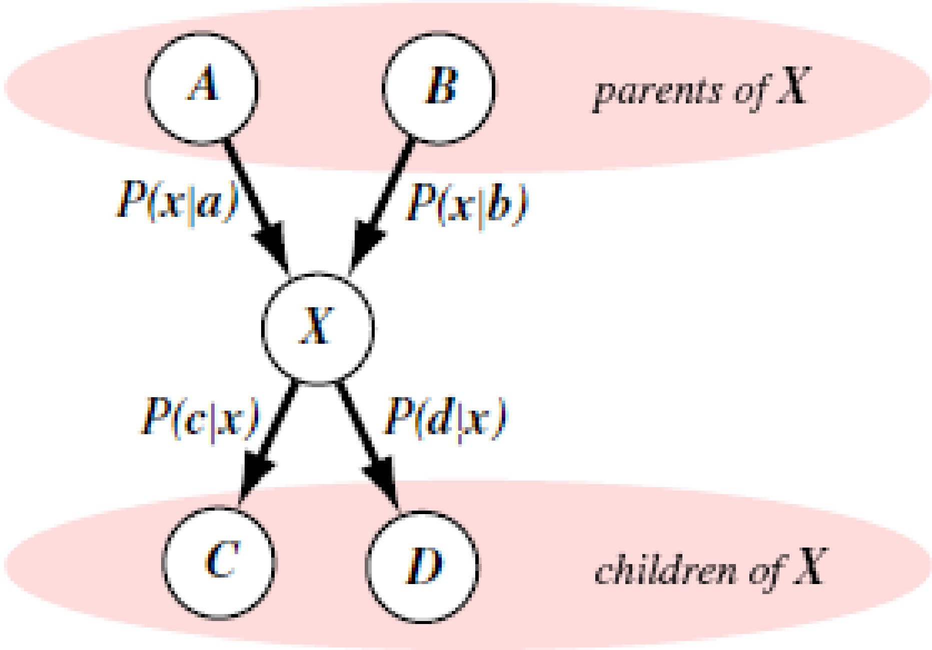


贝叶斯置信网：描述

- 每一个变量与其前辈导出的其他子节点构成**非子代**时，则**子节点间相互独立**。
- x_1 是x_2的一个子孙节点：是指在有向图中存在一条从 x_2 到 x_1的有向路径。例如:
  – “篝火C” 的前辈节点： “暴风雨S”， “自驾旅游团B”。 (C是这两个变量S和B的子孙节点).
  – “篝火C” 独立于由它的前辈节点给出的子孙节点 “闪电L”



- 如图为某置信网的一部分。包含一个节点 ***X***, 变量取值为 ($x_1, x_2, ...$), 其父节点 (**A 和 B**), 子节点 (**C 和 D**)



联合概率分布

- 给定一个贝叶斯置信网 BBN，用下面的公式可以简单地计算出一个变量集合的联合概率密度:

$$P(x_1, x_2, \cdots, x_n) = \prod_i P\left(x_i \middle| parents(x_i)\right)$$

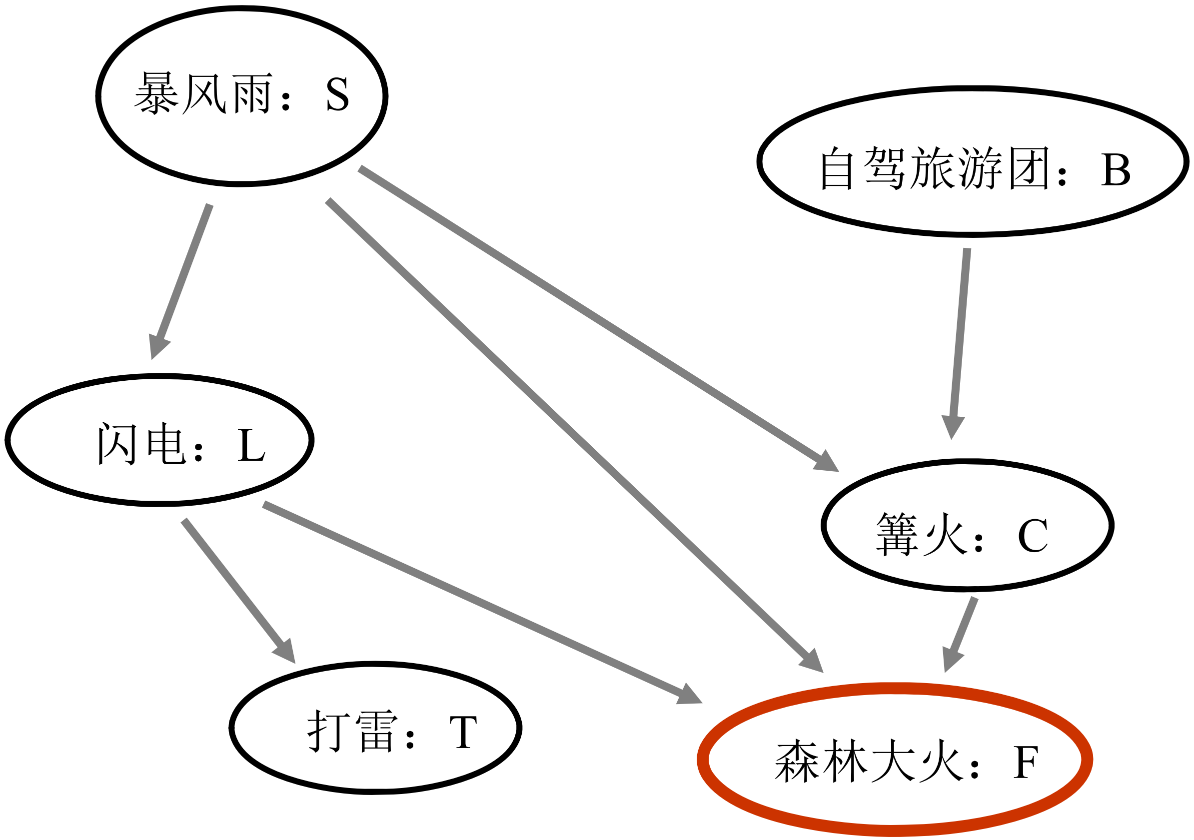
- 这里 parents 是x_i的直接前辈节点。

例如:

P(C, S, B, L, T, F)?

= P(S)P(B)P(C|S,B) P(L|S)P(T|L)P(F|L,S,C).

联合分布：一个例子



$$P(C, S, B, L, T, F) = P(S)P(B)P(C|S,B)P(L|S)P(T|L)P(F|L,S,C).$$

置信网的学习

BBN 可以采用不同的方式学习。下面是两个基本的方法:

1）假设已知网络结构:

– 可以从已知数据的每一个变量估计条件概率

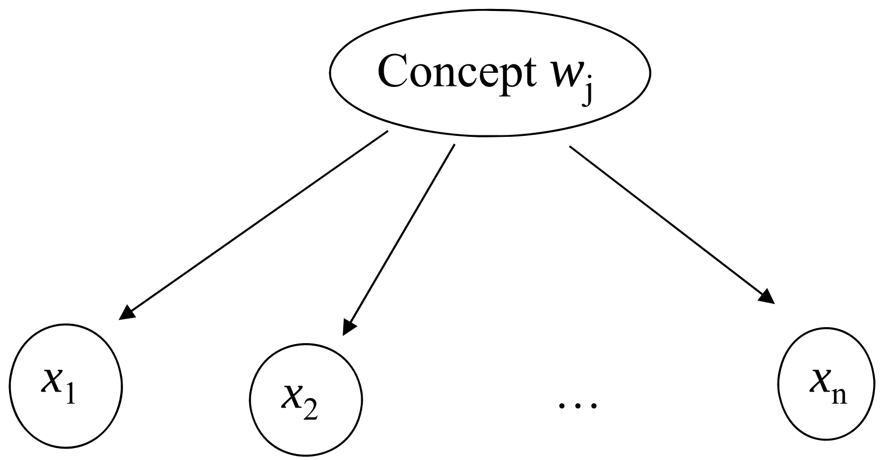
2）假设知道部分的网络结构，但是丢失了一些变量:

– 类似于在神经网络中对隐单元节点的学习。
– 可以用梯度下降的方法训练 BBN.

- 假设什么都是未知的
  – 我们可以通过可能的网络空间中进行观察、学习网络结构和条件概率。

BBN与分类的关系

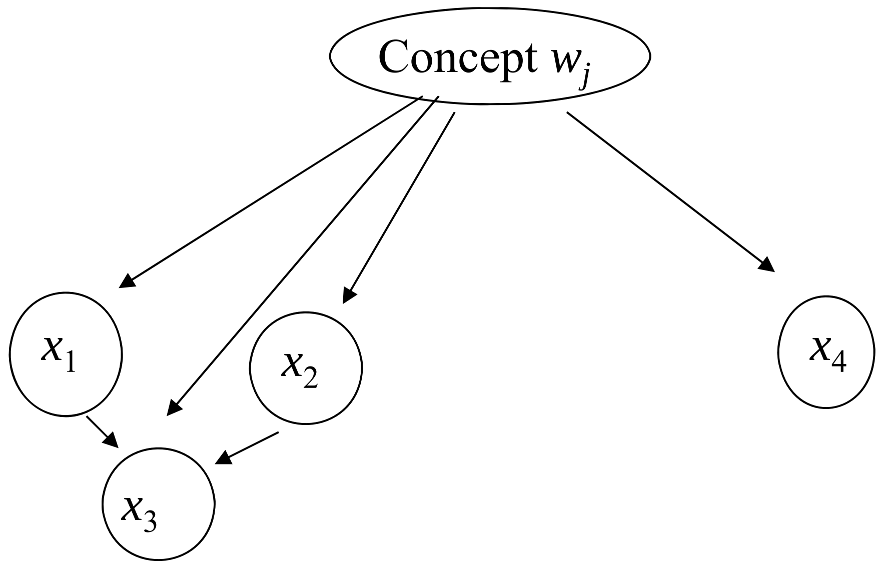
- 假设其中的一个变量是目标变量，可以计算给定其他变量下目标变量的概率。



$$P\left(x_1,x_2,\cdots,x_n,\omega_j\right)=P\left(\omega_j\right)P\left(x_1\middle|\omega_j\right)P\left(x_2\middle|\omega_j\right)\cdots P\left(x_n\middle|\omega_j\right)$$

一般情况

- 在一般情况下，可以使用 BBN 详细说明变量之间的独立性假设.



$$P\left(x_1,x_2,\cdots,x_n,\omega_j\right)=P\left(\omega_j\right)P\left(x_1\middle|\omega_j\right)P\left(x_2\middle|\omega_j\right)P\left(x_3\middle|x_1,x_2,\omega_j\right)P\left(x_4\middle|\omega_j\right)$$

本章要点

① 贝叶斯公式和贝叶斯决策论

② 损失函数 **(e.g., zero-one loss)**

③ 判别函数

④ **ROC** 曲线

⑤ 独立二值特征下的判别函数

⑥ 条件独立

⑦ 贝叶斯置信网 **BBN**