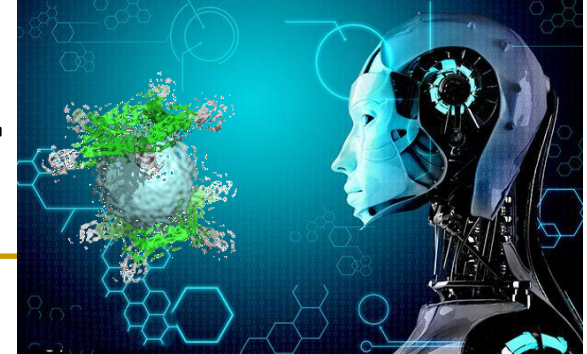


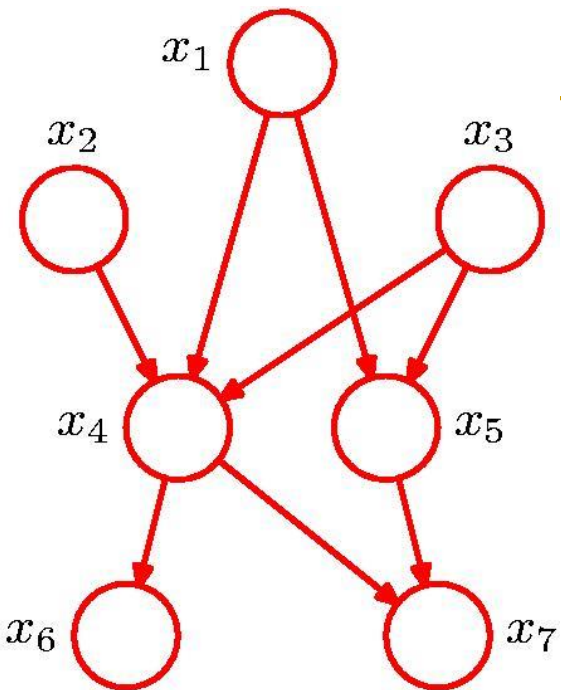
《机器学习及其在化学中的应用》2025年课程



Sec. 7

概率图模型

(贝叶斯网络；朴素贝叶斯分类器；马尔科夫随机场；玻尔兹曼机)



刘志荣 (LiuZhiRong@pku.edu.cn)

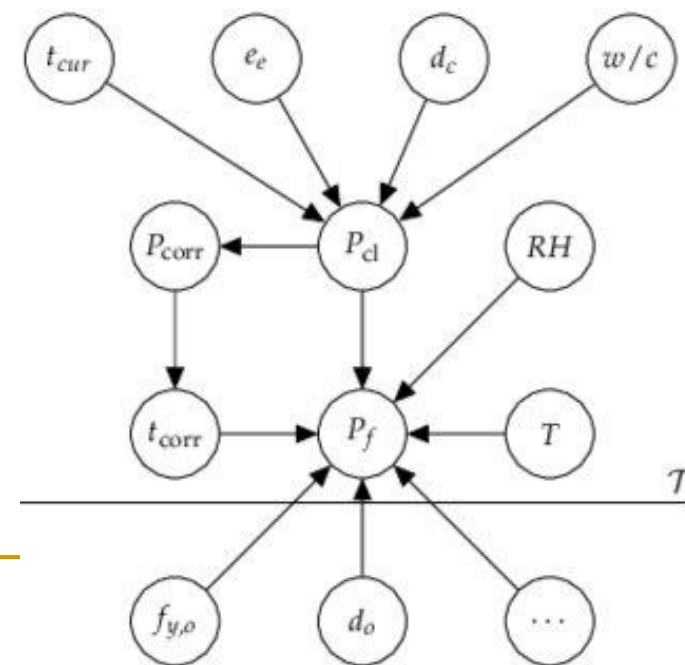
北京大学化学学院

2025.11.3

内容提要

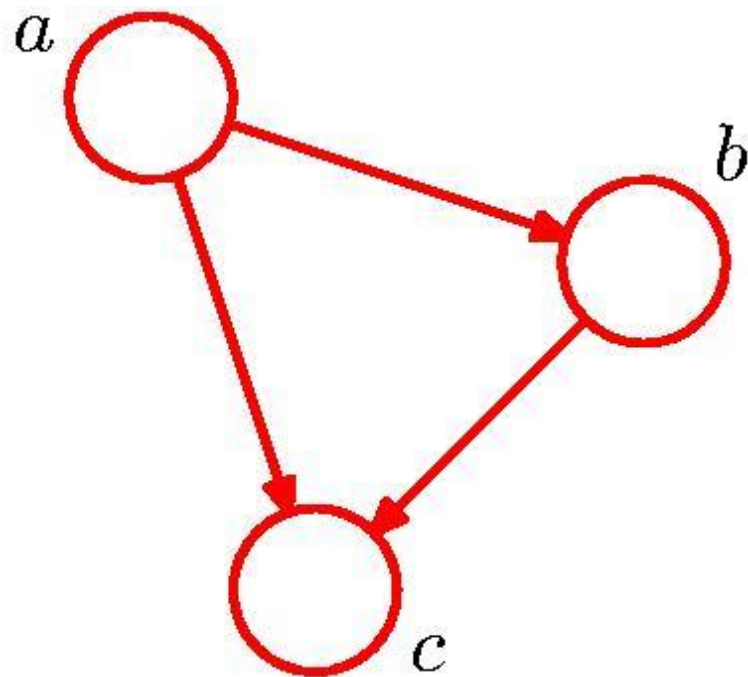
- 有向图：贝叶斯网络
- 朴素贝叶斯
- 无向图：马尔科夫随机场与玻尔兹曼机
- 应用例子

1. 贝叶斯网络... (Bayesian Networks)



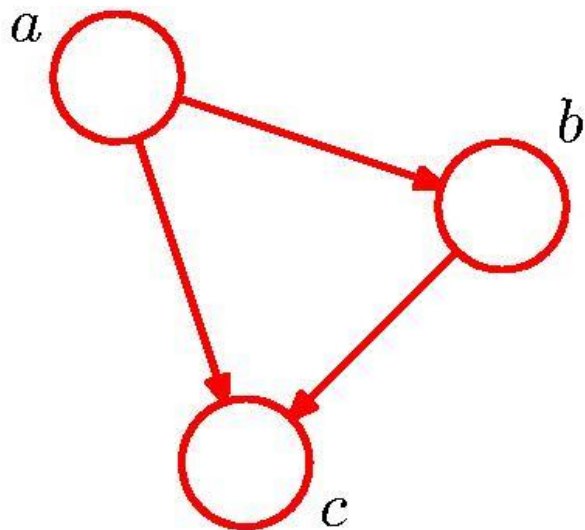
概率图模型与贝叶斯网络

- 利用概率分布来讨论体系性质，是机器学习的一种重要思路。
- 概率图模型（Probabilistic Graphical Model）是用图来表示变量概率依赖关系的模型，通过结合概率论与图论的知识，有效地处理与联合概率分布有关的性质。
 - 节点：随机变量
 - 边：随机变量的依赖或相关关系，严格讲并不代表因果关系。
- 贝叶斯网络（又称信念网络，Belief Network）用有向边表示单向的依赖（条件概率）。
 - 是模拟人类推理过程中因果关系的不确定性处理模型。



贝叶斯网络：普适表达式

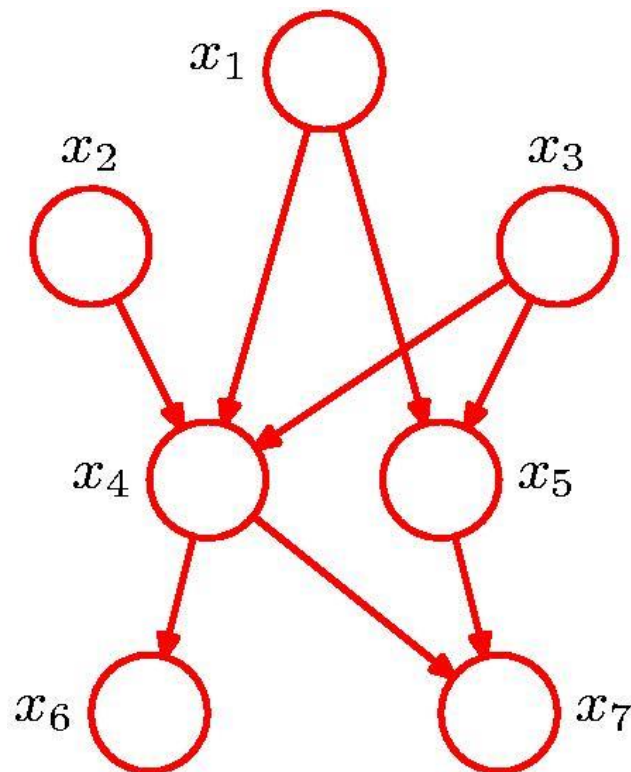
- 有向无环图。
- 因子化形式。



$$\begin{aligned} p(a, b, c) &= p(c|a, b)p(a, b) \\ &= p(c|a, b)p(b|a)p(a) \end{aligned}$$

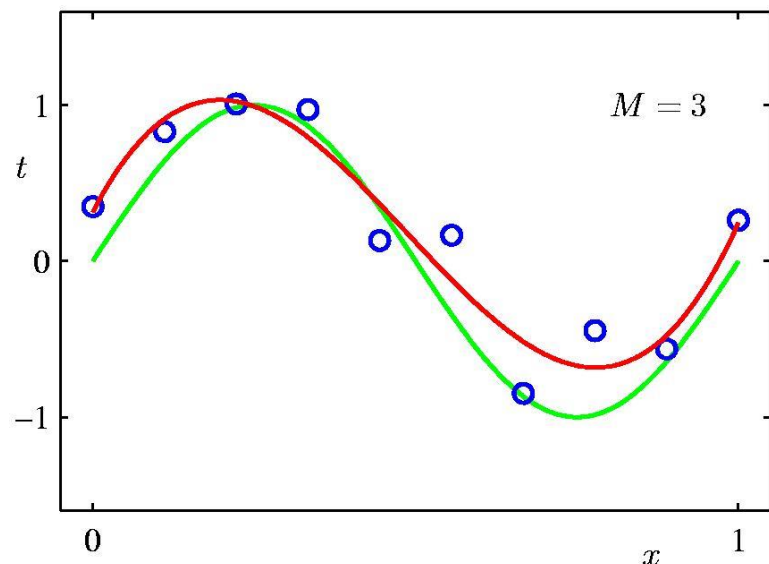
$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k | \text{pa}_k)$$

节点 k 的所有父节点



$$\begin{aligned} p(x_1, \dots, x_7) &= \\ & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\ & p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5) \end{aligned}$$

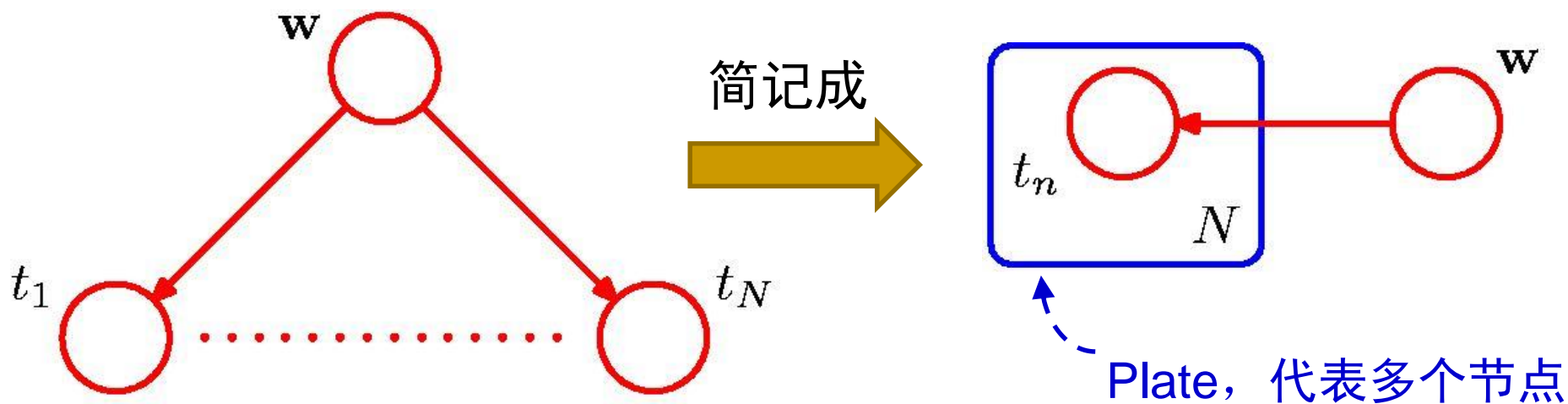
例子：贝叶斯多项式回归



$$y(x, \mathbf{w}) = \sum_{j=0}^M w_j x^j$$

$$t_n = y(x_n, \mathbf{w}) + \epsilon_n$$

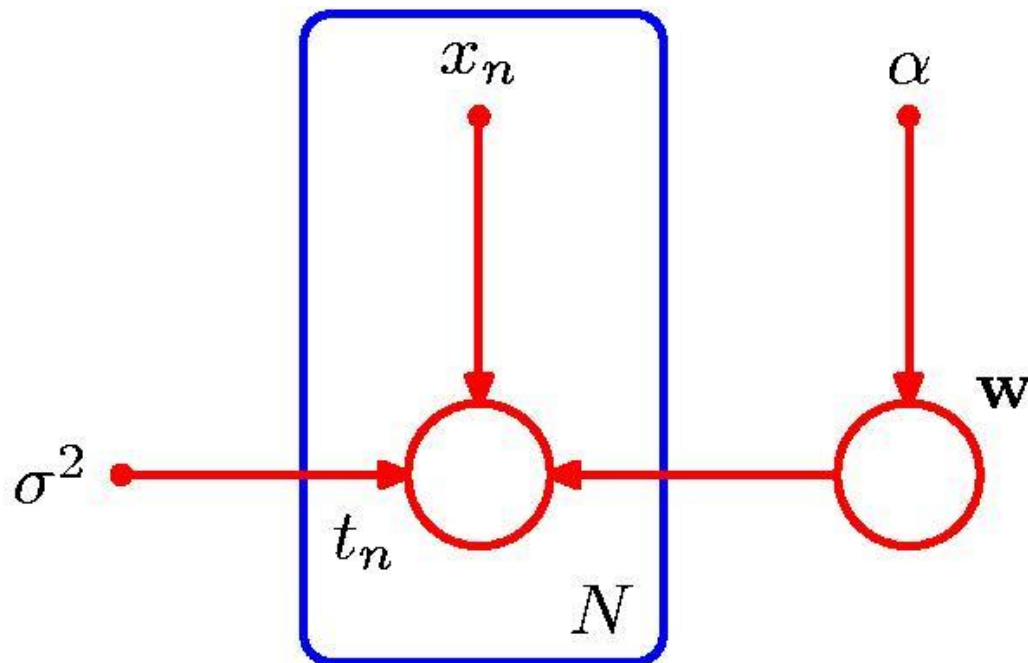
$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^N p(t_n | y(\mathbf{w}, x_n))$$



一些记号...

- 也可用圆点显式表示固定的值或量，例如输入自变量 x_n 与超参数 α

$$p(\mathbf{t}, \mathbf{w} | \mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w} | \alpha) \prod_{n=1}^N p(t_n | \mathbf{w}, x_n, \sigma^2).$$

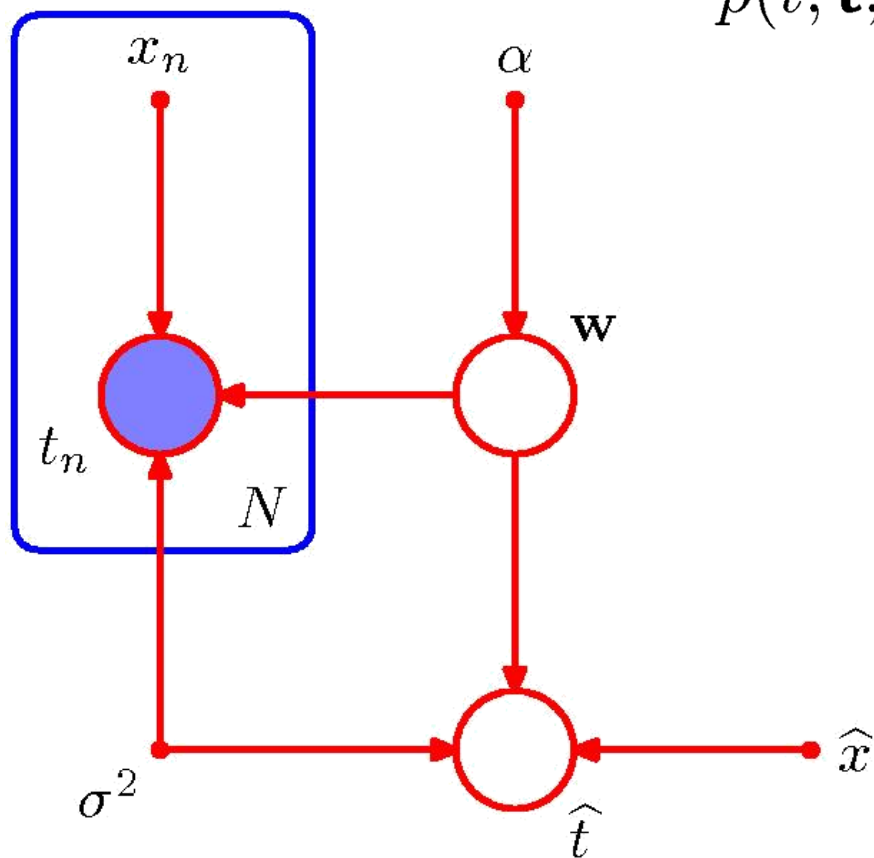


- 用带阴影的节点表示观测量。

- 预测值分布: $p(\hat{t}|\hat{x}, \mathbf{x}, \mathbf{t}, \alpha, \sigma^2) \propto \int p(\hat{t}, \mathbf{t}, \mathbf{w}|\hat{x}, \mathbf{x}, \alpha, \sigma^2) d\mathbf{w}$

其中

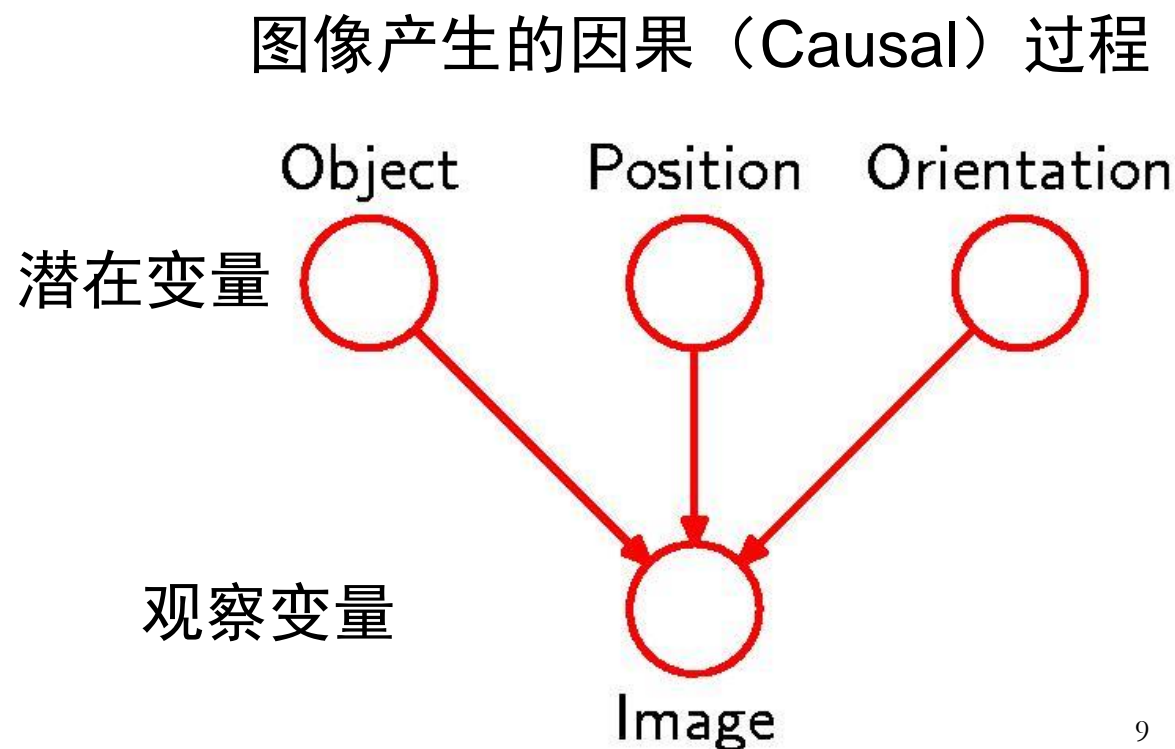
$$p(\hat{t}, \mathbf{t}, \mathbf{w}|\hat{x}, \mathbf{x}, \alpha, \sigma^2) = \left[\prod_{n=1}^N p(t_n|x_n, \mathbf{w}, \sigma^2) \right] p(\mathbf{w}|\alpha)p(\hat{t}|\hat{x}, \mathbf{w}, \sigma^2)$$



证据的出现改变信念!

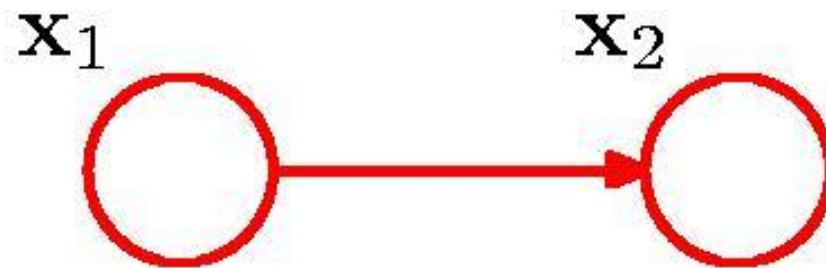
生成模型（Generative Models）

- 通过估计联合分布 $p(\mathbf{x}, t)$ 来预测 $t(\mathbf{x})$ ，例如直方图方法。
 - 也可以用来产生数据点。AI画图；ChatGPT。
- 相对：判别模型（discriminative models）
 - 通过估计 $p(t|\mathbf{x})$ 来预测 $t(\mathbf{x})$ 。
 - 线性回归、逻辑回归、
传统神经网络、支持向量机。
- 考虑数据的某种产生过程。
- 例子：图像识别
 - 回推，预测潜在变量的值，
因此与贝叶斯公式相关。



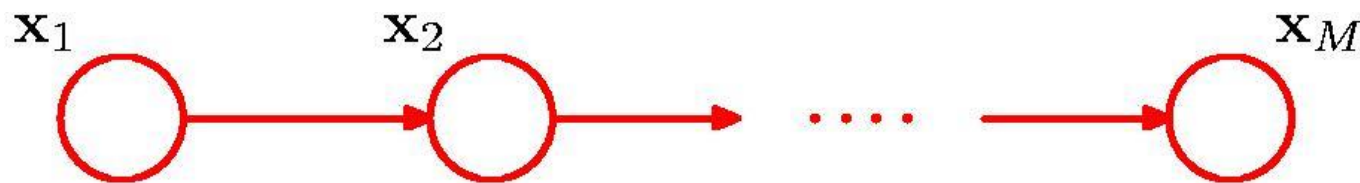
概率分布的参数数目

- 如果一个变量 x 有 K 个离散值，则在普适情况下其分布 $p(x) = \{p_1, p_2, \dots, p_K\}$ 在满足归一化条件 ($\sum_i p_i = 1$) 下需要 $K - 1$ 个参数来描述。
- 两个变量则需 $K^2 - 1$ 个参数。
- 维度灾难： M 个变量需 $K^M - 1$ 个参数。
- 经常引入简化假设以减少参数个数。
- 例如：朴素贝叶斯模型
 - 节点之间没有连线，变量之间是独立的。
 - M 个变量，参数个数： $M(K - 1)$

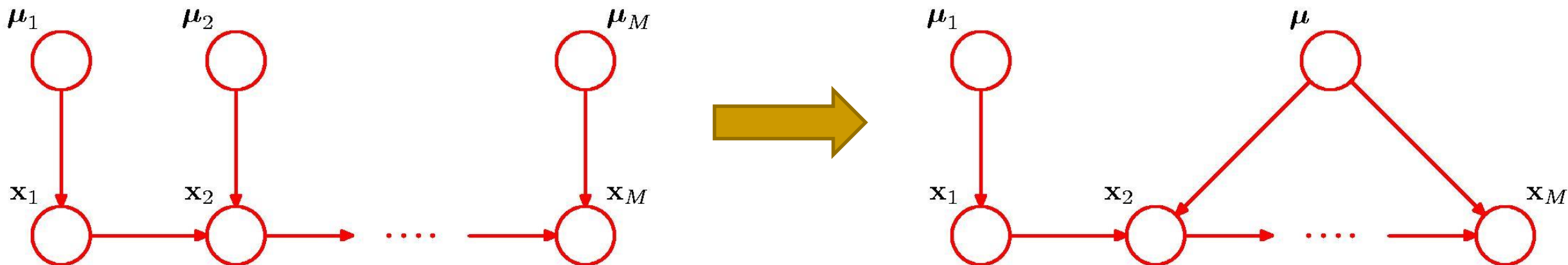


一些简化情形...

- 串联依赖，参数个数： $K - 1 + (M - 1)K(K - 1)$



- 参数共享:



- 另一种减少参数数目的思路：概率的函数（公式）化，这样只需要确定其参数。
- 但代价是对函数形式的假设。
- 例如，逻辑回归值只需要 $M + 1$ 个参数：

$$p(y = 1|x_1, x_2, \dots, x_M) = \sigma \left(w_0 + \sum_{i=1}^M w_i x_i \right)$$

重要性质：条件独立性

- a 与 b 是独立的：

$$p(a, b) = p(a)p(b) \iff p(a|b) = p(a) \iff p(b|a) = p(b)$$

- a 与 b 在给定 c 的条件下是独立的 (a is conditionally independent of b given c) :

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

$$\iff p(a|b, c) = p(a|c)$$

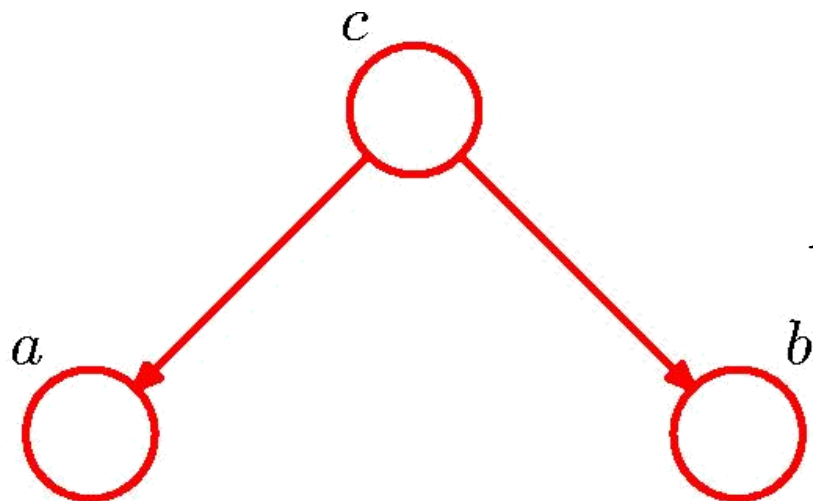
记成

$$a \perp\!\!\!\perp b \mid c$$

情况1: tail-to-tail

■ 例子:

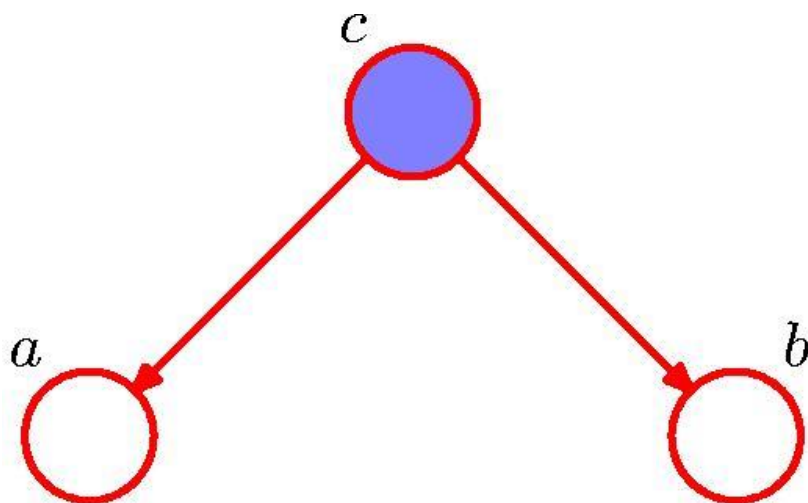
孩子年龄;
鞋子大小;
阅读能力。



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

$$p(a, b) = \sum_c p(a|c)p(b|c)p(c)$$

$$a \not\perp b \mid \emptyset$$

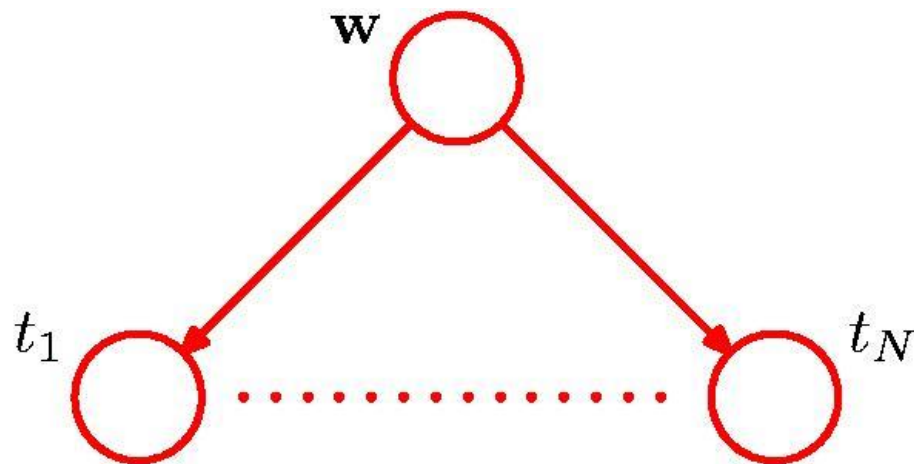


$$p(a, b|c) = \frac{p(a, b, c)}{p(c)}$$

$$= p(a|c)p(b|c)$$

$$a \perp b \mid c$$

例子：贝叶斯多项式回归



$$y(x, \mathbf{w}) = \sum_{j=0}^M w_j x^j$$

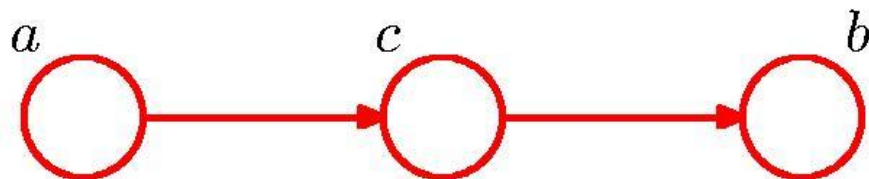
$$t_n = y(x_n, \mathbf{w}) + \epsilon_n$$

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^N p(t_n | y(\mathbf{w}, x_n))$$

- \mathbf{w} 是未知的，因此不同 x_n 的 t_n 不是独立的，知道其中一些值将有助于对其它值的推断。

情况2: head-to-tail

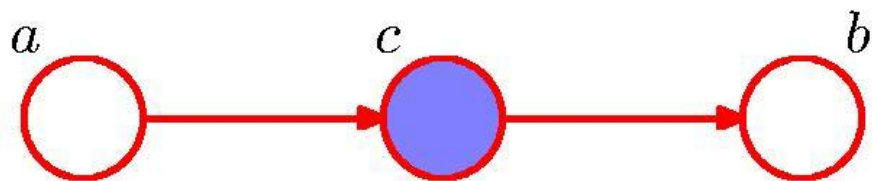
- 例子：火灾->烟雾->报警器。



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

$$p(a, b) = p(a) \sum_c p(c|a)p(b|c) = p(a)p(b|a)$$

$$a \not\perp\!\!\!\perp b \mid \emptyset$$



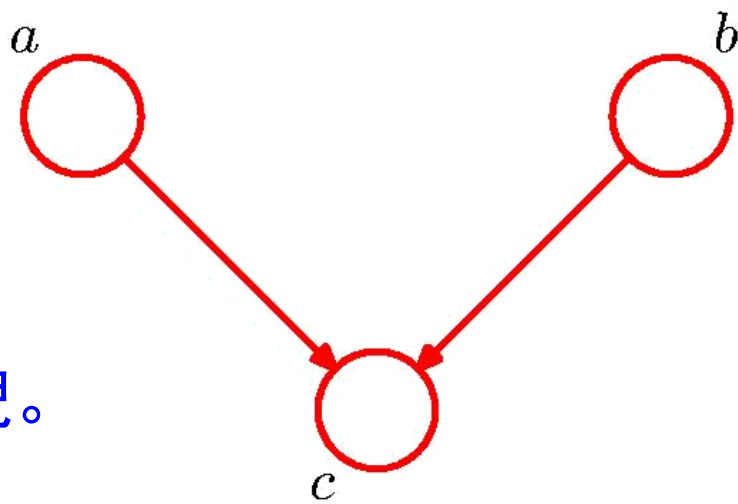
$$\begin{aligned} p(a, b|c) &= \frac{p(a, b, c)}{p(c)} \\ &= \frac{p(a)p(c|a)p(b|c)}{p(c)} \\ &= p(a|c)p(b|c) \end{aligned}$$

$$a \perp\!\!\!\perp b \mid c$$

情况3: head-to-head

■ 例子:

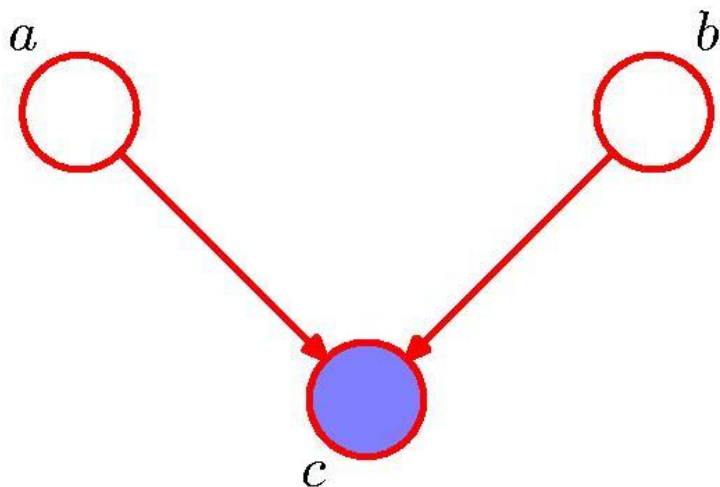
名人: 才能或外貌。



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

$$p(a, b) = p(a)p(b)$$

$$a \perp\!\!\!\perp b \mid \emptyset$$

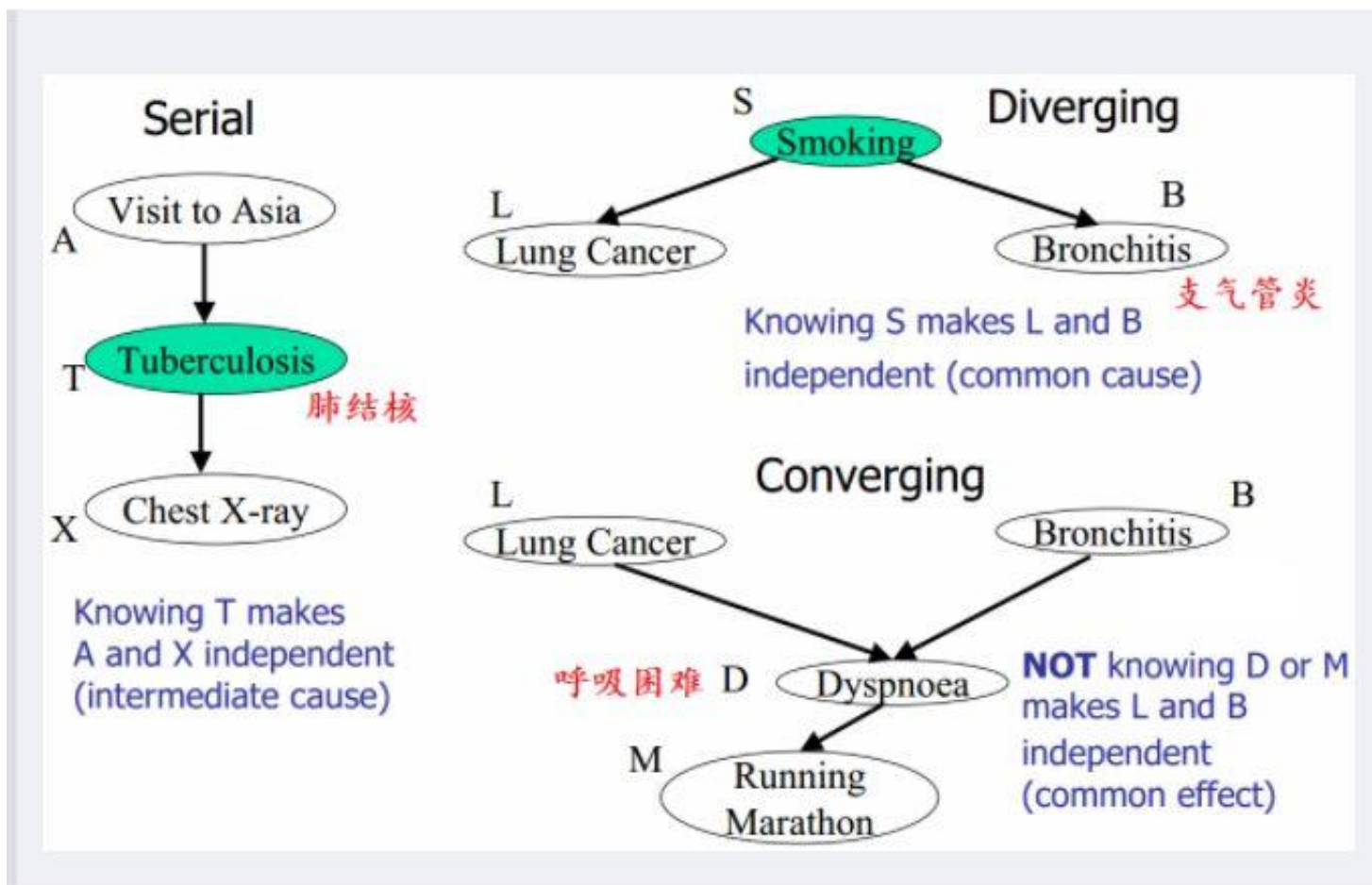


$$\begin{aligned} p(a, b|c) &= \frac{p(a, b, c)}{p(c)} \\ &= \frac{p(a)p(b)p(c|a, b)}{p(c)} \end{aligned}$$

$$a \not\perp\!\!\!\perp b \mid c$$

小结：三种情况

- 知道（固定）共同原因、知道中间原因、或不知道共同结果，将是A与B相互独立。

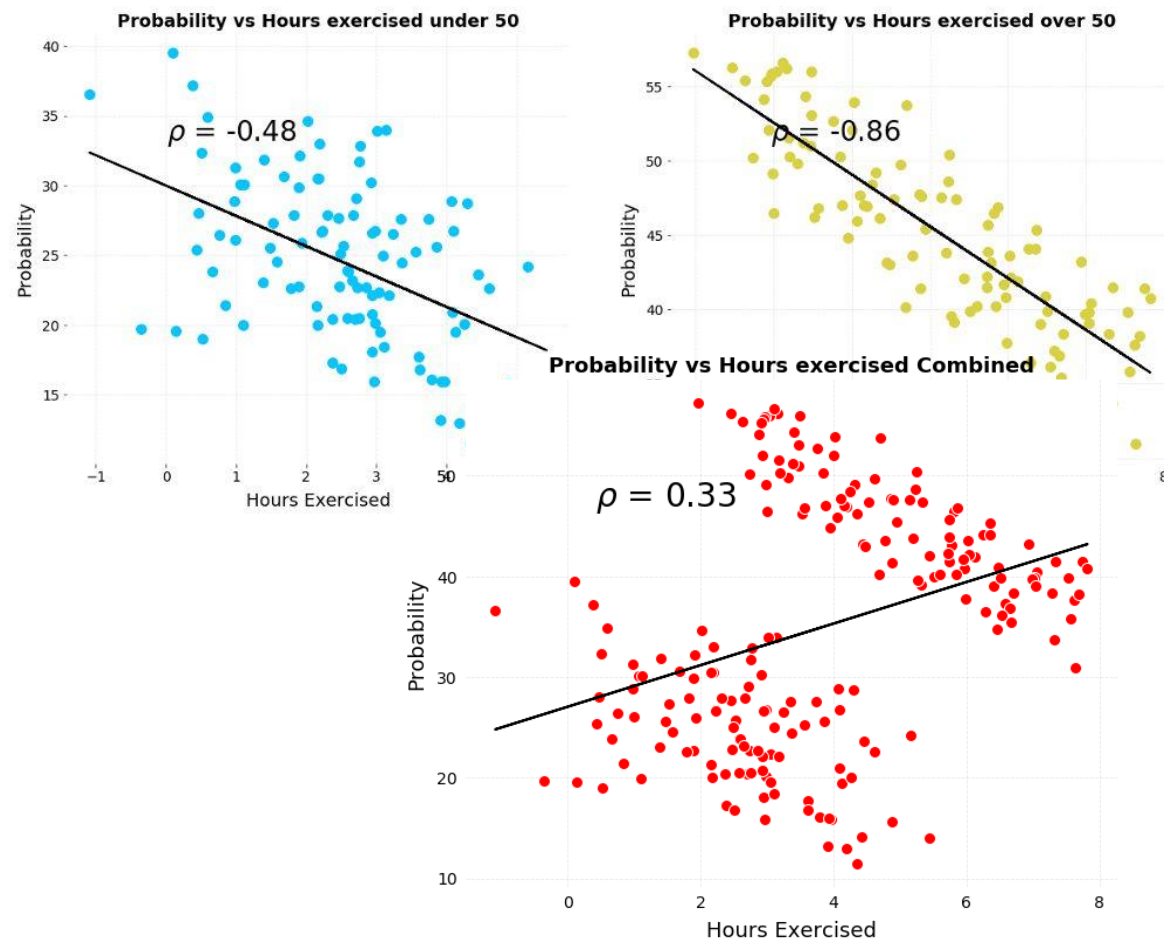


题外：辛普森悖论

- 在分组比较中都占优势的一方，在总评中有时反而是劣势一方。
- 分组引起结论变化，同一个数据集给出完全相反的观点

录取比例

	男生	女生
外语系	$\frac{15}{20} = 75\%$	$\frac{49}{100} = 49\%$
哲学系	$\frac{10}{100} = 10\%$	$\frac{1}{20} = 5\%$
总计	$\frac{15 + 10}{20 + 100} = 21\%$	$\frac{49 + 1}{100 + 20} = 42\%$



用于概率推断的例子：油箱没油了？

■ 电动燃油表的例子

$$p(G = 1 | B = 1, F = 1) = 0.8$$

$$p(G = 1 | B = 1, F = 0) = 0.2$$

$$p(G = 1 | B = 0, F = 1) = 0.2$$

$$p(G = 1 | B = 0, F = 0) = 0.1$$

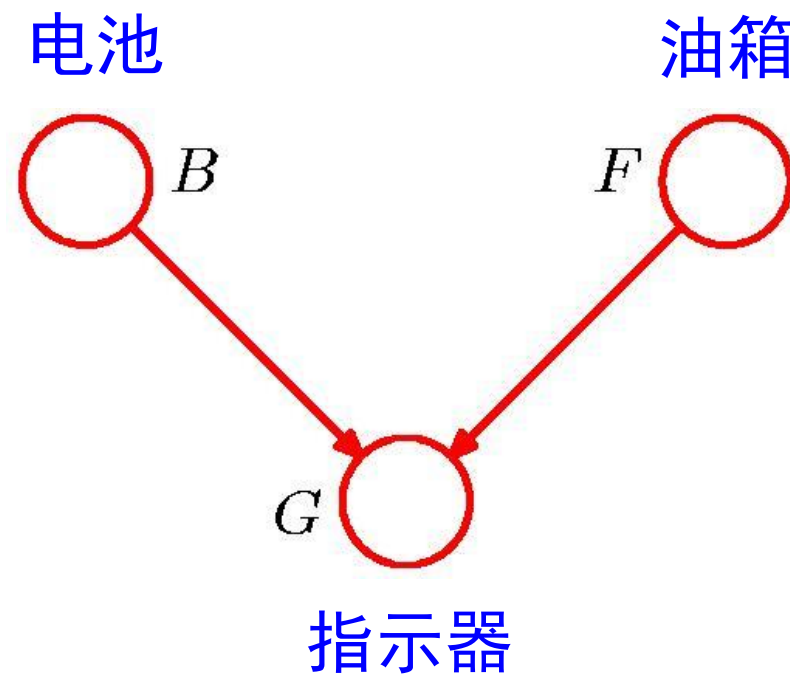
假设先验概率

$$p(B = 1) = 0.9$$

$$p(F = 1) = 0.9$$

and hence

$$p(F = 0) = 0.1$$

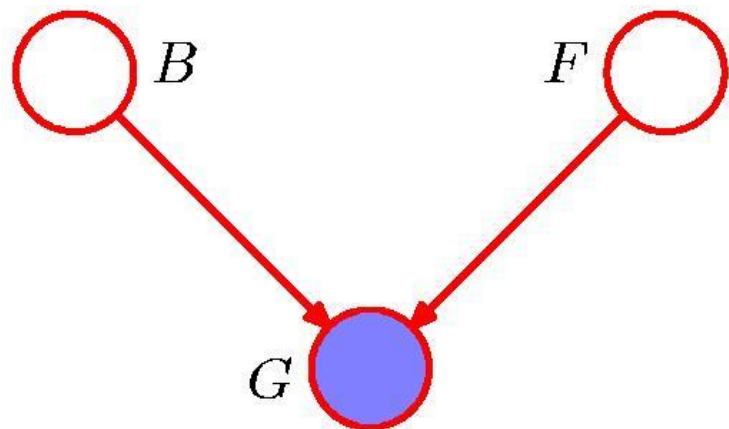


B = Battery (0=flat, 1=fully charged)

F = Fuel Tank (0=empty, 1=full)

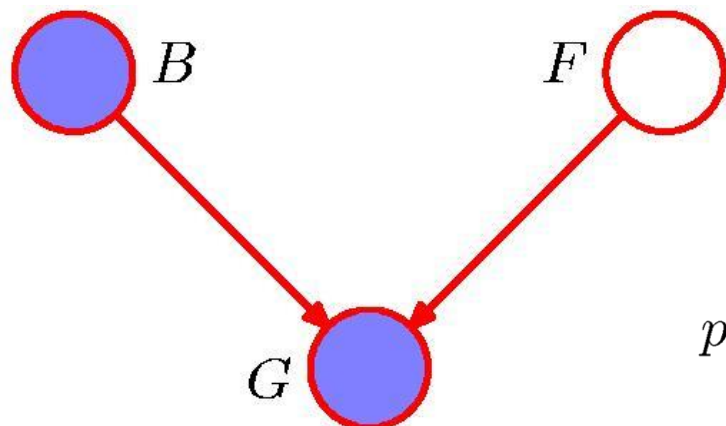
G = Fuel Gauge Reading
(0=empty, 1=full)

证据的出现改变信念！



$$p(F = 0|G = 0) = \frac{p(G = 0|F = 0)p(F = 0)}{p(G = 0)} \\ \simeq 0.257$$

油箱没油的概率随着观察到G=0而增大。



$$p(F = 0|G = 0, B = 0) = \frac{p(G = 0|B = 0, F = 0)p(F = 0)}{\sum_{F \in \{0,1\}} p(G = 0|B = 0, F)p(F)} \\ \simeq 0.111$$

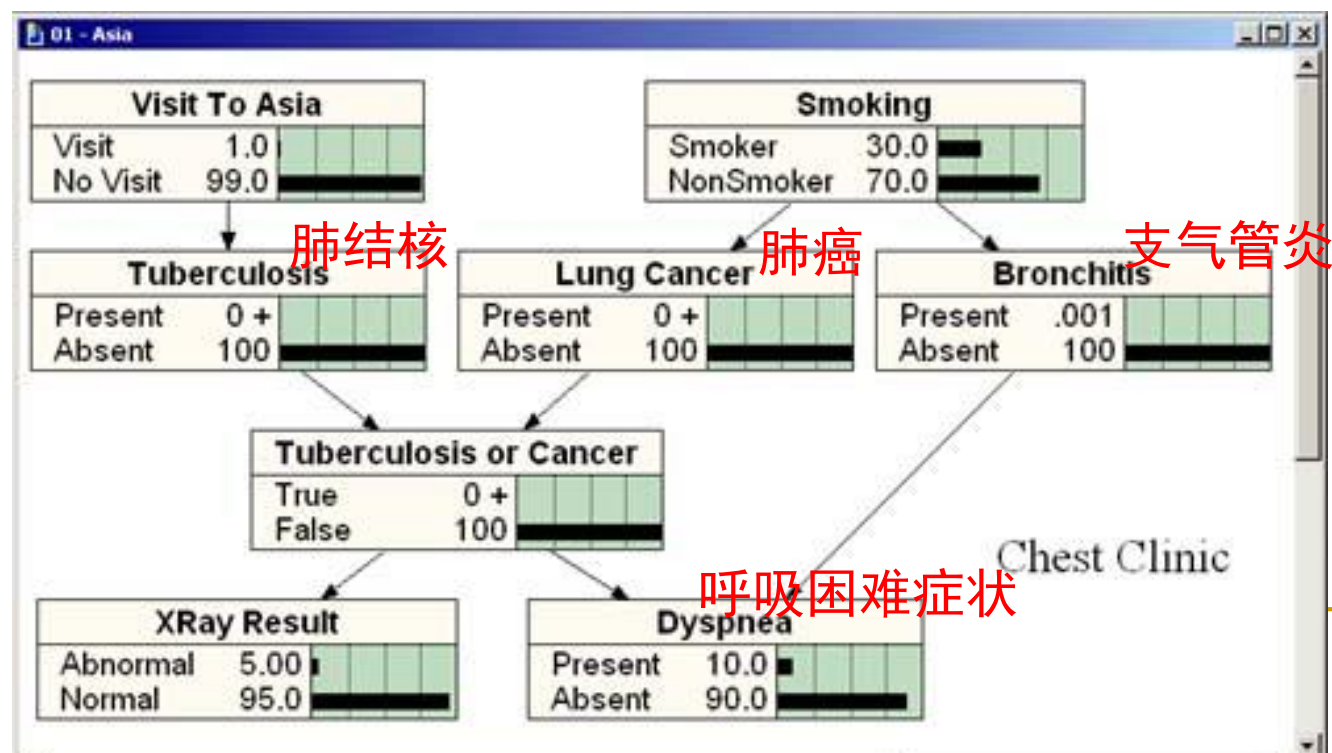
再观察到B=0，油箱没油的概率减小了。

贝叶斯网络为什么有用？

- 只需要分别考虑每个节点与其父母节点的联合分布，而不需所有变量的联合分布。
- 灵活，便于扩展，容易加入新知识。
- 不精确的（概率）信息也能有用。
 - Probabilities need not be exact to be useful!

例子：胸部疾病诊所（Chest Clinic）

- 美国有30%的人吸烟。
- 每10万人中有70人患有肺癌。
- 每10万人中有10人患有肺结核。
- 每10万人中就就有800人患有支气管炎。
- 10%人存在呼吸困难症状。

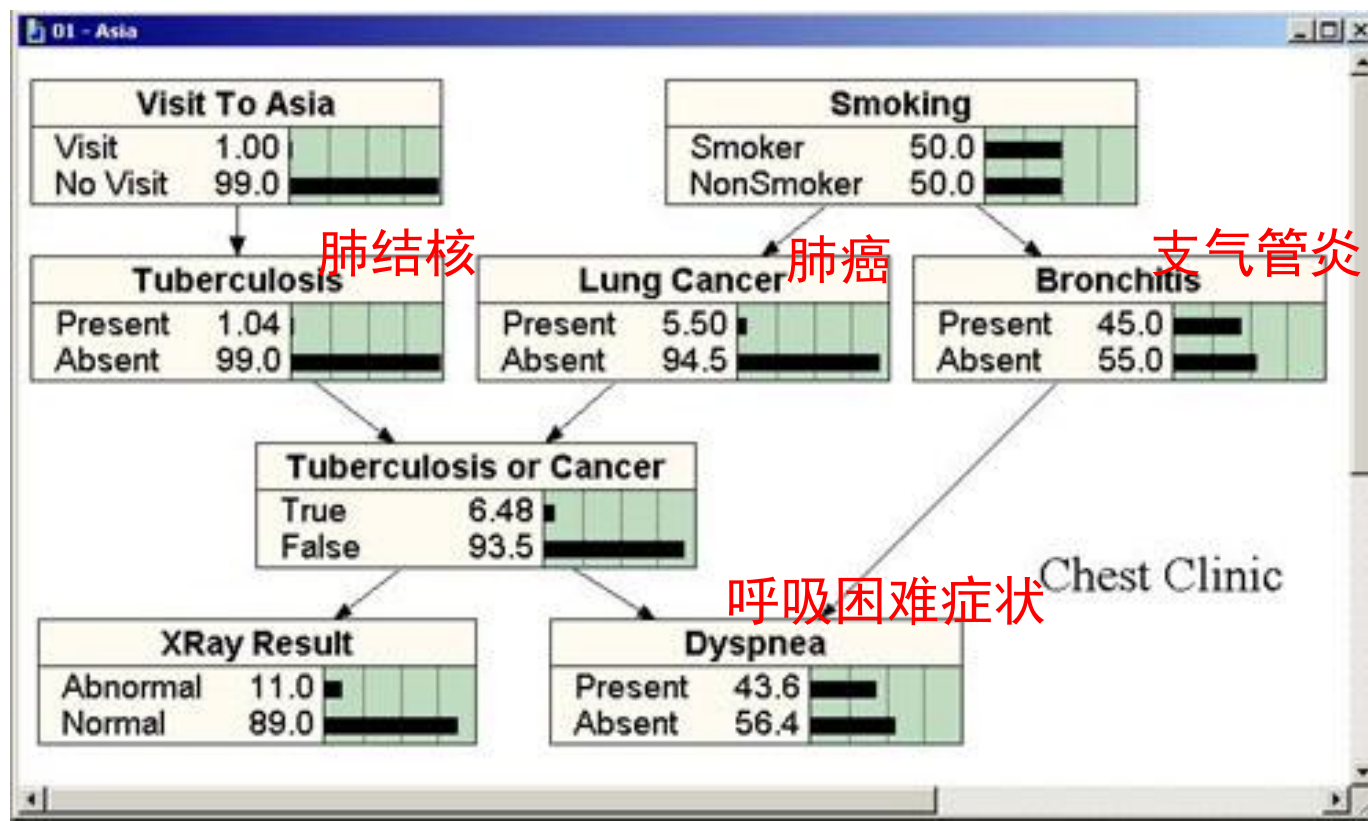


当诊所诊治了数千病人后，

- 50%的病人吸烟.
- 1%患有肺结核.
- 5.5% 得了肺癌.
- 45% 患有不同程度支气管炎.

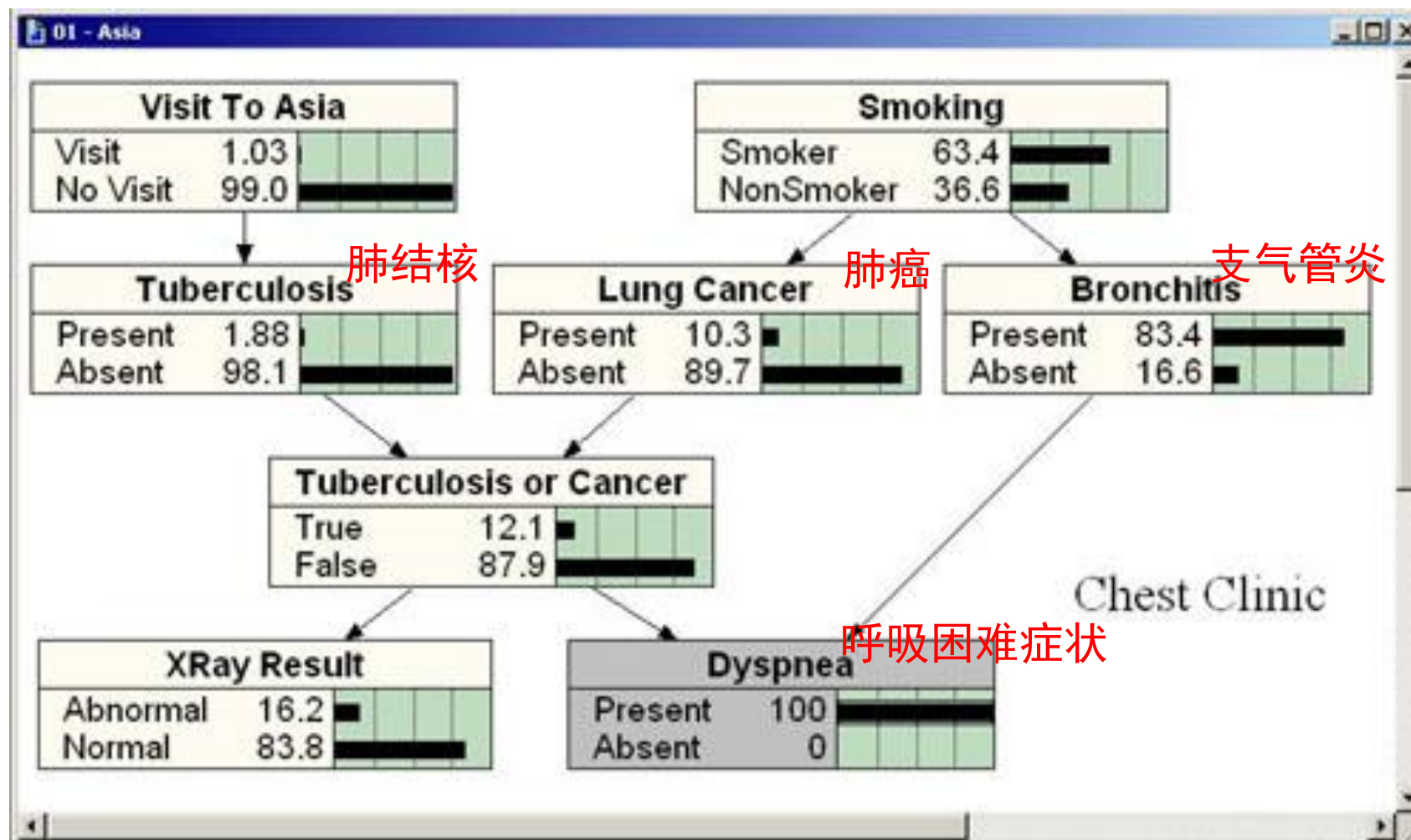
及其（因果）关联关系。

利用这些新数据（及其关联）
输入到BN模型中，才获得了有
意义的实用BN模型。

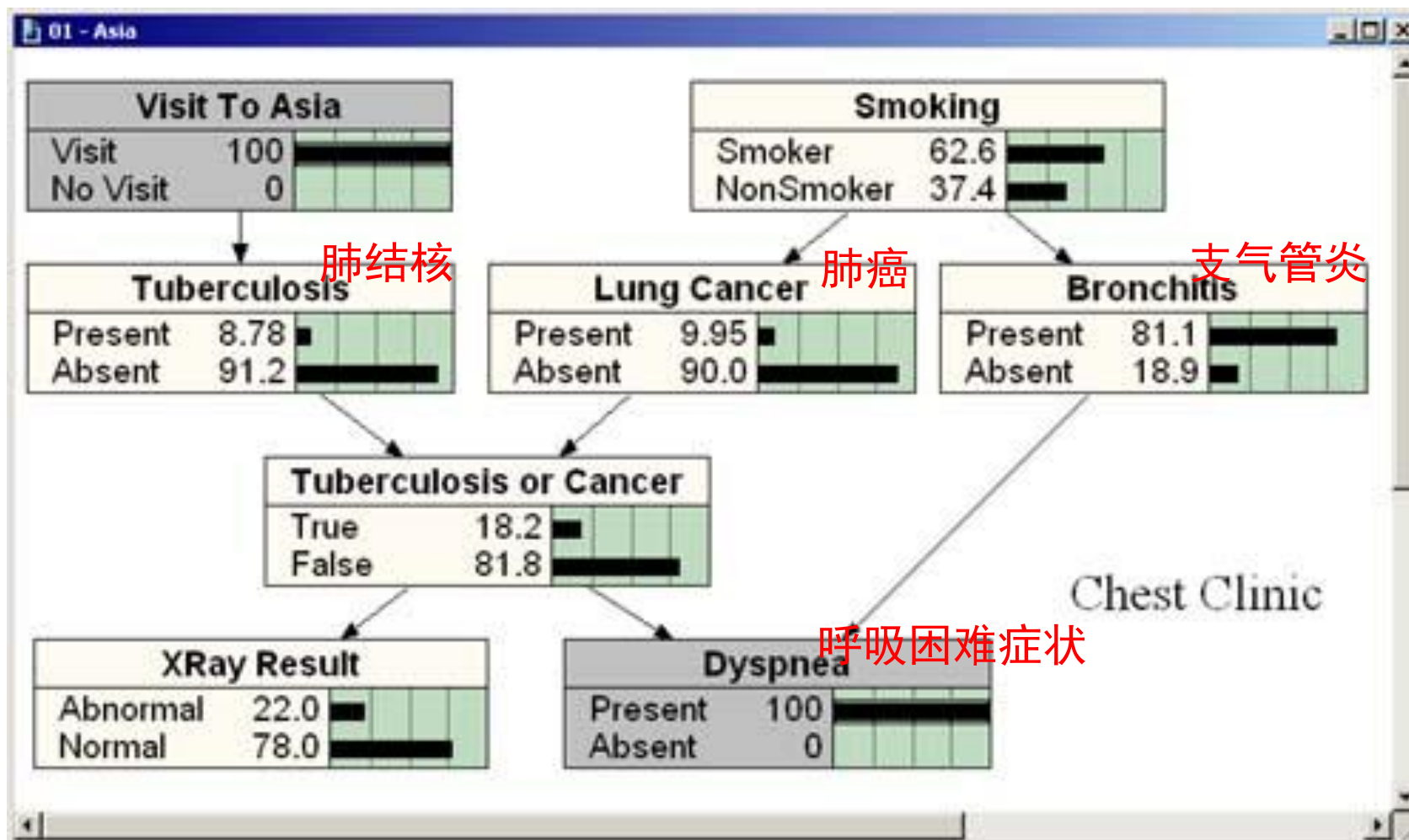


与病人交谈，发现：

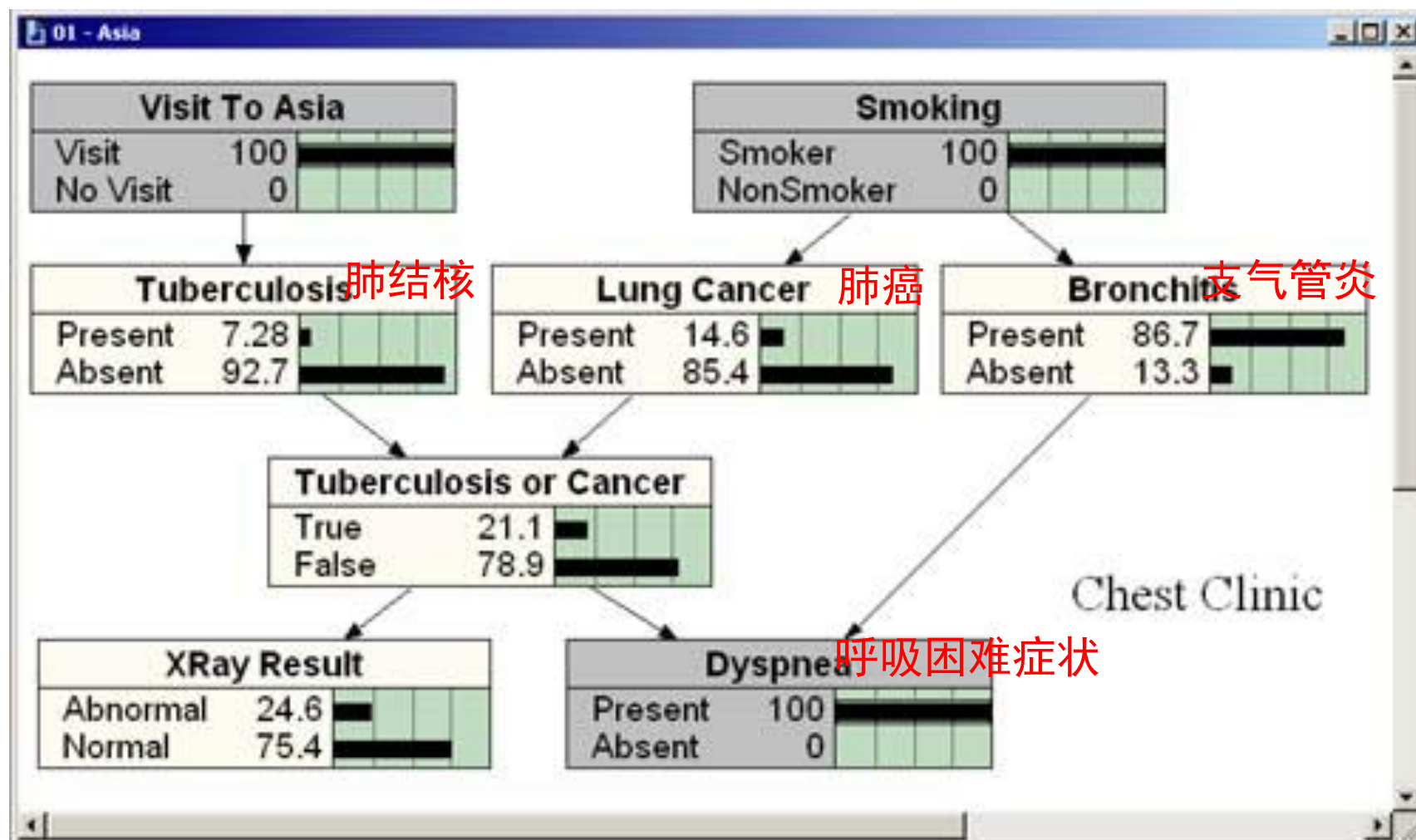
- 有呼吸困难症状。



- 去过亚洲国家？
- 答：是。

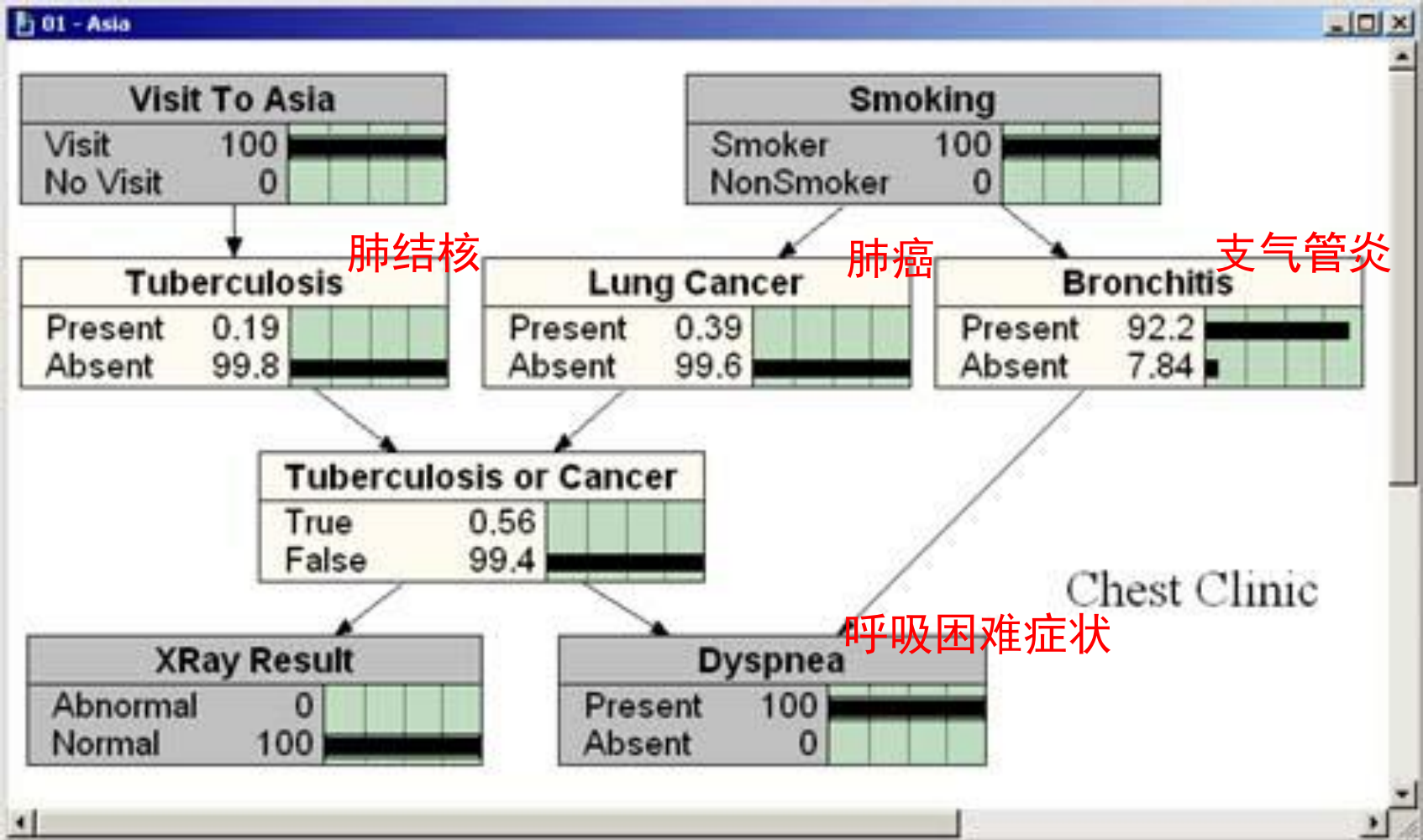


- 是个吸烟者

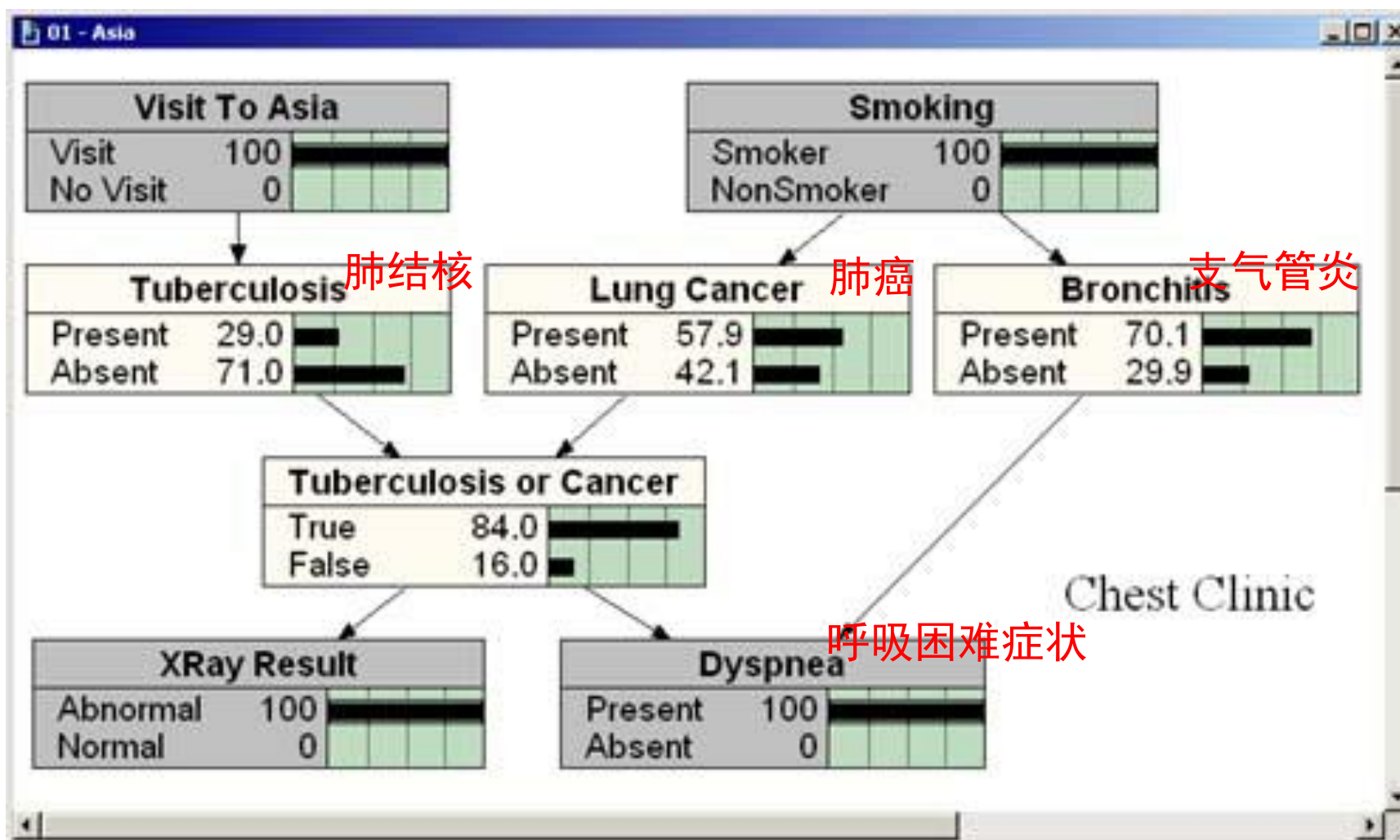


做一个X光透视，

- 结果正常。



- 如果X光透视结果不正常:



题外...

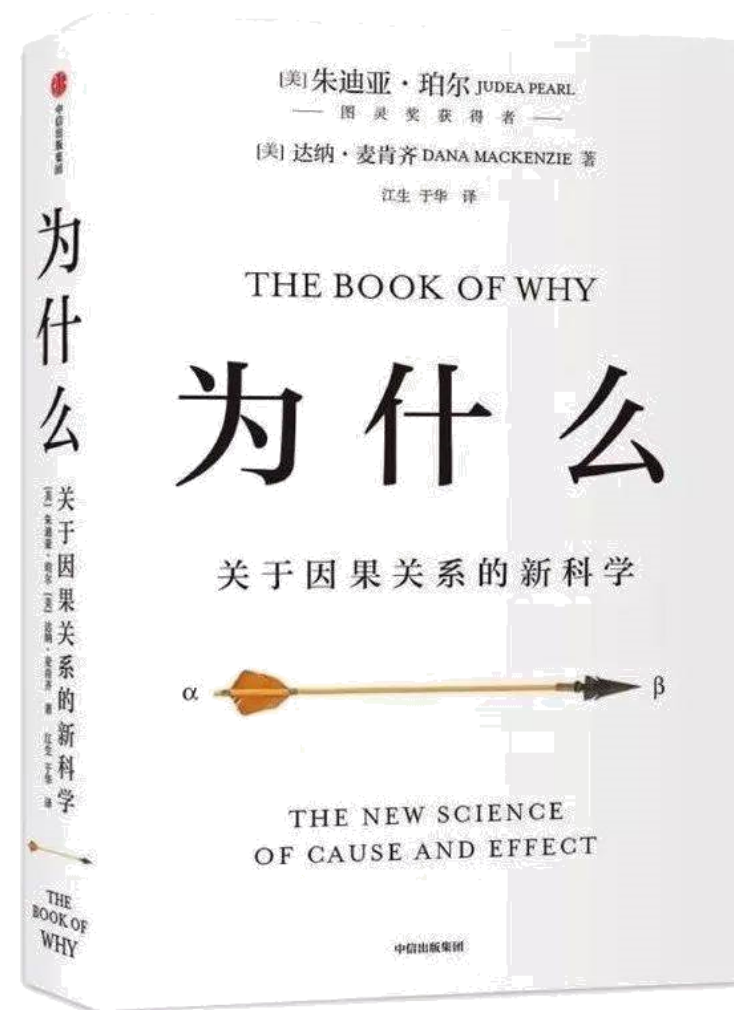
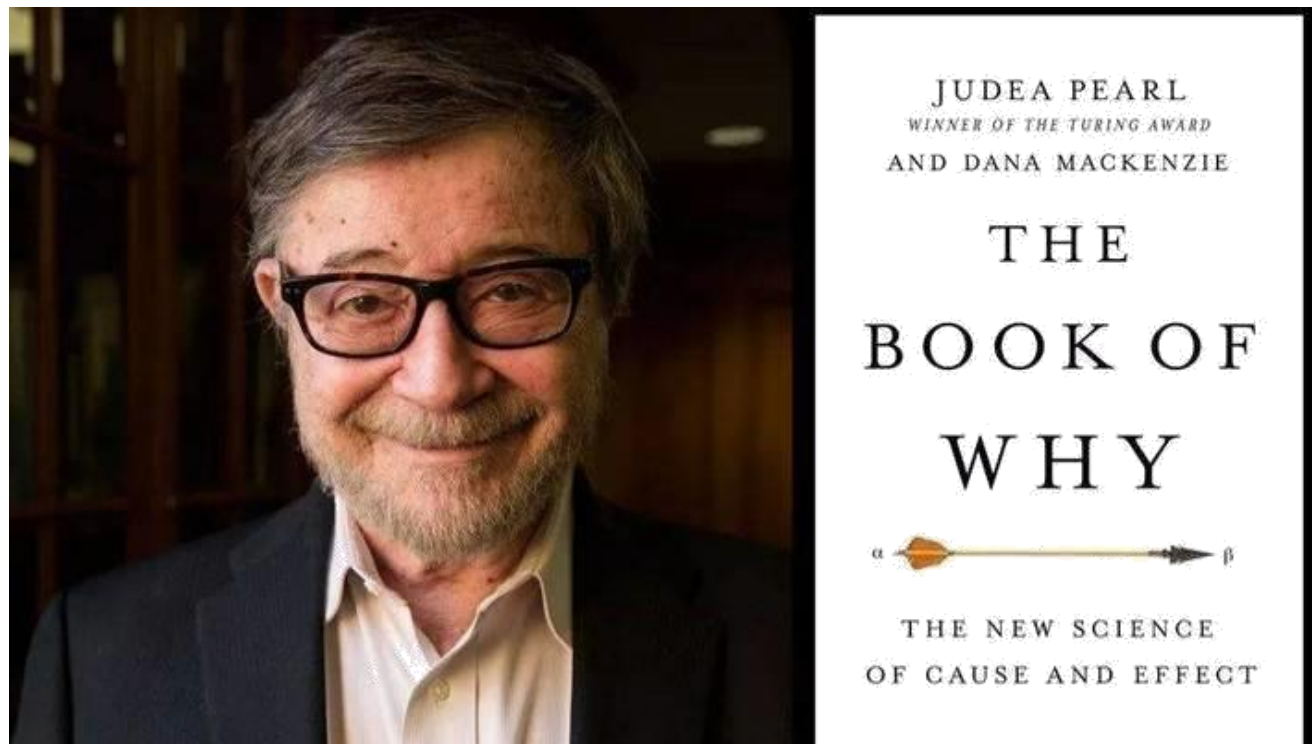
- <https://baike.baidu.com/tashuo/browse/content?id=02c9b57396229adc831a9955>
- 9800 万美元卖掉公司后，他用贝叶斯网络分析数据中的因与果
 - 对事件做出预测相对容易，但分析因果关系则是一件很困难的事。就像路上很多人带着雨伞，代表可能会下雨，但雨伞不是下雨的原因。
 - Via Science公司就希望能通过因果分析技术，帮助电力、能源等公司找到电力系统中的风险所在。
 - 公司的技术核心是因果分析平台 Focus64，它应用了 Judea Pearl 提出的贝叶斯网络，用于理解数据之间的关系。

- 简单来说，贝叶斯网络会引入更多变量来创建一个图（Graph），从而看这些变量之间的相互关系以及如何改变。比如展示雨、雨伞与天色的关系，会看到灰色的天空与下雨有关系，也会看到下雨时没有雨伞这个变量出现，但仍会出现灰色的天空。然后比较两种场景就会发现，灰沉的天空与下雨的关系更强，也就是它出现的原因。
- 贝叶斯网络的优势是用到了图论（Graph Theory），相比于神经网络，其运作过程有透明性，不是黑盒子。
- 贝叶斯网络两个特点：一是能让人类了解到底发生了什么；二是由于知道发生了什么，就可以做出改变。而深度学习网络可以重新训练，但不能改变网络模型。
- 可以在虚拟中模拟可能发生的情况，模拟某个变量可能产生什么样的结果。比如，如果稍微改变电路，某片楼宇中的电力状况会发生什么变化？在现实中不能真的这么做，而在计算机中可以做这种模拟。

题外...

- <https://www.huxiu.com/article/323544.html>

贝叶斯网络发明人珀尔的《为什么》



题外：2021年诺贝尔经济学奖…

- <https://baijiahao.baidu.com/s?id=1713380314221700759>

因果推断：利剑和2021诺贝尔经济学奖三剑客的故事

- ❑ 工资提高会让工人失业吗？
- ❑ 教育会带来更高的回报率吗？



The graphic is a blue rectangular banner for the 2021 Nobel Prize in Economic Sciences. At the top left is a gold Nobel medal. To its right, the text reads "EKONOMIPRISET 2021" and "THE PRIZE IN ECONOMIC SCIENCES 2021". At the top right is the logo of the Royal Swedish Academy of Sciences, "KUNGL. VETENSKAPS-AKADEMIEN". Below this, three portraits of the winners are shown side-by-side. Each portrait has a small vertical photo credit on its left: "Photo: UC Berkeley" for David Card, "Photo: Creative Commons Wiki" for Joshua D. Angrist, and "Photo: Stanford Graduate School of Business" for Guido W. Imbens. Below each portrait is the winner's name and nationality. Underneath the names are their respective contributions in Swedish and English. At the bottom left is the hashtag "#nobelprize". At the bottom right is a circular logo for "THE NOBEL ECONOMICS CIRCLE" with the Chinese text "经济学家圈".

David Card, USA
"för hans empiriska bidrag till arbetsmarknadsekonomi"
"for his empirical contributions to labour economics"
#nobelprize

Joshua D. Angrist, USA
"för deras metodologiska bidrag till analysen av kausala samband"
"for their methodological contributions to the analysis of causal relationships"

Guido W. Imbens, USA

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

THE NOBEL ECONOMICS CIRCLE
经济学家圈

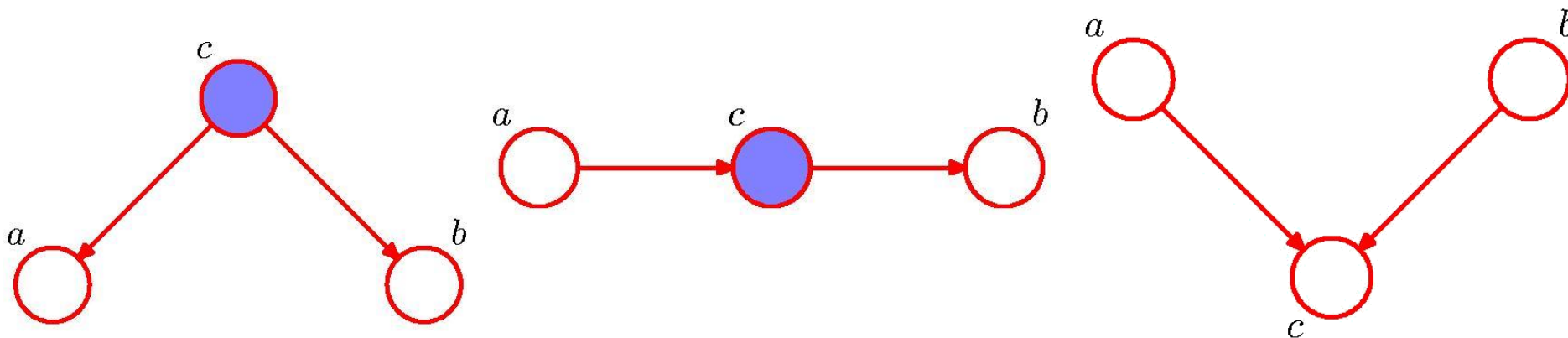
小结

■ 贝叶斯网络

- 用有向边表示单向依赖（条件概率）的有向无环图；概率图模型
- 在一些节点已知的条件下，可对未知节点进行推断。

■ 条件独立性

- 知道（固定）共同原因、知道中间原因、或不知道共同结果，将是A与B相互独立。



2. 朴素贝叶斯... Naive Bayes Classifier...

因为对险恶的现实世界的无知
从而采用了一些天真的假设。

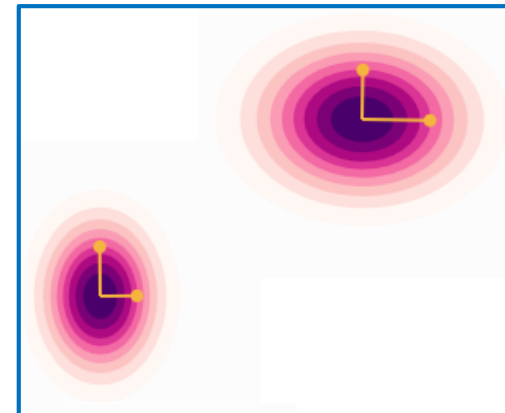
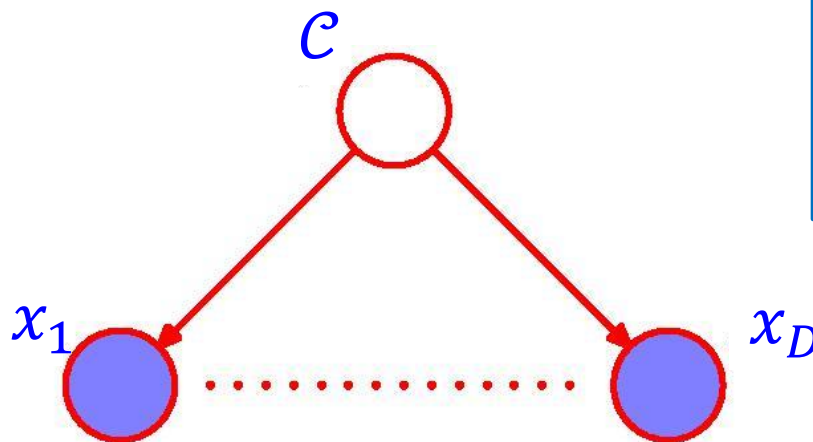
“每次看到朴素这个词，我就仿佛看到贝叶斯穿着一身打满补丁衣服的样子。

naive意思是缺乏经验的、幼稚的、无知的、轻信的。从公式推导过程来看，朴素贝叶斯分类器采用了一些简化条件的假设。

不过，朴素还有一层含义是专一、纯粹，在这个意义上，贝叶斯分类也算大道至简，大智若愚了。”

朴素贝叶斯 (Naive Bayes)

- 把 \mathcal{C} 看做分类, \mathbf{x} 是特征。在 \mathcal{C} 固定时, 特征 x_i 之间是独立的; 但 \mathcal{C} 不固定 (未知) 时, 特征是不独立的。
- 很容易处理高维数据。
 - 这在前一讲中的非参数法中是有麻烦的
- 朴素贝叶斯分类:



$$p(\mathbf{x}|\mathcal{C}) = \prod_{i=1}^D p(x_i|\mathcal{C})$$

$$p(\mathbf{x}) = \sum_{\mathcal{C}} \prod_{i=1}^D p(x_i|\mathcal{C}) p(\mathcal{C}) \neq \prod_{i=1}^D p(x_i)$$

$$p(\mathcal{C}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C})p(\mathcal{C})}{p(\mathbf{x})} = \frac{p(\mathcal{C}) \prod_{i=1}^D p(x_i|\mathcal{C})}{p(\mathbf{x})}$$

参数估计...

- 类别先验概率：用频率估算

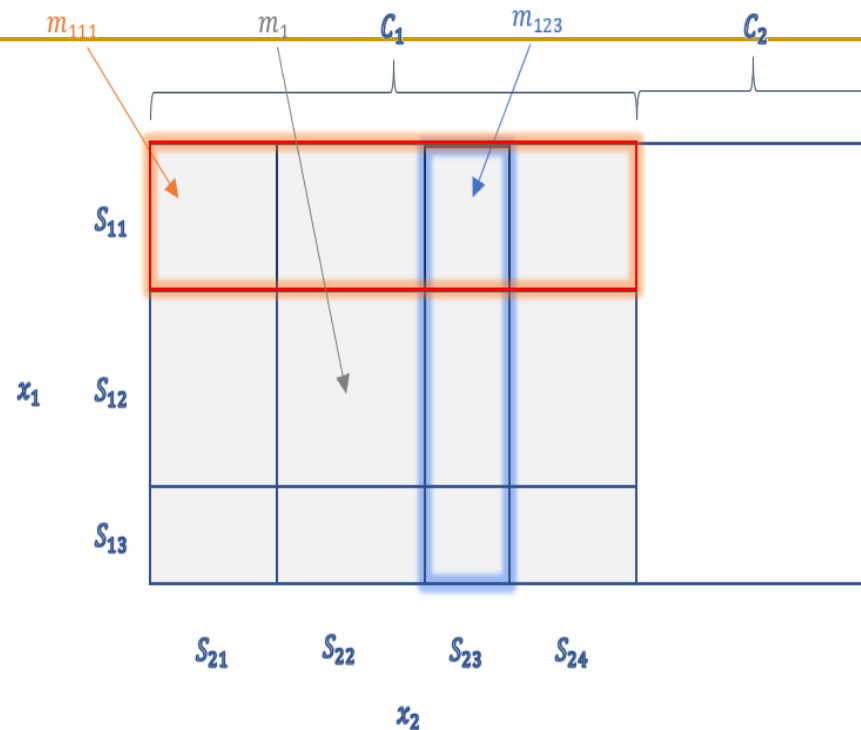
$$p(\mathcal{C}_k) = \frac{N_k}{N}$$

- 特征 x_i （假设有 s_i 个可能的分立值）的条件概率可用频率估算：

$$p(x_{is}|\mathcal{C}_k) = \frac{N_{kis}}{N_k}$$

或者，为了防止 $N_{kis} = 0$ 所导致的 $p(\mathbf{x}|\mathcal{C}_k) = 0$ ，采用拉普拉斯平滑：

$$p(x_{is}|\mathcal{C}_k) = \frac{N_{kis} + 1}{N_k + s_i}$$



例子：根据天气情况决定是否打网球

- 特征取值组合数：36
- (sunny, cool, high, strong), 会打网球吗？

■ 打球：不打球=

$$\frac{9}{14} \left(\frac{2+1}{9+3} \right) \left(\frac{3+1}{9+3} \right) \left(\frac{3+1}{9+2} \right) \left(\frac{3+1}{9+2} \right) :$$

$$\frac{5}{14} \left(\frac{3+1}{5+3} \right) \left(\frac{1+1}{5+3} \right) \left(\frac{4+1}{5+2} \right) \left(\frac{3+1}{5+2} \right)$$

$$= 0.007084: 0.01822 =$$

$$28\%: 72\%$$

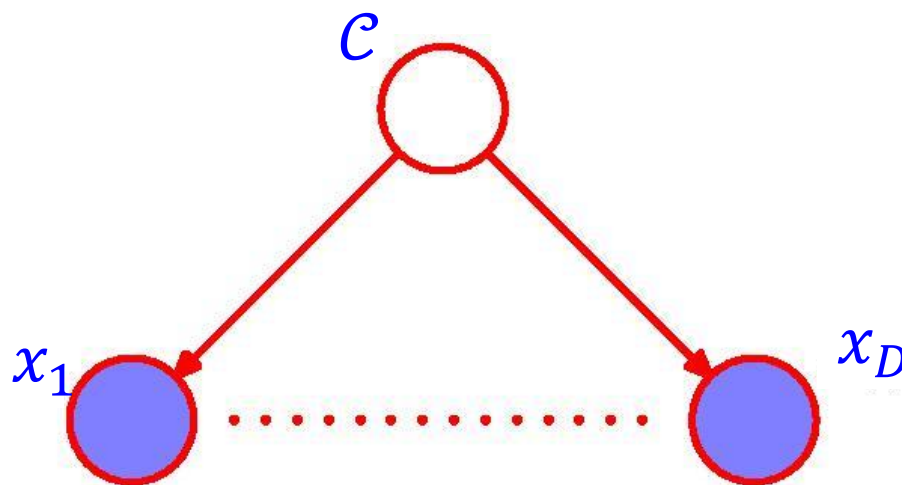
	天气	温度	湿度	风速	
Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

小结

- 朴素贝叶斯模型

- 假设任一类别 \mathcal{C} 的特征 x_i 之间是独立的；

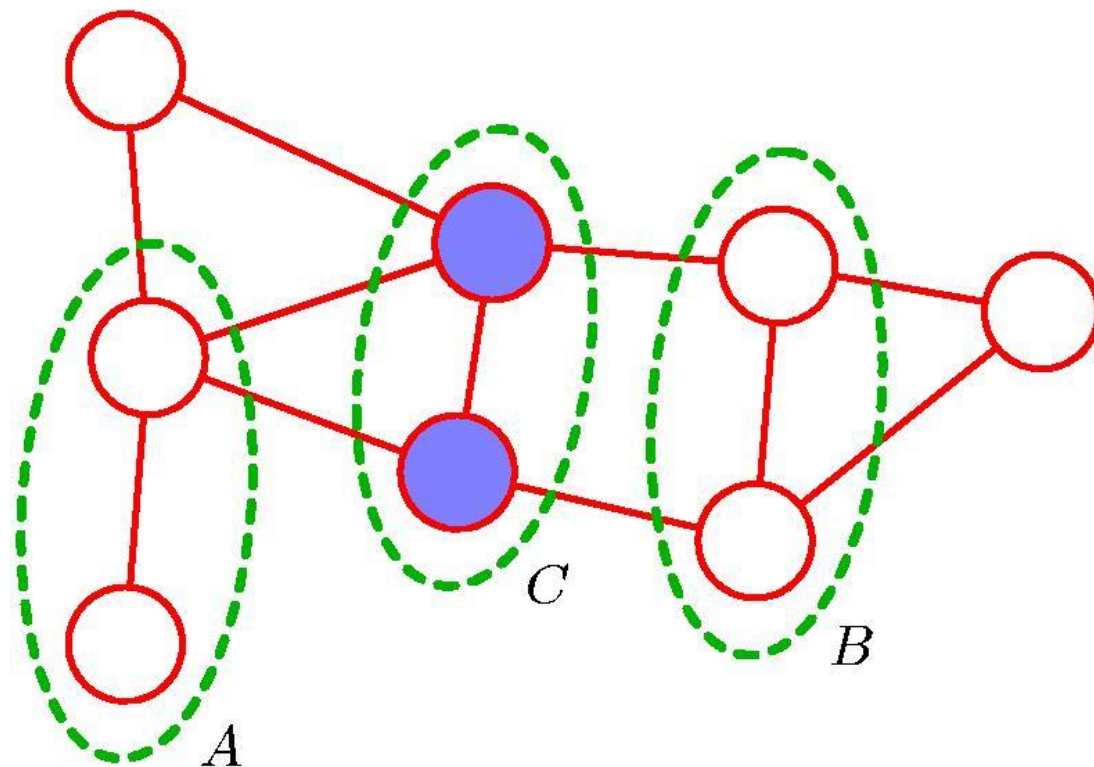
$$p(\mathbf{x}|\mathcal{C}) = \prod_{i=1}^D p(x_i|\mathcal{C})$$



3. 马尔科夫随机场与玻尔兹曼机...

马尔科夫网络

- 又称马尔科夫随机场
(Markov random field)
- 无向图。



- 如果从 A 到 B 的任一条路径都会通过 C ，则称 A 与 B 是被 C 所分离，我们要求其上的所有变量满足

$$A \perp\!\!\!\perp B | C$$

普适表达式（参见Bishop 8.3.2）

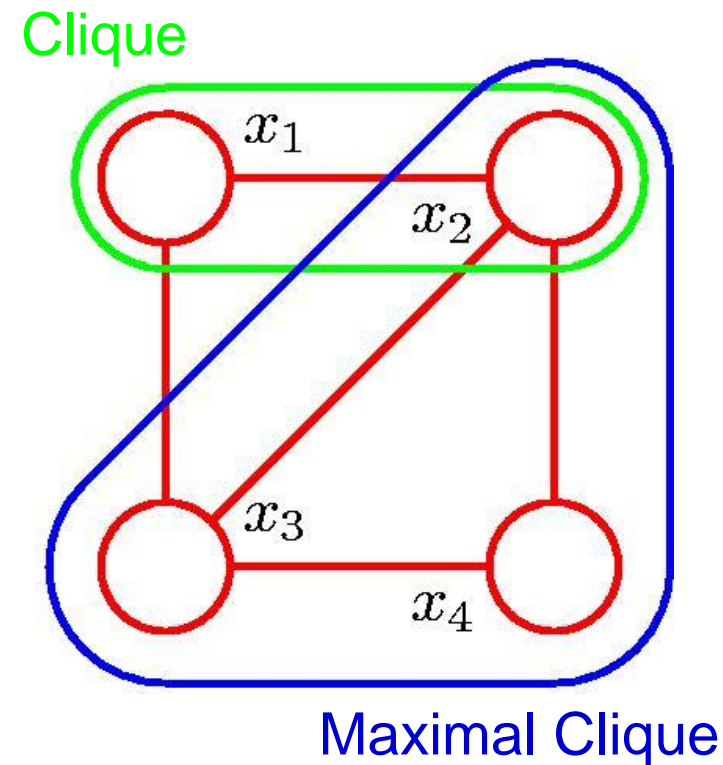
- 全连接的节点集合叫团（clique）
- 联合分布可以写成所有极大clique的乘积：

$$p(\mathbf{x}) = \frac{1}{Z} \prod_c \psi_c(\mathbf{x}_c)$$

- 其中 Z 是归一化常数（配分函数）。
- $\psi_c(\mathbf{x}_c)$ 是非负的，因此常写成类似于玻尔兹曼分布的形式：

$$\psi_c(\mathbf{x}_c) = \exp[-E(\mathbf{x}_c)]$$

- $\psi_c(\mathbf{x}_c)$ 不能直接解释成概率。



玻尔兹曼机 (Boltzmann machine)

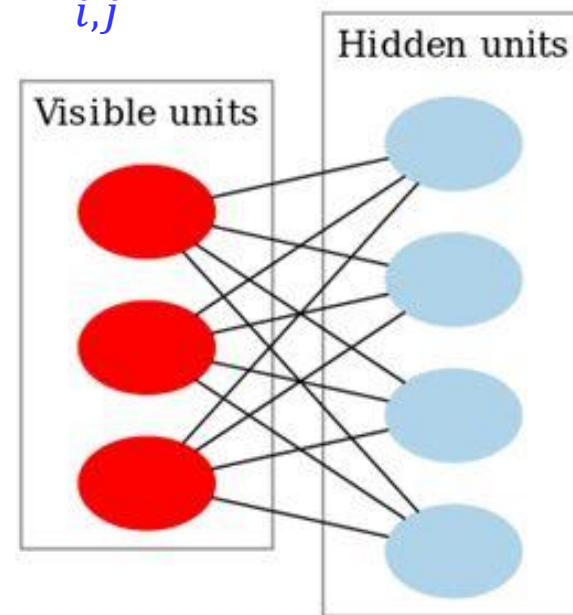
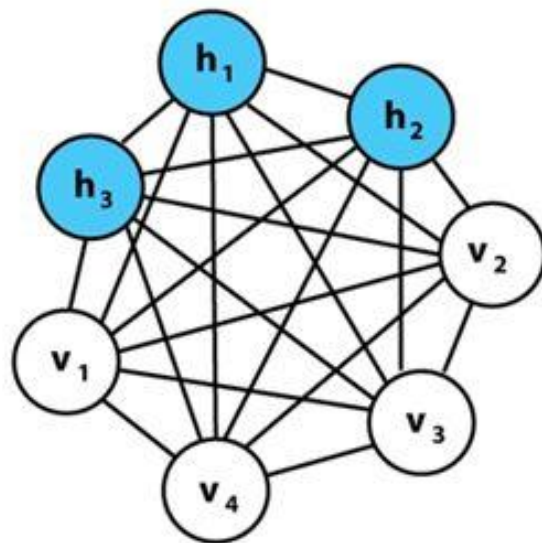
$$E(\mathbf{x}) = - \sum_i b_i x_i - \sum_{i,j} w_{ij} x_i x_j$$

- 玻尔兹曼机可以用来解决两类问题。
 - 搜索问题。当给定变量之间的连接权重，需要找到一组二值向量，使得整个网络的能量最低。
 - 学习问题。当给定一组部分变量的观测值时，计算一组最优的权重。
- 包含隐藏层的模型：
 - 当Boltzmann Machines网络中既包含可被观察到的可见神经元，又包含不可见神经元的时候，网络可以表达更加复杂的模式。

受限玻尔兹曼机 (RBM, Restricted Boltzmann machine)

- 可以用于降维、分类、回归、协同过滤、特征学习以及主题建模的算法。
 - 只在可见层与隐藏层之间有连接：

$$E(\mathbf{v}, \mathbf{h}) = - \sum_i a v_i - \sum_j b_j h_j - \sum_{i,j} w_{ij} v_i h_j$$



玻尔兹曼机(左)与受限玻尔兹曼机(右)的区别

小结

■ 马尔科夫网络

- 无向图
- 联合分布可以写成所有极大clique的乘积

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C), \quad \psi_C(\mathbf{x}_C) = \exp[-E(\mathbf{x}_C)]$$

■ 玻尔兹曼机: $E(\mathbf{x}) = -\sum_i b_i x_i - \sum_{i,j} w_{ij} x_i x_j$

■ 受限玻尔兹曼机

- 只在可见层与隐藏层之间有连接
- $E(\mathbf{v}, \mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} w_{ij} v_i h_j$

应用例子：

Lorillee Tallorin, JiaLei Wang, Woojoo E. Kim, Swagat Sahu, Nicolas M. Kosa, Pu Yang, Matthew Thompson, Michael K. Gilson, Peter I. Frazier, Michael D. Burkart & Nathan C. Gianneschi.

Discovering de novo peptide substrates for enzymes using machine learning.

Nat. Commun. **9**, 5253 (2018).

学习资料： [nc5253.pdf](#), [nc5253s.pdf](#)

问题描述

- 4'-phosphopantetheinyl transferase (PPTase)（磷酸泛酰巯基乙胺基转移酶，简称**磷酰转移酶**）的多肽底物？
- 初始数据：天然蛋白底物的多肽片段，以及已知的其它筛选多肽底物
- 预测：以Ser（丝氨酸）为中央的多肽底物，不多于38个残基，越短越好。

- 为减少过拟合的风险，将残基种类分成8大类（聚类）。

CLASS 1	D E (acidic)		
CLASS 2	N Q (amidic)		
CLASS 3	F W Y (aromatic)		
CLASS 4	H K R (basic)		
CLASS 5	A I L M V (nonpolar)		
CLASS 6	G P (aliphatic)		
CLASS 7	S T (hydroxylic)		
CLASS 8	C (sulfur)		

模型与方法

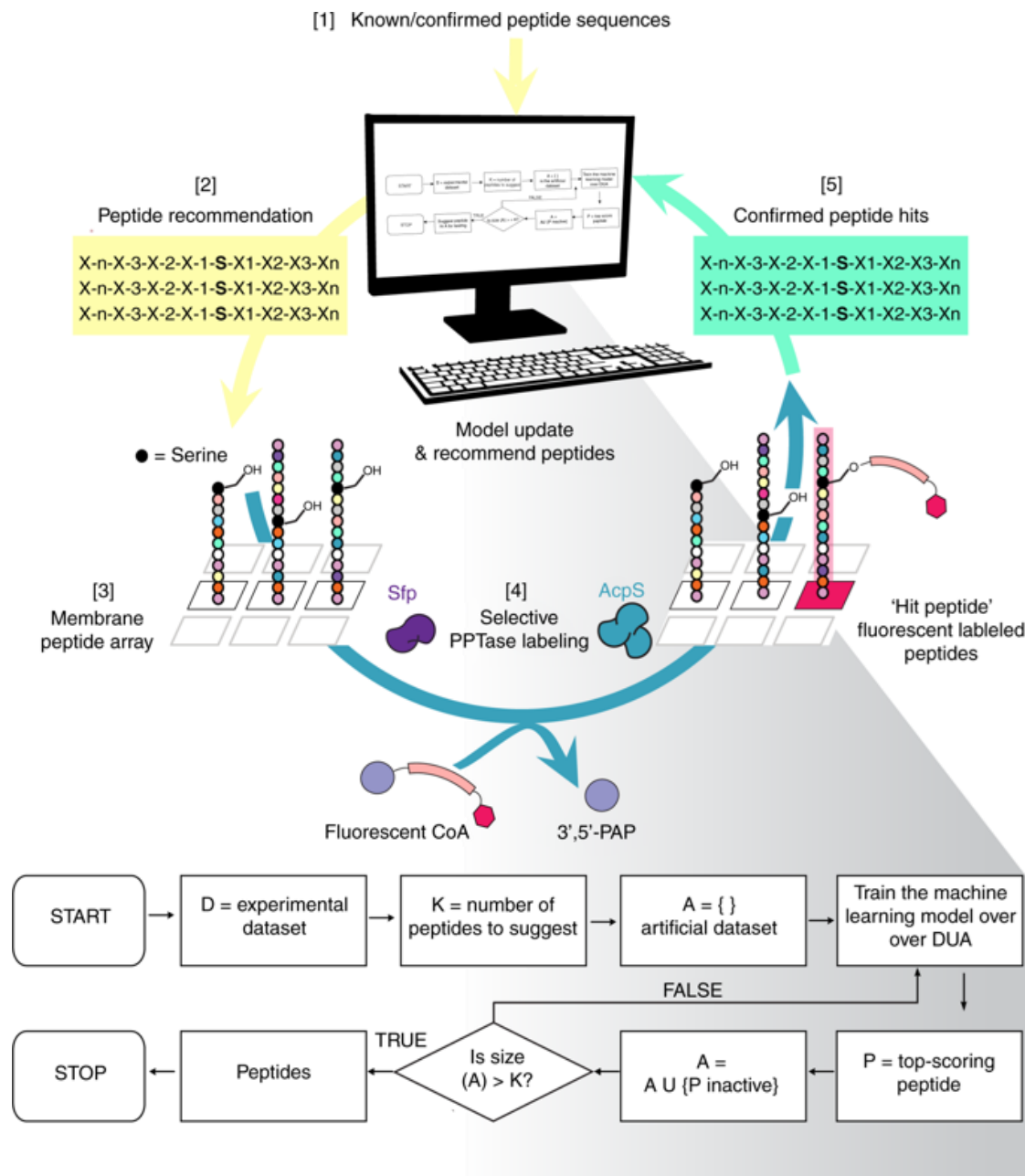
- Naive Bayes classifier

- 参数先验概率采用Dirichlet（狄利克雷）分布（参见Bishop 2.2.1）；
为反映离Ser越远的残基影响越小的性质，

1 We further put a Dirichlet prior over $\theta^{y',j}$, that is, $\theta^{y',j} \sim \text{Dirichlet}(\bar{\alpha}^{y,j})$. We set $\bar{\alpha}^{y,j}$ to be a vector that grows with the square root as we move further from the conversed serine:
 $\alpha_i^{y,j} = \alpha_i^{y,1} \sqrt{|j|}.$

- 对后验概率做平均，以进一步提高准确性

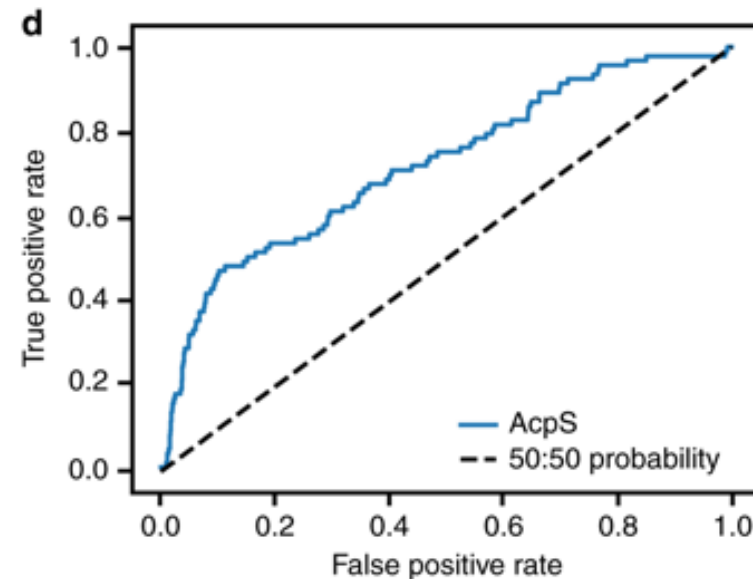
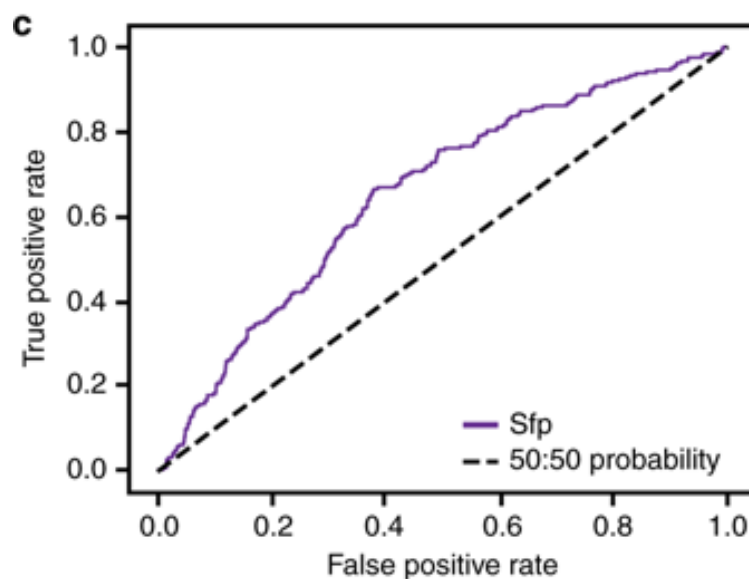
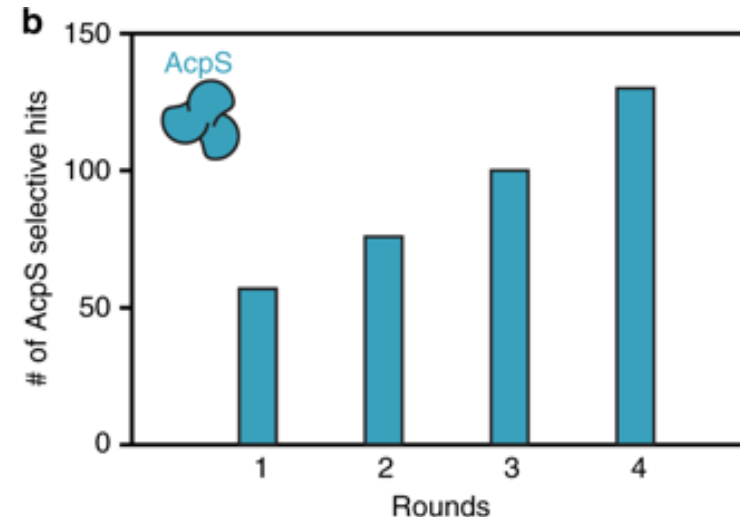
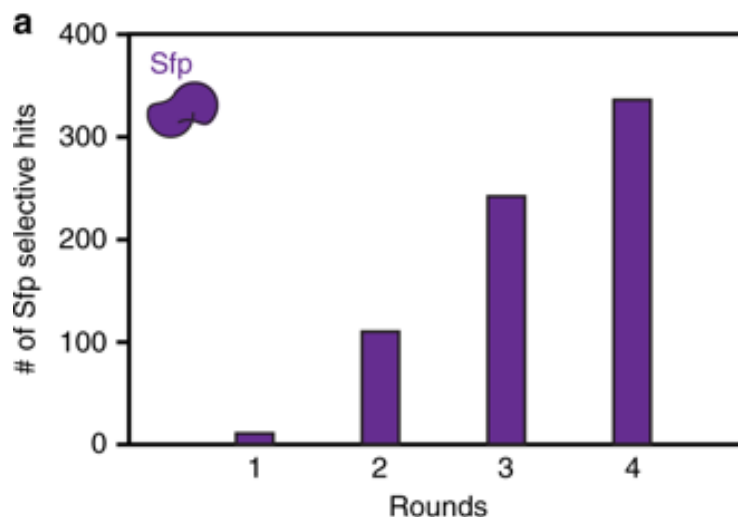
流程



预测一个最佳多肽后，把它作为负样本加入训练集重新训练并预测第二个多肽，以实现正交性（增加预测序列的多样性）。

结果

- Specific peptide hits identified by POOL and model validation.
- 每轮约含600种多肽
- （新闻稿）：命中
率高达**30%**



Scikit-Learn相关内容

<https://scikit-learn.org/>

<https://sklearn.apachecn.org/>

■ 1.9. Naive Bayes

- ❑ `naive_bayes.GaussianNB`
- ❑ `naive_bayes.MultinomialNB`
- ❑ `naive_bayes.ComplementNB`
- ❑ `naive_bayes.BernoulliNB`

■ Reference:

- Bishop 8;
- Elements 17。
- 实战 4。

■ 扩展阅读：

□ <https://baike.baidu.com/tashuo/browse/content?id=02c9b57396229adc831a9955>

9800 万美元卖掉公司后，他用贝叶斯网络分析数据中的因与果.mht

□ <https://www.jiqizhixin.com/articles/2018-10-26-6>

迷人又诡异的辛普森悖论：同一个数据集是如何证明两个完全相反的观点的？.mht

□ <https://zhuanlan.zhihu.com/p/74586507>

机器学习中的判别式模型和生成式模型.pdf

□ <https://baijiahao.baidu.com/s?id=1713380314221700759>

因果推断：利剑和2021诺贝尔经济学奖三剑客的故事.mht

❑ <https://finance.sina.cn/2024-07-14/detail-inceanfx2867140.d.html>

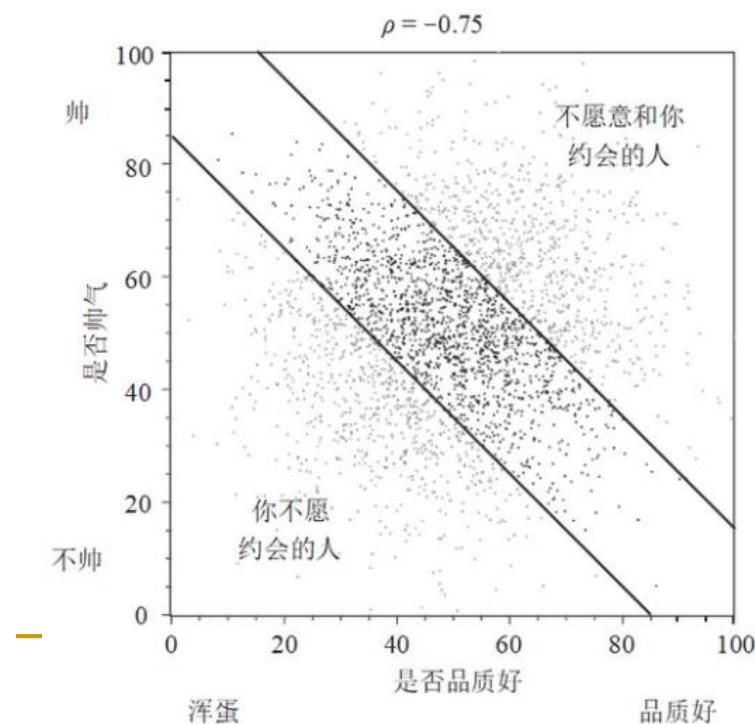
6700万参数比肩万亿巨兽GPT-4！微软MIT等联手破解Transformer推理密码.mhtml

- 通过因果模型构建数据集，直接教模型学习公理，结果只有67M参数的微型Transformer竟能媲美GPT-4的推理能力。

❑ <https://www.huxiu.com/article/508964.html>

帅哥和顶级程序员.pdf

- 谷歌的工程师将机器学习技术应用到他们自己的招聘过程中，试图找出最高效的员工。他们发现了一个令人惊讶的结果：**工作表现与之前在编程竞赛中的成绩呈负相关。**
- 在约会场所非常活跃的朋友们有时抱怨说，当和帅哥出去约会时，会发现这个人原来是个浑蛋；而当感觉某个人品质很好时，却发现这个人长得不帅。
- 这是怎么回事？



图计算的浪漫：图解《长安三万里》.mhtml



谢谢大家!