中山大学AI技术俱乐部
《机器学习》课程

7. 贝叶斯分类器

余伟浩
whyu@whyu.me

AIClub
Sun Yat-sen University

HCP
Human Cyber Physical Intelligence Integration Lab @ SYSU
中山大学人机物智能融合实验室
Our laboratory seeks general and task-driven ways to build intelligence, embracing the access to the ternary human-cyber-physical universe.

参考课本：周志华. (2016). *机器学习*. 清华大学出版社.

何为分类器

X为输入，c为类别（class）

对于各个模型，明确输入输出才不会乱

思路

- 7. 贝叶斯分类器
  - 区别：
    - 贝叶斯分类器：通过最大后验概率进行**单点估计**
    - 贝叶斯学习：通过最大后验概率进行**分布估计**
  - 7.0 入门知识（补充）
    - 贝叶斯定理
    - 相互独立
    - 判别式与生成式模型
    - 期望
  - 7.1 贝叶斯决策论
  - 7.2 极大似然估计（不讲）
  - 7.3 朴素贝叶斯分类器
  - 7.4 半朴素贝叶斯分类器
  - 7.5 贝叶斯网络

7.0 入门知识（补充）

已知：

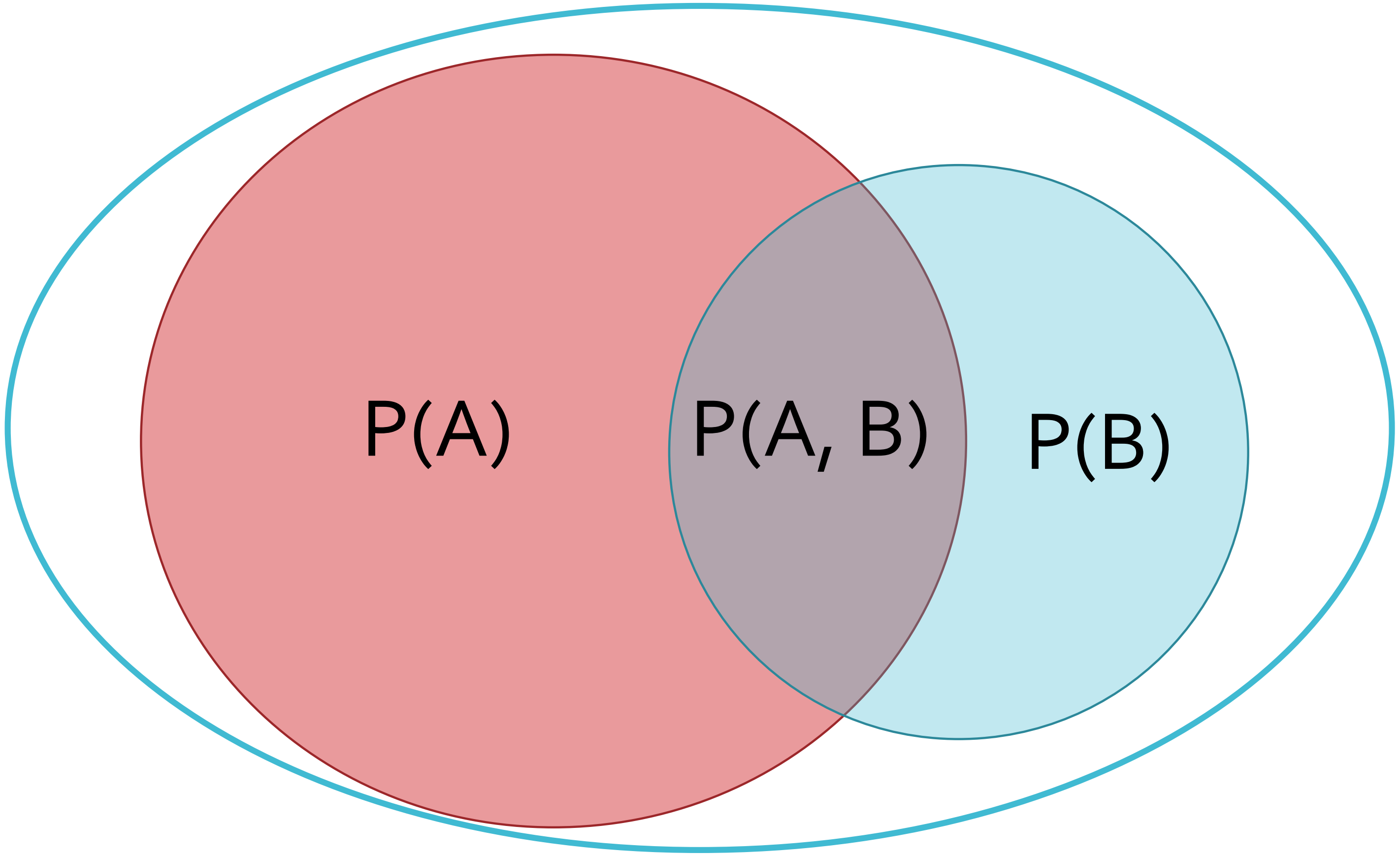
$P(A) = 1/2$

$P(B|A) = 1/3$

$P(B) = 1/3$

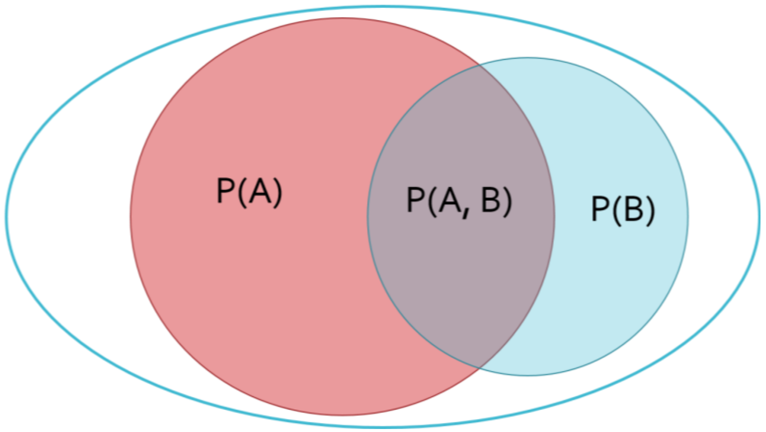
$P(A|B) = 1/2$

求$P(A, B)$？



贝叶斯定理

- P(A, B) = P(A) × P(B|A)
- P(A, B) = P(B) × P(A|B)
- 又因为P(A, B) = P(A, B)
- 所以：
- P(A) × P(B|A) = P(B) × P(A|B)
- 所以 P(A) , P(B|A), P(B), P(A|B) 四个值，可以知三求一
- 比如要求P(B|A) , 只需移一下项便可得
- $P(B|A) = \frac{P(B)\times P(A|B)}{P(A)}$
- 上面这个式子便是**贝叶斯定理**



贝叶斯定理

- 应用题：
- 上海交大武筱林教授的论文《基于面部图像的自动犯罪性概率推断》（Automated Inference on Criminality using Face Images)

C/N → 模型 → +/-

实际犯罪/无犯罪 → 模型 → 检测犯罪/无犯罪

模型性能

C → + 89%

C → - 11%

N → + 7%

N → - 93%

已知中国的犯罪率是0.36%，请问当被模型检测出犯罪，实际有罪犯的概率有多大？

←左边还有

N=0.9964

S

←左边还有

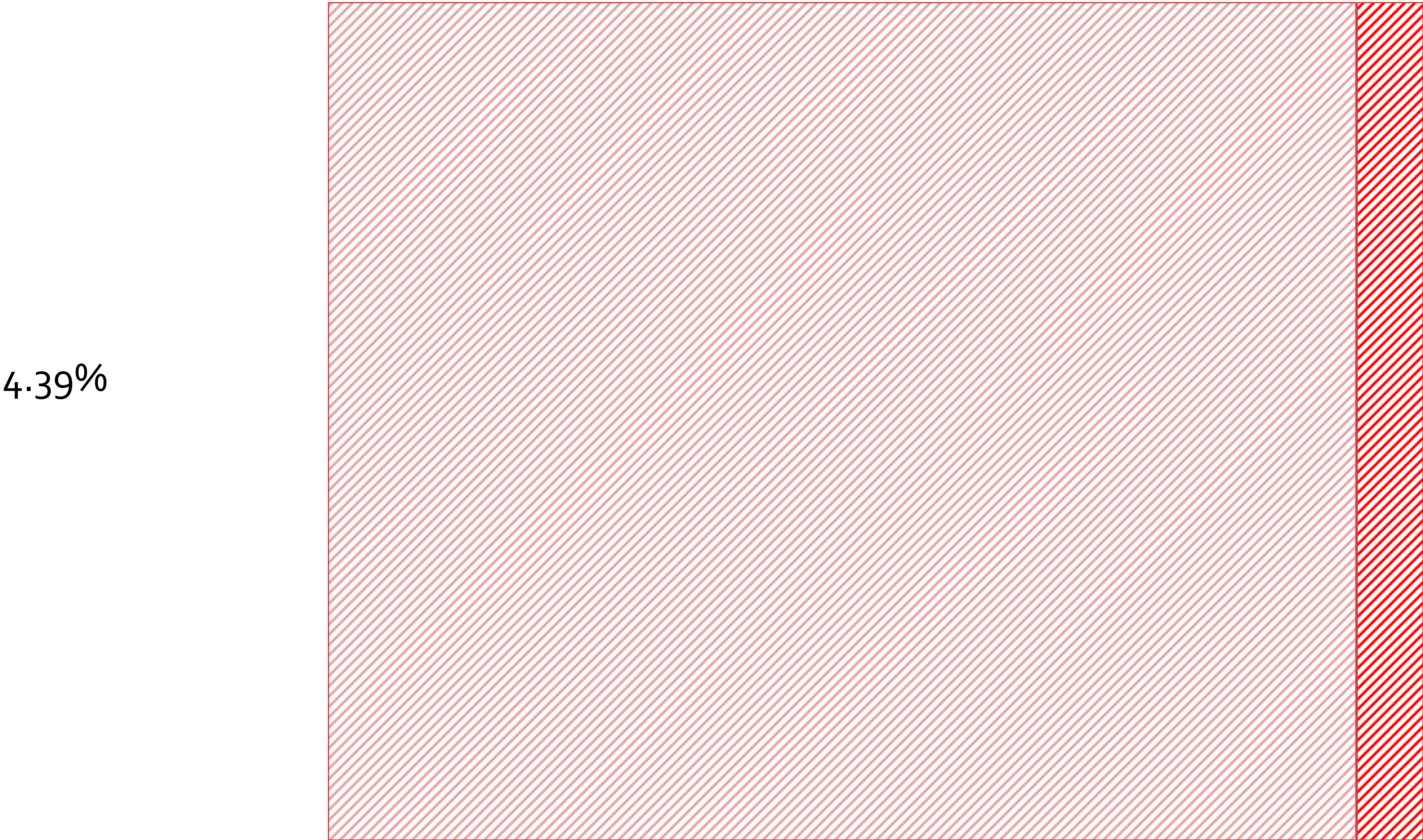
N=0.9964

←左边还有

N=0.9964

←左边还有

N=0.9964


4·39%

已知：

$P(A) = 1/2$

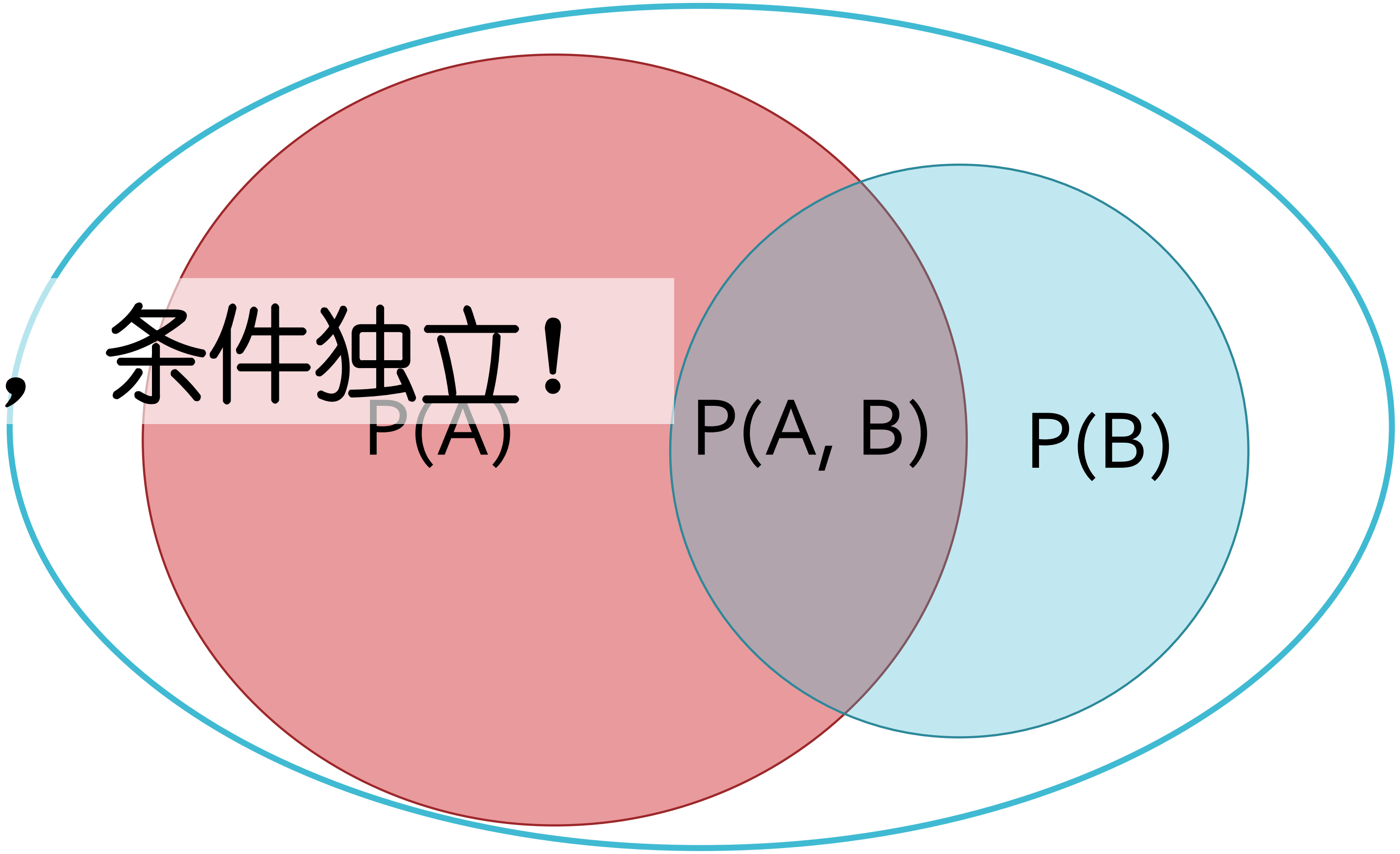
$P(B|A) = 1/3$

$P(B) = 1/3$

$P(A|B) = 1/2$

相等，条件独立！

求$P(A, B)$？



已知：

$P(A) = 1/2$

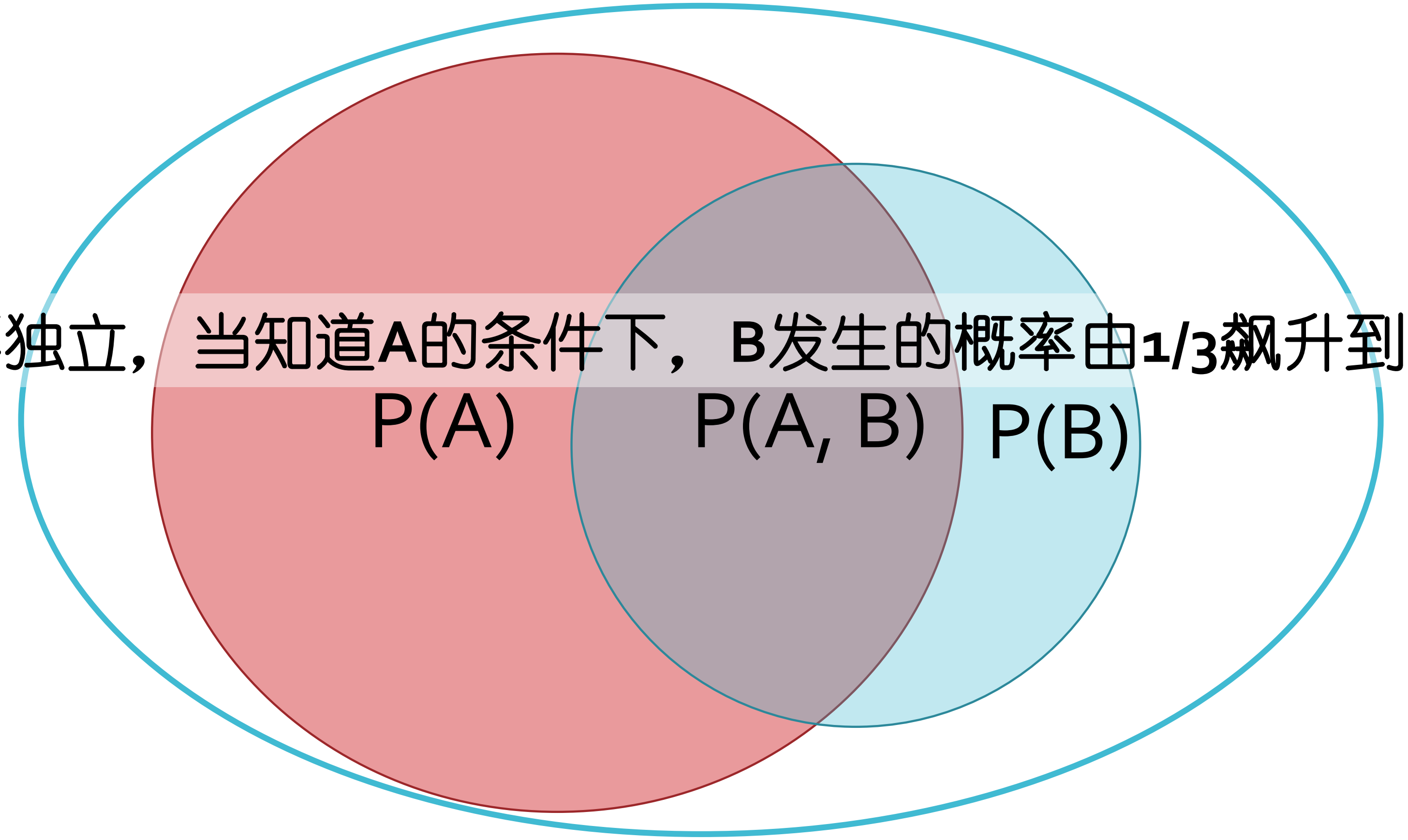
$P(B|A) = 1/2$

$P(B) = 1/3$

$P(A|B) = 3/4$

求$P(A, B)$？

条件不独立，当知道A的条件下，B发生的概率由1/3飙升到1/2

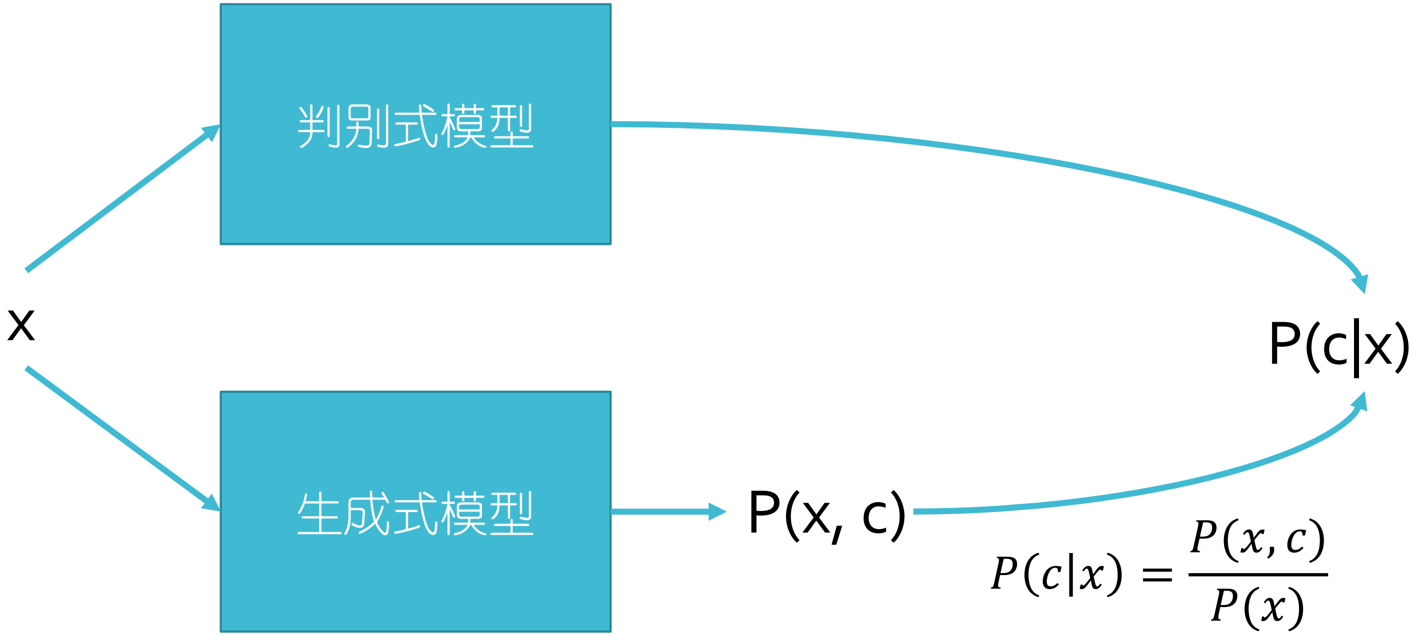


相互独立

- $P(A, B) = P(A) \times P(B|A) = P(A) \times P(B)$
- $P(A, B) = P(B) \times P(B|A) = P(B) \times P(A)$
- 所以A，B相互独立的条件为
- $P(A, B) = P(A) \times P(B)$
- 推广到多维
- $P(A, B, C, D \ldots\ldots) = P(A) \times P(B) \times P(C) \times P(D) \times \ldots\ldots$
- 推广到条件概率
- $P(A, B, C, D \ldots\ldots|Y) = P(A|Y) \times P(B|Y) \times P(C|Y) \times P(D|Y) \times \ldots\ldots$

判别式与生成式模型

X是输入，c是标签



$$P(c|x) = \frac{P(x,c)}{P(x)}$$

总结：在分类任务中，
判别式模型：一步到位，直接输出P(c|x), 模型中间运算不会出现P(x, c)
生成式模型：两步走，先算出P(x, c)，再由概率公式算出P(c|x)

期望

- 例子：
- 买一张彩票金额为1元，中奖奖金1000元，中奖率万分之一，请问选择买一张彩票得到的奖金期望是多少？不买彩票得到的奖金期望是多少？

期望

- 例子：
- 买一张彩票金额为1元，中奖奖金1000元，中奖率万分之一，请问选择买一张彩票得到的奖金期望是多少？不买彩票得到的奖金期望是多少？
- E(买彩票) = -1 * 0.9999 + （1000 – 1） * 0.0001 = -0.9
- E(不买彩票） = 0 * 0.9999 + 0 * 0.0001 = 0

期望

- 例子：
- 买一张彩票金额为1元，中奖奖金1000元，中奖率万分之一，请问选择买一张彩票得到的奖金期望是多少？不买彩票得到的奖金期望是多少？
- E(买彩票) = -1 * 0.9999 + （1000 – 1） * 0.0001 = -0.9
- E(不买彩票） = 0 * 0.9999 + 0 * 0.0001 = 0
- 转换成矩阵预算

奖励矩阵

$$\begin{array}{cc} & \begin{array}{cc}\text{不中奖} & \text{中奖}\end{array} \\ \begin{array}{c}\text{买} \\ \text{不买}\end{array} & \begin{bmatrix} -1 & 999 \\ 0 & 0 \end{bmatrix} \end{array}$$

期望

- 例子：
- 买一张彩票金额为1元，中奖奖金1000元，中奖率万分之一，请问选择买一张彩票得到的奖金期望是多少？不买彩票得到的奖金期望是多少？
- E(买彩票) = -1 * 0.9999 + （1000 – 1） * 0.0001 = -0.9
- E(不买彩票） = 0 * 0.9999 + 0 * 0.0001 = 0
- 转换成矩阵预算

奖励矩阵

$$\begin{array}{cc} & \begin{array}{cc} 不中奖 & 中奖 \end{array} \\ \begin{array}{c} 买 \\ 不买 \end{array} & \begin{bmatrix} -1 & 999 \\ 0 & 0 \end{bmatrix} \end{array}$$

概率矩阵

$$\begin{array}{cc} \begin{array}{c} 不中奖 \\ 中奖 \end{array} & \begin{bmatrix} 0.9999 \\ 0.0001 \end{bmatrix} \end{array}$$

期望

$$\begin{bmatrix} E(买) & \\ E(不买) & = \end{bmatrix} = \begin{bmatrix} -1 & 999 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.9999 \\ 0.0001 \end{bmatrix} = \begin{bmatrix} -0.9 \\ 0 \end{bmatrix}$$

7.1 贝叶斯决策论

何为贝叶斯决策理论？

- 天气预报明天
- 30%的概率下雨
- 60%的概率阴天
- 10%的概率晴天

- 我要选择明天天气情况（决定了我要不要带雨伞）
-

何为贝叶斯决策论？

- 天气预报明天
- 30%的概率下雨
- 60%的概率阴天
- 10%的概率晴天

- 我要选择明天天气情况（决定了我要不要带雨伞）
- 贝叶斯理决策理论：**给定概率做决策**

贝叶斯决策论

输入

输出

P1 = XX %
P2 = XX %
P3 = XX %
P4 = XX %
……

$\sum_n P_n = 1$

贝叶斯决策论

选择

贝叶斯决策论

0代表猫，1代表狗

$p(c=0|x) = 0.9$

$P(c=1|x) = 0.1$

基于概率的决策模型

$c = ?$

x

假如这个决策模型就是你的大脑，给定了以上概率，你是选择“猫”还是“狗”？

贝叶斯决策论

假设我们去医院做检查
拍了B超照片后B超机器给出了以下概率

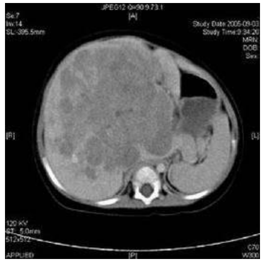
0代表正常，1代表癌症

$p(c=0|x) = 0.9$

$P(c=1|x) = 0.1$

基于概率的决策模型

c = ?



x

我们是选择c=0, 回家洗洗睡
还是选择c=1，请人类医生再进一步检查呢
对比上一页ppt的猫狗问题，同样0.9， 0.1的概率分布，选择却不一样？怎样对不同情况进行建模？

贝叶斯决策论

要计算损失期望，有必要引入损失矩阵

对于猫狗问题

$$\begin{array}{cc} \text{选择}\backslash\text{实际} & \begin{matrix}\text{猫} & \text{狗}\end{matrix} \\ \begin{matrix}\text{猫}\\ \text{狗}\end{matrix} & \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\end{array}$$

选择猫的损失期望值为

$R(0|x) = 0 \times P(0|x) + 1 \times P(1|x) = 0 \times 0.9 + 1 \times 0.1 = 0.1$

选择狗的损失期望值为

$R(1|x) = 1 \times P(0|x) + 0 \times P(1|x) = 1 \times 0.9 + 0 \times 0.1 = 0.9$

对于癌症问题

$$\begin{array}{cc} \text{选择}\backslash\text{实际} & \begin{matrix}\text{无癌症} & \text{癌症}\end{matrix} \\ \begin{matrix}\text{无癌症}\\ \text{癌症}\end{matrix} & \begin{bmatrix}0 & 10000\\ 1 & 0\end{bmatrix}\end{array}$$

选择无癌症的损失期望值为

$R(0|x) = 0 \times P(0|x) + 10000 \times P(1|x) = 0 \times 0.9 + 10000 \times 0.1 = 1000$

选择癌症的损失期望值为

$R(1|x) = 1 \times P(0|x) + 0 \times P(1|x) = 1 \times 0.9 + 0 \times 0.1 = 0.9$

贝叶斯决策论

对于癌症问题

$$\begin{array}{c|cc} \text{选择} \backslash \text{实际} & \text{无癌症} & \text{癌症} \\ \hline \text{无癌症} & 0 & 10000 \\ \text{癌症} & 1 & 0 \end{array}$$

多样本

x^1：
- P(y=0|x^1) = 0.9
- p(y=1|x^1) = 0.1

x^2：
- P(y=0|x^2) = 0.99
- p(y=1|x^2) = 0.01

X^3：
- P(y=0|x^3) = 0.9999
- p(y=1|x^3) = 0.0001

X^4：
- P(y=0|x^4) = 0.99999
- p(y=1|x^4) = 0.00001

找判定标准X→Y，使得对所有样本，总体损失最小

R总 = R^1 + R^2 + R^3 + R^4

→ 找判定标准X→Y，使得每个个体的损失函数最小

→ $$h^*(x) = \arg\min_{c \in y} R(c|x)$$

贝叶斯决策论

幸运的是，在分类任务中，我们一般遇到的是猫狗问题，
损失矩阵为1-I，其中I为单位矩阵

对于猫狗问题 $1-\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$

对于猫狗兔问题 $1-\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}=\begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\\1 & 1 & 0\end{bmatrix}$

此时条件风险R(y=c|x) = 1 – P(y=c|x)，比如猫狗兔问题

X（图片）
$$\begin{cases}P(y=猫|x) = 0.7\\p(y=狗|x) = 0.2\\p(y=兔|x) = 0.1\end{cases} \rightarrow \begin{bmatrix}0.7\\0.2\\0.3\end{bmatrix} \rightarrow R=\begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\\1 & 1 & 0\end{bmatrix}\begin{bmatrix}0.7\\0.2\\0.3\end{bmatrix}=\begin{bmatrix}0.3\\0.8\\0.1\end{bmatrix}$$

得 R(y=猫|x) = 0.3, R(y=狗|x) = 0.8, R(y=兔|x)=0.9
所以为了最小化损失期望，应该选择y=猫

$$h^*(x)=\underset{c\in y}{\arg\min}\, R(c|x)=\underset{c\in y}{\arg\min}(1-\mathrm{P}(c|x))=\underset{c\in y}{\arg\,max}\,\mathrm{P}(c|x)$$

所以，结论是为了最小化损失期望，选择那个概率最大的。
其实我们一开始就是这么想的。。。这就是贝叶斯决策论！

贝叶斯决策论总结

输入

输出

P1 = XX %
P2 = XX %
P3 = XX %
P4 = XX %
……

$\sum_n P_n = 1$

贝叶斯决策理论

选择

根据损失矩阵计算损失期望值，选择使得损失期望值最小的那个类

说白了就是算期望，再比大小

当损失矩阵为1-I时，选择那个概率最大的类

这种情况连期望都不用算了，直接比大小

贝叶斯决策论总结

这些概率要怎么得到？

输入

$P_1 = XX\%$
$P_2 = XX\%$
$P_3 = XX\%$
$P_4 = XX\%$
……

$\sum_n P_n = 1$

贝叶斯决策理论

输出

选择

根据损失矩阵计算损失期望值，选择使得损失期望值最小的那个类

说白了就是算期望，再比大小

当损失矩阵为1-l时，选择那个概率最大的类

这种情况连期望都不用算了，直接比大小

贝叶斯决策论总结

这些概率要怎么得到？

答：利用**概率模型**

输入

P1 = XX %
P2 = XX %
P3 = XX %
P4 = XX %
……

$\sum_n P_n = 1$

贝叶斯决策理论

输出

选择

根据损失矩阵计算损失期望值，选择使得损失期望值最小的那个类

说白了就是算期望，再比大小

当损失矩阵为1-l时，选择那个概率最大的类

这种情况连期望都不用算了，直接比大小

何为概率模型？

何为概率模型？

x → 概率模型 →

$P_1 = XX\ \%$

$P_2 = XX\ \%$

$P_3 = XX\ \%$

$P_4 = XX\ \%$

……

$$\sum_n P_n = 1$$

何为概率模型？

X → 概率模型 →

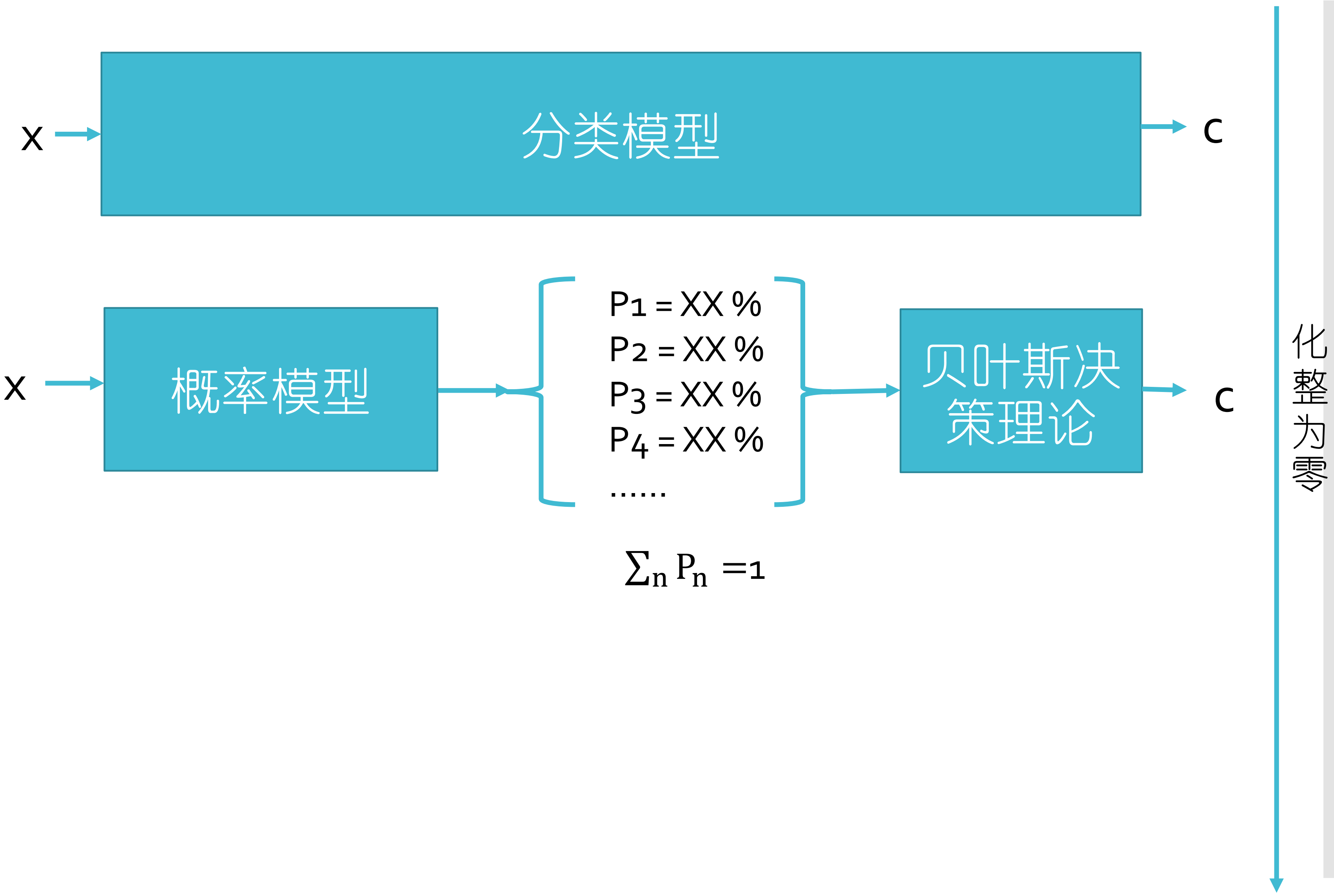
- $P_1 = XX\,\%$
- $P_2 = XX\,\%$
- $P_3 = XX\,\%$
- $P_4 = XX\,\%$
- ……

$\sum_n P_n = 1$

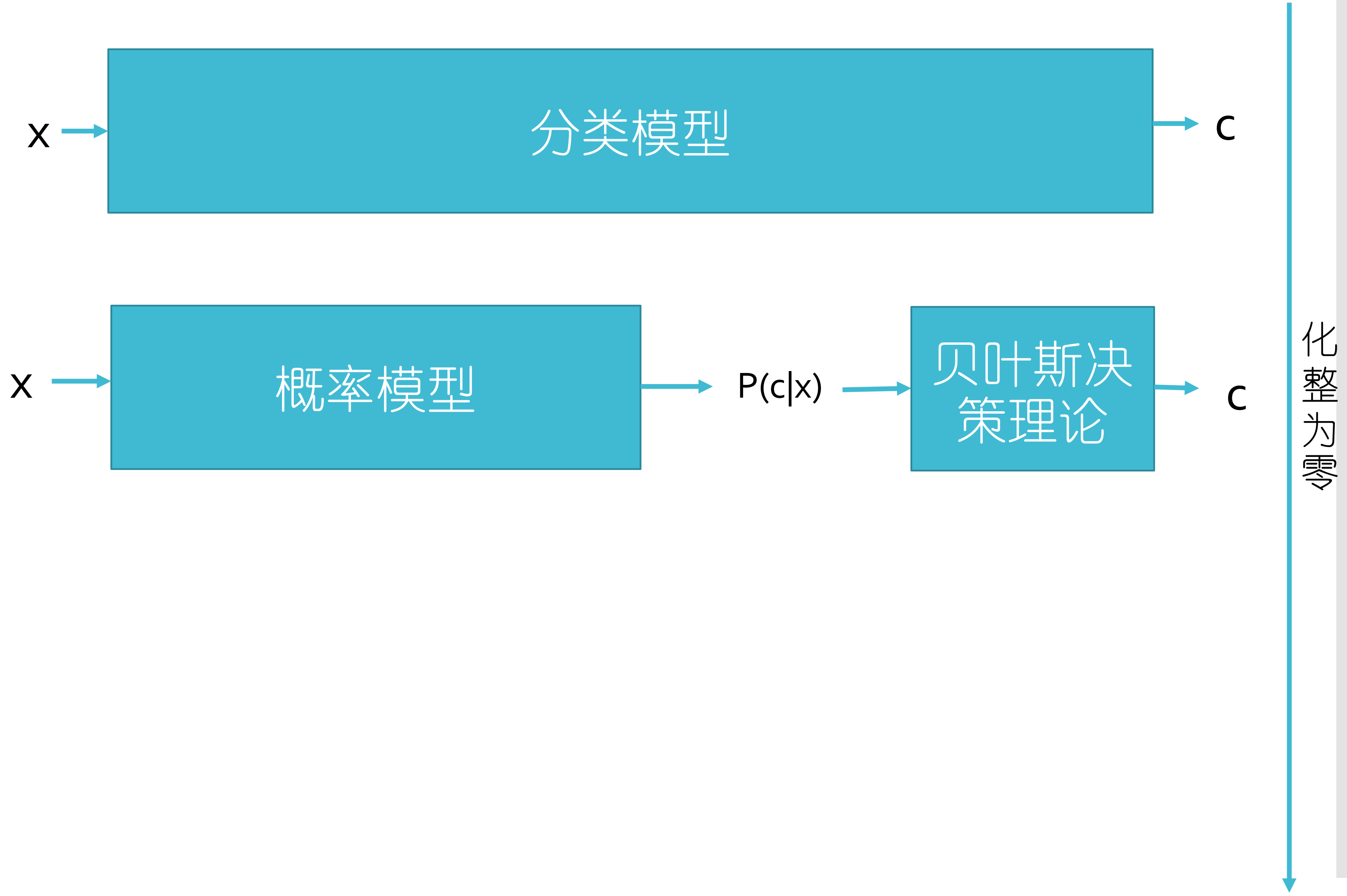
明天 → 天气预报系统 →

- P（下雨）= 60 %
- P（阴天）= 30 %
- P（晴天）=10 %

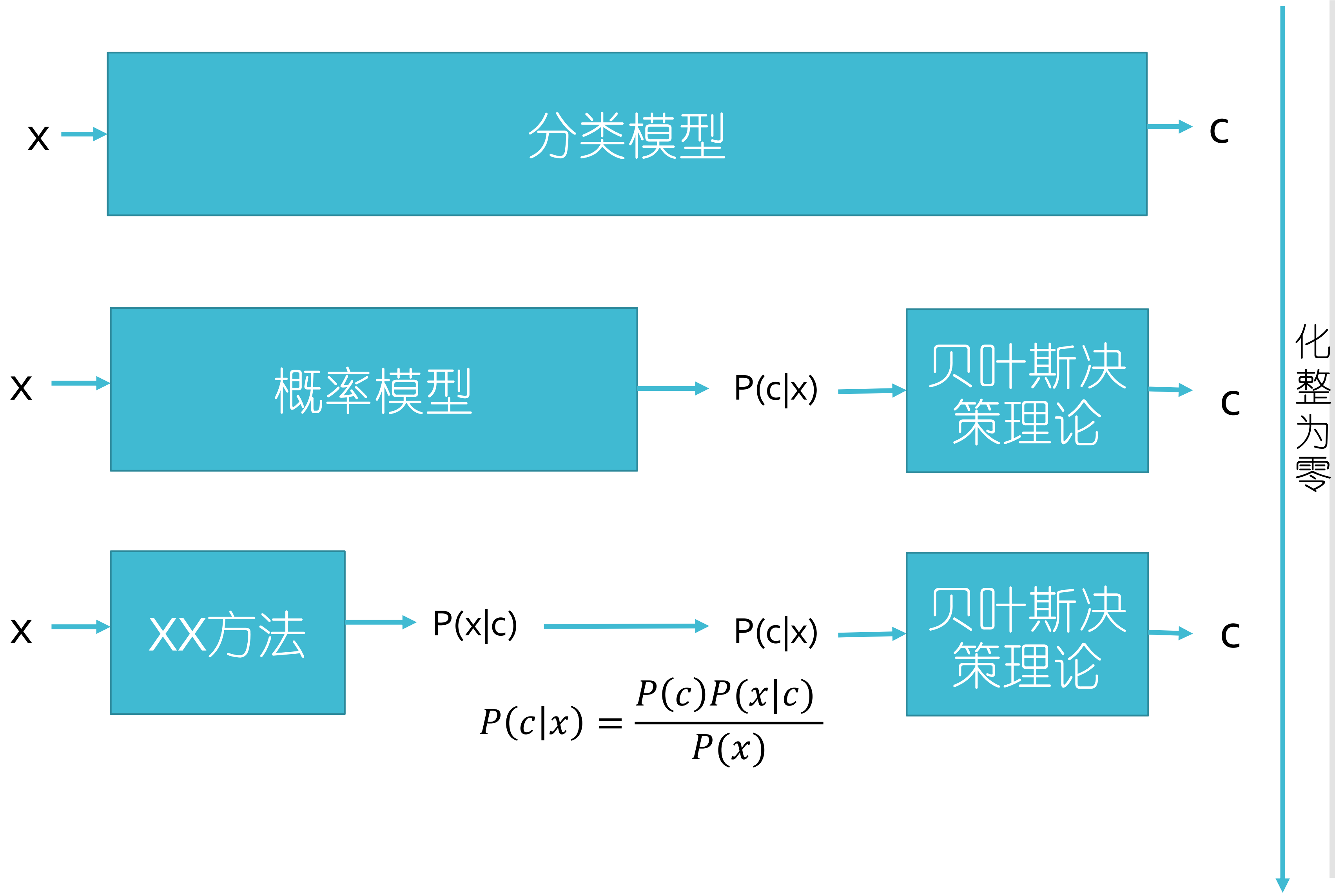
分类器的化整为零



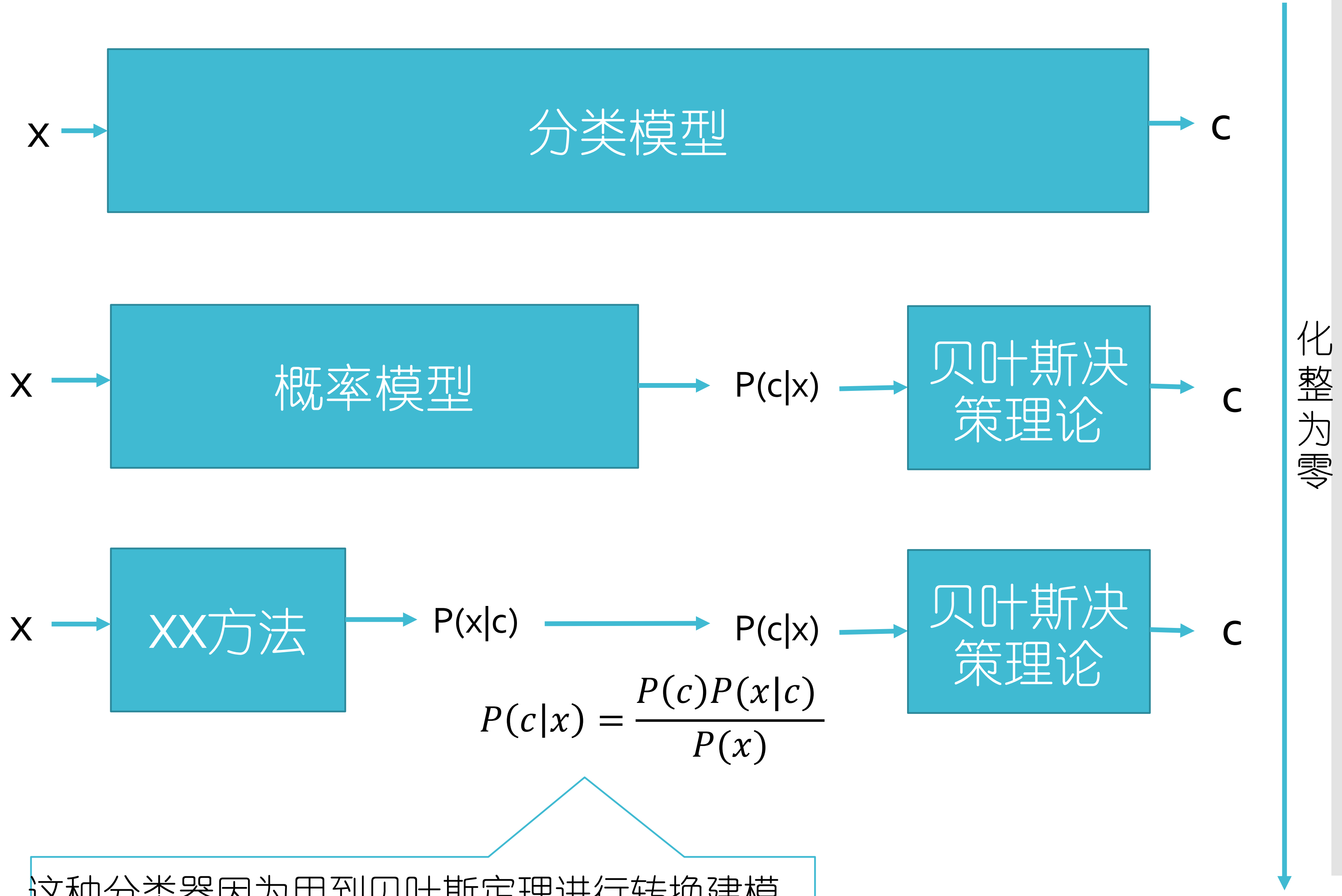
分类器的化整为零



分类器的化整为零


x
分类模型
c
x
概率模型
P(c|x)
贝叶斯决策理论
c
x
XX方法
P(x|c)
P(c|x)
贝叶斯决策理论
c
$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$
化整为零

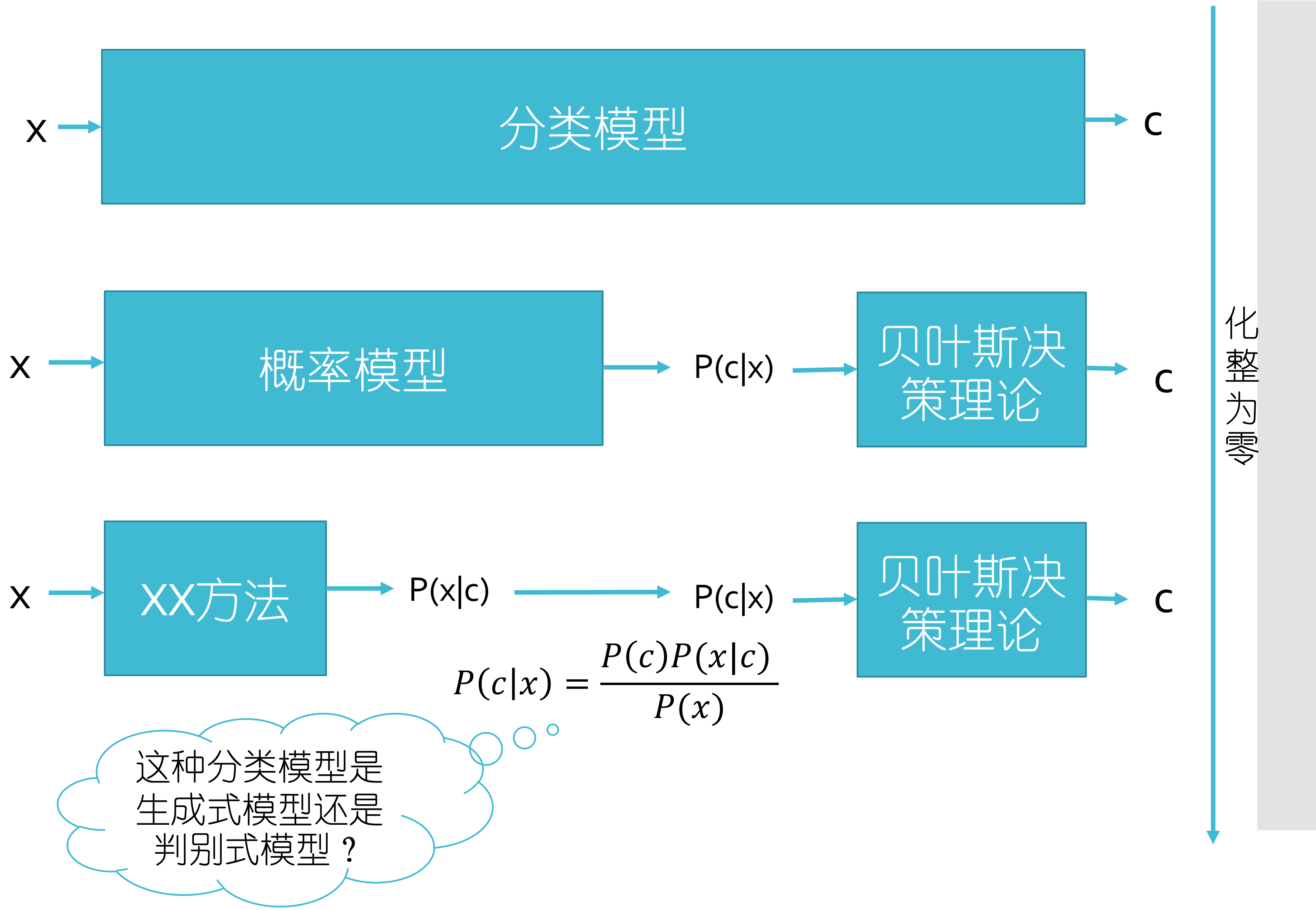
分类器的化整为零



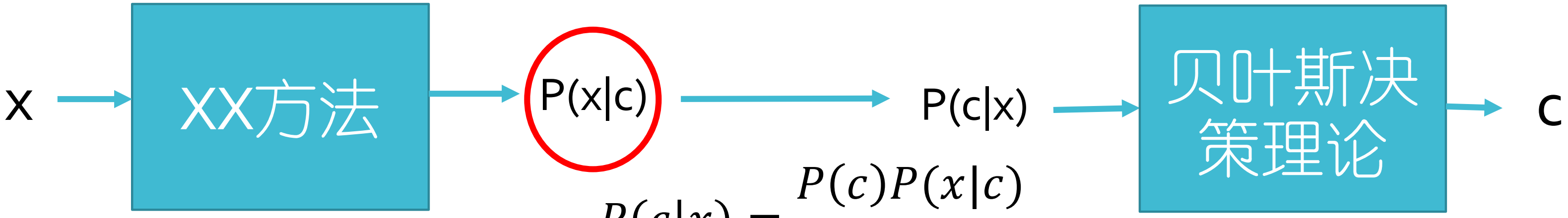
$$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$$

这种分类器因为用到贝叶斯定理进行转换建模，所以这种被成为**贝叶斯分类器**

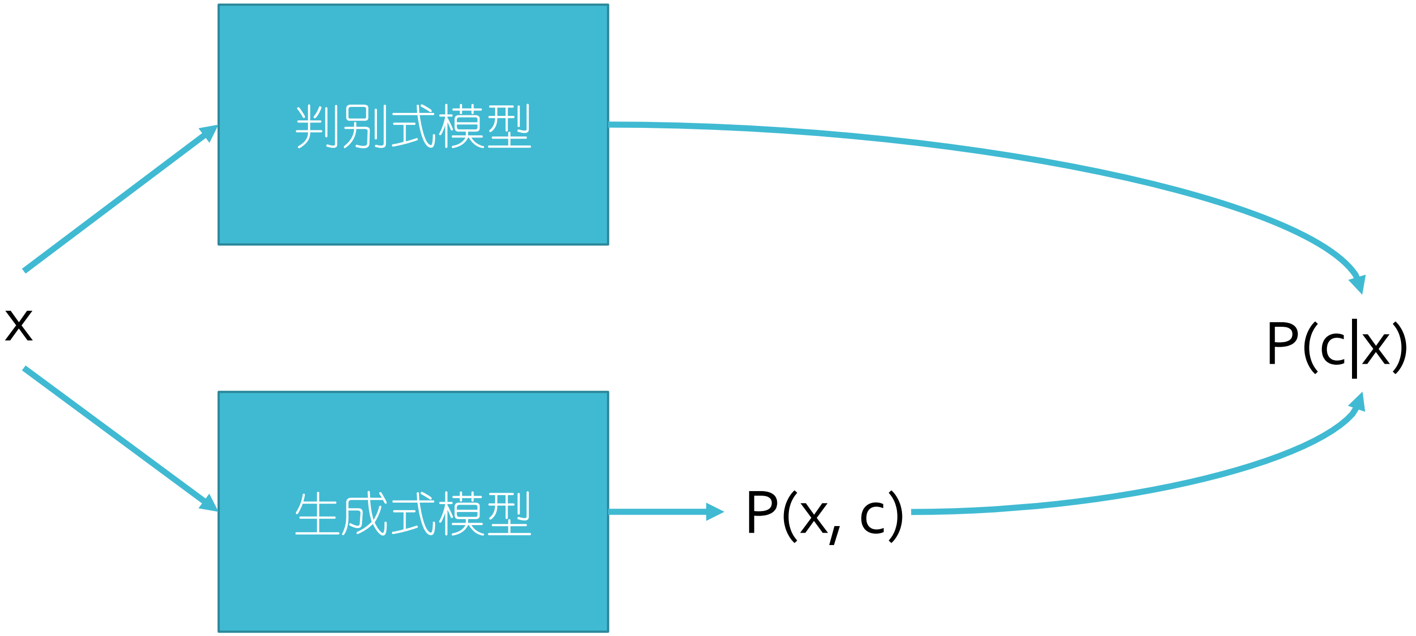
分类器的化整为零

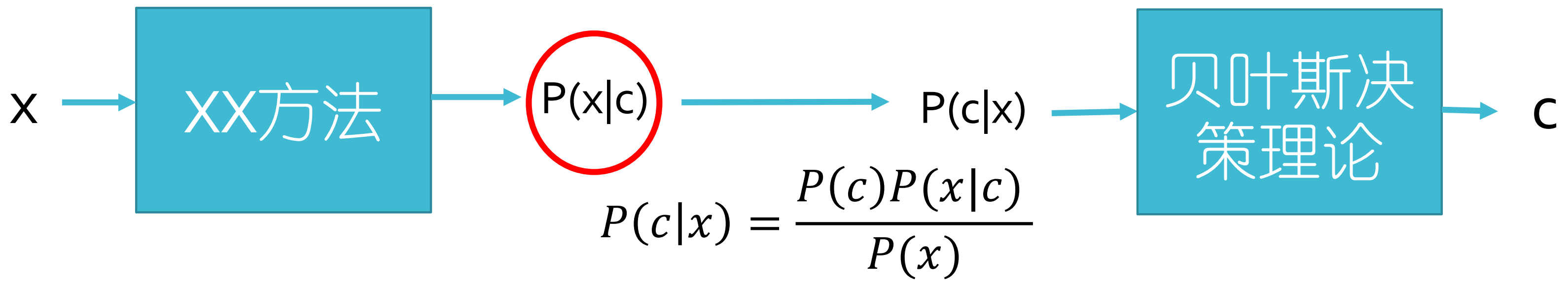


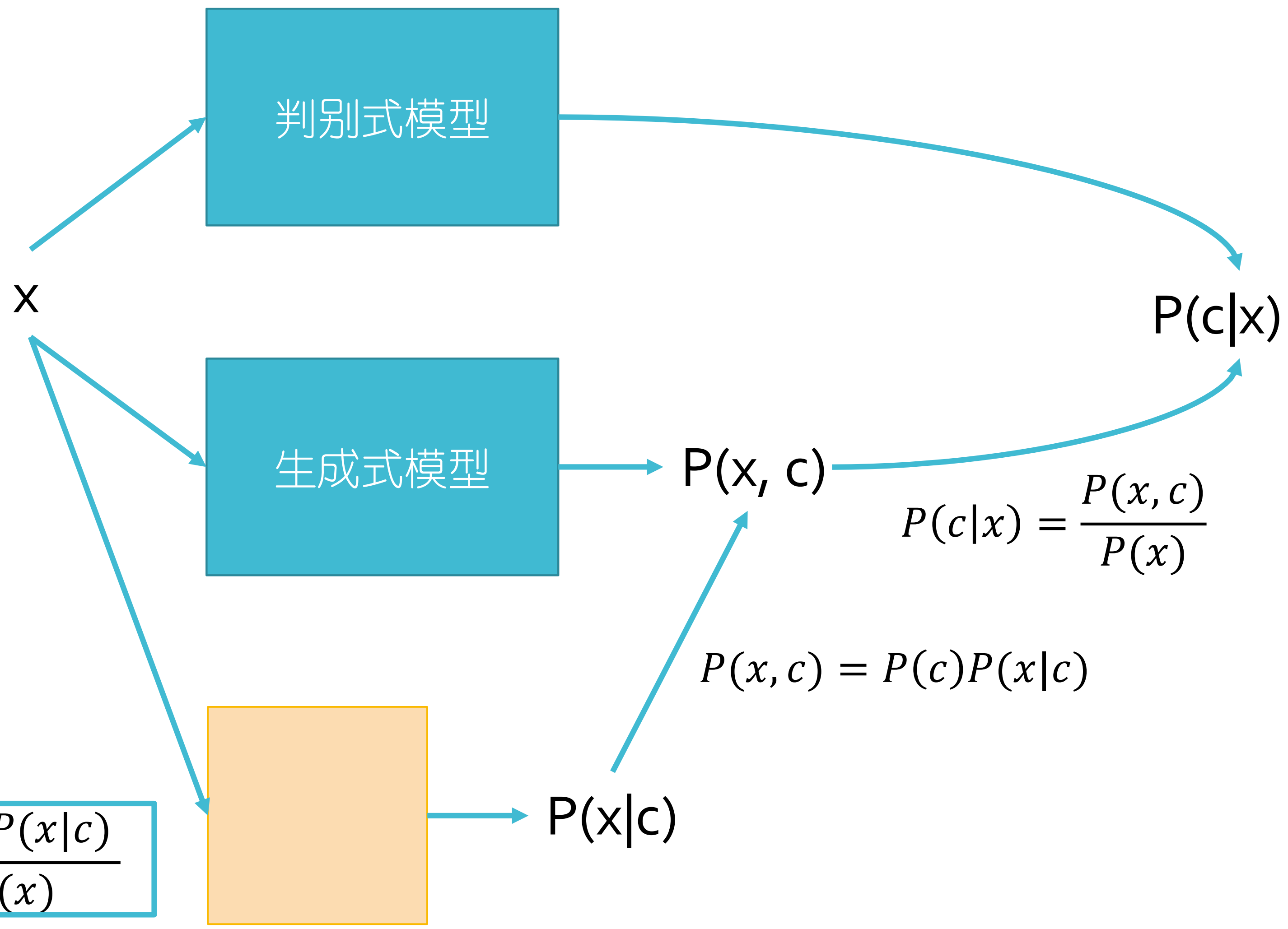
x
分类模型
c
x
概率模型
P(c|x)
贝叶斯决策理论
c
x
XX方法
P(x|c)
P(c|x)
贝叶斯决策理论
c
$P(c|x)=\frac{P(c)P(x|c)}{P(x)}$
这种分类模型是生成式模型还是判别式模型？
化整为零



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

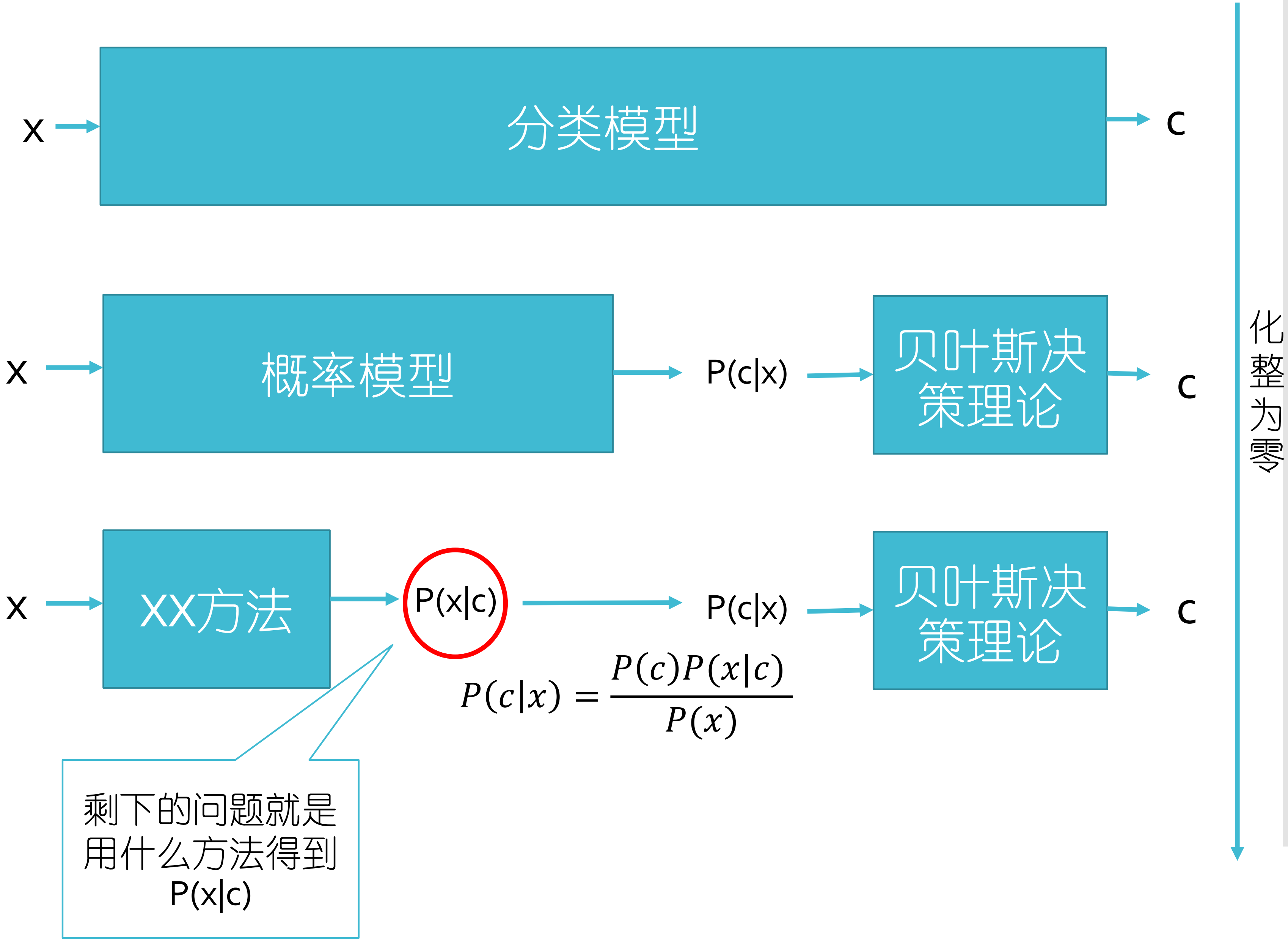






贝叶斯定理 $P(c|x) = \frac{P(x,c)}{P(x)} = \frac{P(c)P(x|c)}{P(x)}$

分类器的化整为零


x
分类模型
c
x
概率模型
P(c|x)
贝叶斯决策理论
c
x
XX方法
P(x|c)
P(c|x)
贝叶斯决策理论
c
$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$
剩下的问题就是用什么方法得到P(x|c)
化整为零

加假设条件简化计算

x → XX方法 → P(x|c) → P(c|x) → 贝叶斯决策理论 → c

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

剩下的问题就是用什么方法得到P(x|c)

x = (x1, x2, x3, x4, x5, ...)

P(x1, x2, x3, x4, x5, ...|c) = ?

强

假设	方法
x1, x2, x3, x4, x5 , ...之间相互独立 →	朴素贝叶斯分类器
x1, x2, x3, x4, x5 , ...之间部分相互依赖 →	半朴素贝叶斯分类器
x1, x2, x3, x4, x5 , ...之间任意相互依赖 →	贝叶斯网路

复杂

7.3 朴素贝叶斯分类器

朴素贝叶斯分类器

x → XX方法 → P(x|c) → P(c|x) → 贝叶斯决策理论 → c

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

剩下的问题就是用什么方法得到P(x|c)

x = (x1, x2, x3, x4, x5, ……)

P(x1, x2, x3, x4, x5, ……|c) = ?

假设 x1, x2, x3, x4, x5 , ……之间相互独立

P(x1, x2, x3, x4, x5, ……|c) = P(x1|c) * P(x2|c) * P(x3|c) * P(x4|c) * P(x5|c)

请同学们照着周志华老师的《机器学习》的p152-154页的题目演算一遍

7.4 半朴素贝叶斯分类器

半朴素贝叶斯

- 讲课时借用周志华老师的ppt进行讲解，请同学们参照周志华老师的ppt和课本进行复习

7.5 贝叶斯网络

贝叶斯网络

- 讲课时借用周志华老师的ppt进行讲解，请同学们参照周志华老师的ppt和课本进行复习