# 6 贝叶斯学习
# (Bayesian learning)

代启国

大连民族大学
计算机科学与技术系

2018 年 12 月 14 日

# 贝叶斯（Bayes）其人

贝叶斯 (Thomas Bayes,1701—1761) 英国牧师、数学家。

# 贝叶斯（Bayes）其人

贝叶斯 (Thomas Bayes,1701—1761) 英国牧师、数学家。

- 贝叶斯在数学方面主要研究概率论。为了证明上帝的存在，他发明了概率统计学原理，遗憾的是，他的这一美好愿望至死也未能实现。

# 贝叶斯（Bayes）其人

贝叶斯 (Thomas Bayes,1701—1761) 英国牧师、数学家。

- 贝叶斯在数学方面主要研究概率论。为了证明上帝的存在，他发明了概率统计学原理，遗憾的是，他的这一美好愿望至死也未能实现。
- 他创立了贝叶斯统计理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。1763 年发表了这方面的论著，对于现代概率论和数理统计都有很重要的作用。

# 贝叶斯（Bayes）其人

贝叶斯 (Thomas Bayes,1701—1761) 英国牧师、数学家。

- 贝叶斯在数学方面主要研究概率论。为了证明上帝的存在，他发明了概率统计学原理，遗憾的是，他的这一美好愿望至死也未能实现。
- 他创立了贝叶斯统计理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。1763 年发表了这方面的论著，对于现代概率论和数理统计都有很重要的作用。
- 贝叶斯思想和方法对概率统计的发展产生了深远的影响。今天，贝叶斯思想和方法在许多领域都获得了广泛的应用。

# 贝叶斯公式

- 先验概率：$P(A)$，事件 $A$ 发生的概率

# 贝叶斯公式

- 先验概率：$P(A)$，事件 $A$ 发生的概率
- 后验概率（条件概率）：$P(A|B)$，事件 $B$ 发生条件下事件 $A$ 发生的概率

# 贝叶斯公式

- 先验概率：$P(A)$，事件 $A$ 发生的概率
- 后验概率（条件概率）：$P(A|B)$，事件 $B$ 发生条件下事件 $A$ 发生的概率
- 联合概率：$P(A,B)$ 事件 A 和 B 同时发生的概率

$$P(A,B) = P(A|B) \cdot P(B)$$

$$P(A,B) = P(B|A) \cdot P(A)$$

# 贝叶斯公式

- 先验概率：$P(A)$，事件 $A$ 发生的概率
- 后验概率（条件概率）：$P(A|B)$，事件 $B$ 发生条件下事件 $A$ 发生的概率
- 联合概率：$P(A,B)$ 事件 A 和 B 同时发生的概率

$$P(A,B) = P(A|B) \cdot P(B)$$

$$P(A,B) = P(B|A) \cdot P(A)$$

## 贝叶斯公式

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

# 贝叶斯学习基本思想

假设一个分类任务有 $N$ 种类别，即 $y = \{c_1, c_2, ..., c_N\}$
在现实分类任务中，给定一个样本 $\mathbf{x}$，分类任务的本质就是计算 $P(c|\mathbf{x})$

# 贝叶斯学习基本思想

假设一个分类任务有 $N$ 种类别，即 $y=\{c_1,c_2,...,c_N\}$
在现实分类任务中，给定一个样本 $\mathbf{x}$，分类任务的本质就是计算 $P(c|\mathbf{x})$

- 机器学习：在有限的训练样本集上尽可能准确地估计后验概率 $P(c|\mathbf{x})$

# 贝叶斯学习基本思想

假设一个分类任务有 $N$ 种类别，即 $y=\{c_1,c_2,...,c_N\}$
在现实分类任务中，给定一个样本 $\mathbf{x}$，分类任务的本质就是计算 $P(c|\mathbf{x})$

- 机器学习：在有限的训练样本集上尽可能准确地估计后验概率 $P(c|\mathbf{x})$
- 判别式模型（discriminative models）：通过建模后验概率 $p(c|x)$ 分布模型来预测 $c$
  - 决策树、BP 神经网络等
- 生成式模型（generative models）：先对联合概率 $p(c,x)$ 分布模型建模，然后在通过后验概率来预测 $c$

# 贝叶斯学习基本思想

假设一个分类任务有 $N$ 种类别，即 $y = \{c_1, c_2, ..., c_N\}$
在现实分类任务中，给定一个样本 $\mathbf{x}$，分类任务的本质就是计算 $P(c|\mathbf{x})$

- 机器学习：在有限的训练样本集上尽可能准确地估计后验概率 $P(c|\mathbf{x})$
- 判别式模型（discriminative models）：通过建模后验概率 $p(c|x)$ 分布模型来预测 $c$
  - 决策树、BP 神经网络等
- 生成式模型（generative models）：先对联合概率 $p(c, x)$ 分布模型建模，然后在通过后验概率来预测 $c$
- 生成式模型：

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

- 根据贝叶斯定理：

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

# 贝叶斯学习基本思想

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

- 类先验（prior）概率：$P(c)$
- 类条件概率：$P(\mathbf{x}|c)$，样本集 $\mathbf{x}$ 相对于类标记 $c$ 的 (class-conditional probability)（亦称 “似然”）
- $P(\mathbf{x})$：归一化证据因子，与类标记 $c$ 无关

$$P(c|\mathbf{x}) \propto P(c)P(\mathbf{x}|c)$$

# 贝叶斯学习基本思想

$$P(c|\mathbf{x}) \propto P(c)P(\mathbf{x}|c)$$

- $P(c)$：表达了样本空间中各类样本所占的比例，根据大数定律，当训练集包含充足的独立同分布样本时，P(c) 可通过各类样本出现的频率来过估计。

# 贝叶斯学习基本思想

$$P(c|\mathbf{x}) \propto P(c)P(\mathbf{x}|c)$$

- $P(c)$：表达了样本空间中各类样本所占的比例，根据大数定律，当训练集包含充足的独立同分布样本时，P(c) 可通过各类样本出现的频率来过估计。
- $P(\mathbf{x}|c)$：由于设计 $\mathbf{x}$ 所有属性的联合概率，假设样本的 $d$ 个属性都是二值的，则会有 $2^d$ 个可能的取值，通常远远高于训练样本数》》》无法通过频率来进行有效估计

# 贝叶斯学习基本思想

$$P(c|\mathbf{x}) \propto P(c)P(\mathbf{x}|c)$$

- $P(c)$：表达了样本空间中各类样本所占的比例，根据大数定律，当训练集包含充足的独立同分布样本时，P(c) 可通过各类样本出现的频率来过估计。
- $P(\mathbf{x}|c)$：由于设计 $\mathbf{x}$ 所有属性的联合概率，假设样本的 $d$ 个属性都是二值的，则会有 $2^d$ 个可能的取值，通常远远高于训练样本数》》》无法通过频率来进行有效估计

**贝叶斯学习的关键**

如何有效地估计类条件概率???

# 朴素贝叶斯分类 (Naive Bayesian classification)

- 贝叶斯学习的关键: 如何有效地估计类条件概率???
- 类条件概率：在所有属性上的联合概率

$$P(\mathbf{x}|c) = P(x_1, x_2, ..., x_d|c)$$

# 朴素贝叶斯分类 (Naive Bayesian classification)

- 贝叶斯学习的关键: 如何有效地估计类条件概率???
- 类条件概率：在所有属性上的联合概率

$$P(\mathbf{x}|c) = P(x_1, x_2, ..., x_d|c)$$

- 朴素贝叶斯分类器：条件独立性假设（conditional independence assumption），在类别已知条件下，所有属性均相互独立，每个属性独立地对分类结果发生影响

# 朴素贝叶斯分类 (Naive Bayesian classification)

- 贝叶斯学习的关键: 如何有效地估计类条件概率???
- 类条件概率：在所有属性上的联合概率

$$P(\mathbf{x}|c) = P(x_1, x_2, ..., x_d|c)$$

- 朴素贝叶斯分类器：条件独立性假设（conditional independence assumption），在类别已知条件下，所有属性均相互独立，每个属性独立地对分类结果发生影响
- $$P(\mathbf{x}|c) = \prod_{i=1}^{d} P(x_i|c)$$

# 朴素贝叶斯分类 (Naive Bayesian classification)

- 贝叶斯学习的关键: 如何有效地估计类条件概率???
- 类条件概率：在所有属性上的联合概率

$$P(\mathbf{x}|c) = P(x_1, x_2, ..., x_d|c)$$

- 朴素贝叶斯分类器：条件独立性假设（conditional independence assumption），在类别已知条件下，所有属性均相互独立，每个属性独立地对分类结果发生影响
- $$P(\mathbf{x}|c) = \prod_{i=1}^{d} P(x_i|c)$$
- $$P(c|\mathbf{x}) = \frac{P(c)P(\mathbf{x}|c)}{P(\mathbf{x})} = \frac{P(c)}{P(\mathbf{x})}\prod_{i=1}^{d} P(x_i|c)$$

朴素贝叶斯判定准则：

$$h_{nb}(\mathbf{x}) = \arg\max_{c \in y} P(c) \prod_{i=1}^{d} P(x_i|c)$$

令 $D_c$ 表示训练集 $D$ 中第 $c$ 类样本的集合，在假设有充足的独立同分布样本的条件下，我们可以计算第 $c$ 类别的先验概率

$$P(c) = \frac{|D_c|}{|D|}$$

对于样本的属性是离散值，令 $D_{c,x_i}$ 表示 $D_c$ 中第 $i$ 个属性上取值为 $x_i$ 的样本组成的集合，则

$$P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|}$$

# 朴素贝叶斯分类 (Naive Bayesian classification)

朴素贝叶斯判定准则:

$$h_{nb}(\mathbf{x}) = \arg\max_{c \in y} P(c) \prod_{i=1}^{d} P(x_i | c)$$

对于样本的属性是连续值，假定

$$p(x_i|c) \sim N(\mu_{c,i}, \sigma_{c,i}^2)$$

其中，$N(\mu_{c,i}, \sigma_{c,i}^2)$ 表示第 $c$ 类样本第 $i$ 个属性上取值的均值和方差

$$p(x_i|c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right)$$

拉普拉斯平滑（修正）

- 对离散属性计算概率时，经常容易出现概率值为 0 的情况
- 导致溢出等问题
- 为了避免这一问题，通常要采用平滑处理（拉普拉斯修正）

$$P(c) = \frac{|D_c| + 1}{|D| + d}$$

$$P(x_i|c) = \frac{|D_{c,x_i}| + 1}{|D_c| + d}$$

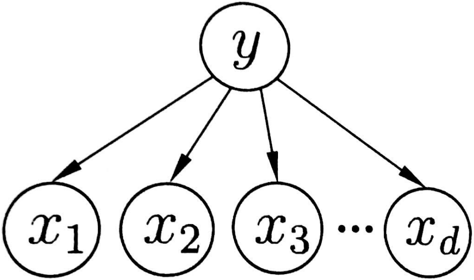
# 朴素贝叶斯分类 (Naive Bayesian classification)

## 算法

1. 输入: 样本 $\mathbf{x}$ 和训练集 D
2. 计算:$y = \{c_1, c_2, ..., c_N\}$ 中每一个类别 $c_n$ 的先验概率 $P(c_n)$，及其在不同样本属性上的类条件概率 $P(x_i|c_n)$
3. 输出:

$$h_{nb}(\mathbf{x}) = \arg\max_{c \in y} P(c) \prod_{i=1}^{d} P(x_i|c)$$

# 朴素贝叶斯分类 (Naive Bayesian classification)



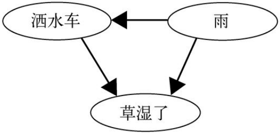

朴素贝叶斯中属性与类别之间的依赖关系

# 贝叶斯网（Bayesian network）

贝叶斯网：也称信念网

- 借助有向无环图（Directed Acyclic Graph，DAG）刻画属性间依赖关系
- 采用条件概率表（Conditional Probability Table，CPT）描述属性间的联合概率分布

| 雨 | 洒水车 T | 洒水车 F |
|---|---|---|
| F | 0.4 | 0.6 |
| T | 0.01 | 0.99 |




| 雨 T | 雨 F |
|---|---|
| 0.2 | 0.8 |

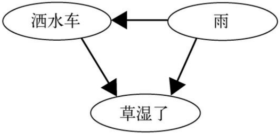
| 洒水车 | 雨 | 草湿了 T | 草湿了 F |
|---|---|---|---|
| F | F | 0.0 | 1.0 |
| F | T | 0.8 | 0.2 |
| T | F | 0.9 | 0.1 |
| T | T | 0.99 | 0.01 |

# 贝叶斯网（Bayesian network）

## 贝叶斯网

- 贝叶斯网络 $B$ 由网络结构 $G$ 和参数 $\theta$ 两部分构成，即 $B=(G,\theta)$
- $G$ 中每个结点对应于一个属性，结点连接表示依赖关系
- 参数 $\theta$ 描述了这种依赖关系的强度

| 雨 | 洒水车 T | 洒水车 F |
|---|---|---|
| F | 0.4 | 0.6 |
| T | 0.01 | 0.99 |



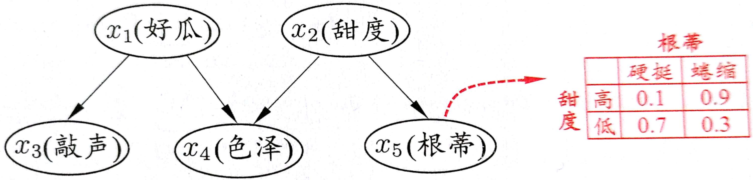

| 雨 T | 雨 F |
|---|---|
| 0.2 | 0.8 |

| 洒水车 | 雨 | 草湿了 T | 草湿了 F |
|---|---|---|---|
| F | F | 0.0 | 1.0 |
| F | T | 0.8 | 0.2 |
| T | F | 0.9 | 0.1 |
| T | T | 0.99 | 0.01 |

## 贝叶斯网

贝叶斯网络将各属性（变量）的联合分布概率定义为：

$$P_B\left(x_1, x_2, ..., x_d\right) = \prod_{i=1}^{d} P_B\left(x_i | \pi_i\right)$$



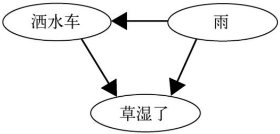

$$P\left(x_1, x_2, x_3, x_4, x_5\right) = P\left(x_1\right) P\left(x_2\right) P\left(x_3 | x_1 P(x_4 | x_1, x_2) P(x_5 | x_2)\right)$$

# 贝叶斯网（Bayesian network）

贝叶斯网络的学习分为两个方面

- 网络结构学习（G）（MDL 算法等）
- 参数学习 $(\theta)$（BIC 评分等）

| 雨 | 洒水车 T | 洒水车 F |
|---|---|---|
| F | 0.4 | 0.6 |
| T | 0.01 | 0.99 |




| 雨 T | 雨 F |
|---|---|
| 0.2 | 0.8 |

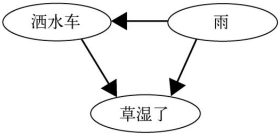
| 洒水车 | 雨 | 草湿了 T | 草湿了 F |
|---|---|---|---|
| F | F | 0.0 | 1.0 |
| F | T | 0.8 | 0.2 |
| T | F | 0.9 | 0.1 |
| T | T | 0.99 | 0.01 |

# 贝叶斯网（Bayesian network）

## 贝叶斯网络的推理

- 训练好的贝叶斯网可以用来回答问题 (推断)
- 通过已知（观测）的属性变量值 (证据)，推测其他未知属性变量的值
- 常用方法有吉布斯采样算法等

| 雨 | 洒水车 T | 洒水车 F |
|---|---|---|
| F | 0.4 | 0.6 |
| T | 0.01 | 0.99 |




| 雨 T | 雨 F |
|---|---|
| 0.2 | 0.8 |

| 洒水车 | 雨 | 草湿了 T | 草湿了 F |
|---|---|---|---|
| F | F | 0.0 | 1.0 |
| F | T | 0.8 | 0.2 |
| T | F | 0.9 | 0.1 |
| T | T | 0.99 | 0.01 |

# 小结

- 贝叶斯学习以贝叶斯理论（公式）为基础
- 通过一定的假设来构建依赖关系和强度的模型
- 朴素贝叶斯（NB）假设属性之间相互独立
- 贝叶斯网络（BN）假设属性之间是条件独立的（部分依赖）