

一、朴素贝叶斯法 00:04

1. 贝叶斯思维 00:24

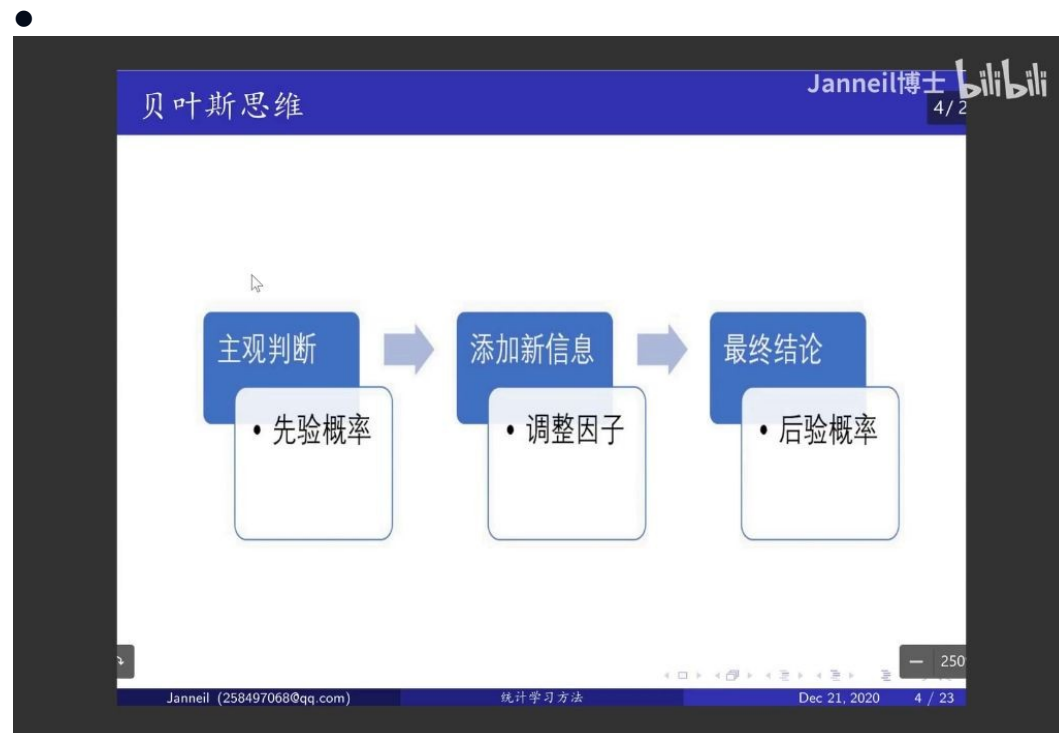


- **核心基础**：朴素贝叶斯法的核心是贝叶斯定理，其思维方式与人脑认知过程高度相似，是机器学习中最基础的方法之一。
 - **生活实例**：通过冬至餐馆点餐场景展示贝叶斯思维的实际应用过程。
- 1) 贝叶斯思维过程分解



- **先验概率**：初始的主观判断（如冬至日老板默认推荐饺子），对应事件发生的初始概率 $P(A)$ 。
- **调整因子**：新获得的信息（如顾客用四川话拒绝），对应条件概率 $P(B|A)$ 。
- **后验概率**：修正后的结论（改推荐羊肉汤），即 $P(A|B)$ ，通过贝叶斯公式计算得出。

2) 完整推理链条



- **典型场景**：北方餐馆冬至日顾客行为预测
 - 初始判断：基于节气（冬至）推荐饺子（先验概率）
 - 信息更新：顾客使用四川方言拒绝（似然函数）
 - 结论修正：改推荐羊肉汤（后验概率）
 - **思维特点**：通过不断吸收新证据来修正原有认知，体现“概率动态更新”的核心思想。
2. 知识要点 02:04
- 1) 条件概率 02:16
- 条件概率的引入
 - **生活实例**：通过《阿甘正传》中“生活就像一盒巧克力糖”的比喻，说明统计可以预测事件发生的可能性，引出条件概率的概念。
 - **核心思想**：在已知部分信息（条件）的情况下，计算特定事件发生的概率。
 - 例题 1：随机取出一块黑色巧克力和红色巧克力的可能性 02:47



已知：现有一盒巧克力，一种装了16块。其中黑色、白色、棕色巧克力各4块，红色、黄色巧克力各2块

问：

- 随机取出一块黑色巧克力的可能性是多少？ $\frac{4}{16}$
- 随机取出一块红色巧克力的可能性是多少？ $\frac{2}{16}$

题目解析

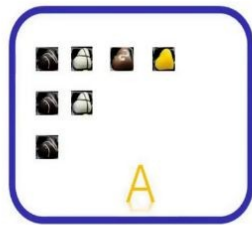
■ 已知条件：16块巧克力中，黑/白/棕各4块，红/黄各2块

■ 计算过程：

- 黑色巧克力概率： $\frac{4}{16} = \frac{1}{4}$
- 红色巧克力概率： $\frac{2}{16} = \frac{1}{8}$

■ 关键点：基础概率计算为事件发生次数与总可能性的比值

- 例题2：已知巧克力出自A盒，取出黑色巧克力的概率 03:59



问：已知巧克力出自A盒，取出黑色巧克力的概率

解：

$$P(\text{Black}|\text{BoxA}) = \frac{P(\text{Black and BoxA})}{P(\text{BoxA})}$$

$$P(\text{Black and BoxA}) = \frac{3}{16}$$

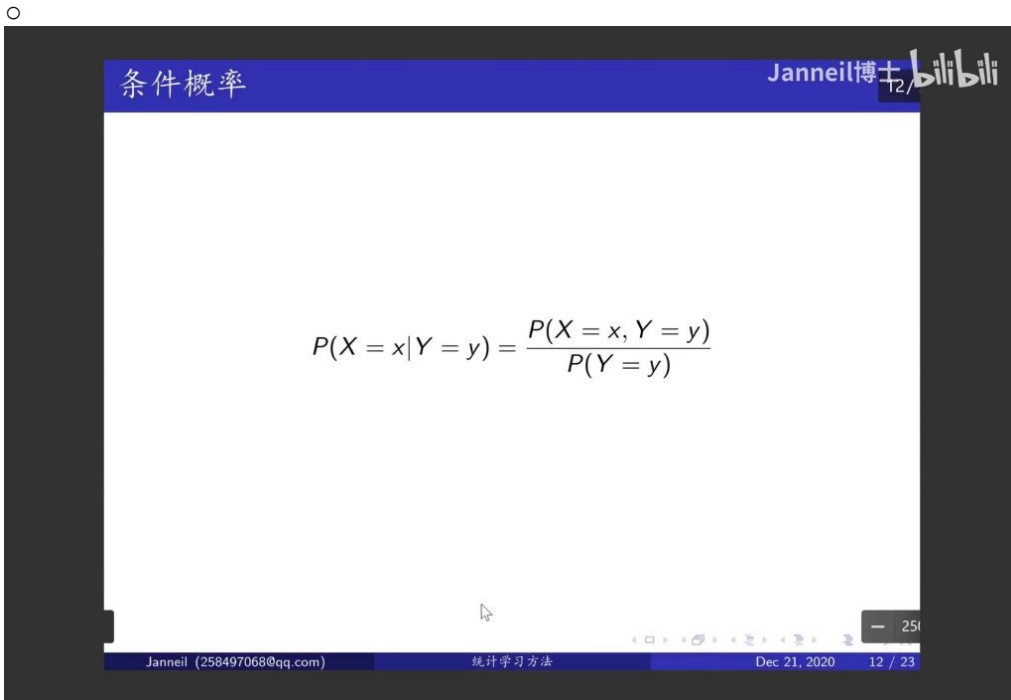
$$P(\text{BoxA}) = \frac{7}{16}$$

题目解析

■ 分布情况：

- A盒：3黑+2白+1棕+1黄（共7块）

- B 盒：1 黑+2 白+3 棕+1 黄+2 红（共 9 块）
- **直观解法**：A 盒中黑色概率= $\frac{3}{7}$
- **条件概率公式**：
 - $P(\text{Black} \vee \text{BoxA}) = \frac{P(\text{Black} \cap \text{BoxA})}{P(\text{BoxA})} = \frac{3/16}{7/16} = \frac{3}{7}$
- **公式理解**：分子为联合概率，分母为条件事件的概率
- 条件概率公式 05:41



- **准形式**： $P(X=x \vee Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$
- **解释**：
 - 分子表示两个事件同生的合概率
 - 分母表示条件事件发生的概率
- **应用意义**：为后续贝叶斯定理的学习奠定基础
- 2) 贝叶斯定理 06:22
- 例题 1：已知取出一块黑色巧克力，它来自 A 盒的概率 06:25

问：已知取出一块黑色巧克力，它来自A盒的概率

解：

$$P(\text{BoxA}|\text{Black}) = \frac{P(\text{Black and BoxA})}{P(\text{Black})}$$

$$P(\text{Black and BoxA}) = P(\text{Black}|\text{BoxA}) \cdot P(\text{BoxA})$$

$$P(\text{Black}) = P(\text{Black}|\text{BoxA}) \cdot P(\text{BoxA}) + P(\text{Black}|\text{BoxB}) \cdot P(\text{BoxB})$$

Janneil (258497068@qq.com)

统计学习方法

Dec 21, 2020

15 / 23

题目解析

■ **问题转换**：与条件概率例题形成"逆问题"（已知颜色求来源）

■ **计算过程**：

- 分子： $P(\text{Black} \vee \text{BoxA}) \cdot P(\text{BoxA}) = \frac{3}{7} \cdot \frac{7}{16} = \frac{3}{16}$

- 分母（全概率公式）：

$$P(\text{Black}) = P(\text{Black} \vee \text{BoxA}) P(\text{BoxA}) + P(\text{Black} \vee \text{BoxB}) P(\text{BoxB}) = \frac{3}{16} + \frac{1}{16} = \frac{4}{16}$$

- 最终结果： $\frac{3/16}{4/16} = \frac{3}{4}$

■ **验证**：与直接观察结果一致（总黑色4块，A盒占3块）

● 贝叶斯定理：逆概率思维 07:03

○ **核心特点**：将条件概率中的"条件"与"结果"互换

○ **应用场景**：已知结果反推原因的概率分布

○ **与条件概率关系**：贝叶斯定理是条件概率公式的逆向应用

● 例题2：已知存在K类，求实例归属第Ci类的可能性 10:04

已知:

存在 K 类 $c_1, c_2, \dots, c_K$, 给定一个新的实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$

问: 该实例归属第 c_i 类的可能性有多大?

$$P(Y = c_i | X = x) = \frac{P(X = x | Y = c_i) \cdot P(Y = c_i)}{P(X = x)}$$

即,

$$P(Y = c_i | X = x) = \frac{P(X = x | Y = c_i) \cdot P(Y = c_i)}{\sum_{i=1}^K P(X = x | Y = c_i) \cdot P(Y = c_i)}$$

一般形式:

$$P(Y = c_i | X = x) = \frac{P(X = x \vee Y = c_i) \cdot P(Y = c_i)}{\sum_{i=1}^K P(X = x \vee Y = c_i) \cdot P(Y = c_i)}$$

成部分:

■ 先验概率: $P(Y = c_i)$ (各类别初始概率)

■ 似然度: $P(X = x \vee Y = c_i)$ (某类别下特征出现的概率)

■ 证据因子: 分母的全概率展开

应用扩展: 适用于多特征 $(x^{(1)} \dots x^{(n)})$ 和多类别 $(c_1 \dots c_K)$ 的分类问题

3) 朴素贝叶斯分类 12:09

● 例题 1: 巧克力归属的贝叶斯分类问题

已知: 有 A 和 B 两盒巧克力, 现在拿到一块黑色的巧克力。

问: 该巧克力最有可能是哪个盒子的?

$$P(\text{BoxA} | \text{Black}) = \frac{P(\text{Black} | \text{BoxA}) \cdot P(\text{BoxA})}{P(\text{Black})}$$

$$P(\text{BoxB} | \text{Black}) = \frac{P(\text{Black} | \text{BoxB}) \cdot P(\text{BoxB})}{P(\text{Black})}$$

- **问题描述**：已知 A 和 B 两盒巧克力，现拿到一块黑色巧克力，求最可能来自哪个盒子
- **解法**：

- 计算来自 A 盒的概率： $P(\text{BoxA} \vee \text{Black}) = \frac{P(\text{Black} \vee \text{BoxA}) \cdot P(\text{BoxA})}{P(\text{Black})}$,

结果为 $\frac{3}{4}$

- 计算来自 B 盒的概率： $P(\text{BoxB} \vee \text{Black}) = \frac{P(\text{Black} \vee \text{BoxB}) \cdot P(\text{BoxB})}{P(\text{Black})}$,

结果为 $\frac{1}{4}$

- **结论**：因 $\frac{3}{4} > \frac{1}{4}$ ，故最可能来自 A 盒

- 贝叶斯分类原理 13:09

贝叶斯分类
Janneil博士

已知：

存在 K 类 $c_1, c_2, \dots, c_K$ ，给定一个新的实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$

问：该实例归属哪一类？

- $$P(Y = c_i | X = x) = \frac{P(X = x | Y = c_i) \cdot P(Y = c_i)}{\sum_{i=1}^K P(X = x | Y = c_i) \cdot P(Y = c_i)}$$
- $$\arg \max_{c_i} P(X = x | Y = c_i) \cdot P(Y = c_i)$$

Janneil (258497068@qq.com)
统计学习方法
Dec 21, 2020 18 / 23

- **核心思想**：通过比较各类别的后验概率确定实例归属
- 公式表达：

- 后验概率： $P(Y = c_i \vee X = x) = \frac{P(X = x \vee Y = c_i) \cdot P(Y = c_i)}{\sum_{i=1}^K P(X = x \vee Y = c_i) \cdot P(Y = c_i)}$

- 决策规则：

- **应用步骤**：

- 计算实例归属于每类的概率
 - 选择概率最大的类别作为最终分类结果

- 朴素贝叶斯法的假设 13:55

- **关键假设**：实例特征之间相互独立
- 数学表达： $P(X = x \vee Y = c_i) = \prod_{j=1}^n P(X^{(j)} = x^{(j)} \vee Y = c_i)$
- **意义**：将联合概率分解为各特征条件概率的乘积，简化计算

- 朴素贝叶斯法的公式推导 14:10

假设：实例特征之间相互独立

$$\begin{aligned}
 P(X = x | Y = c_i) &= \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i) \\
 \Rightarrow P(X = x) &= \sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i) \\
 \Rightarrow P(Y = c_i | X = x) &= \frac{P(X = x | Y = c_i) \cdot P(Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)} \\
 \Rightarrow P(Y = c_i | X = x) &= \frac{P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}
 \end{aligned}$$

Janneil (258497068@qq.com)

统计学习方法

Dec 21, 2020

20 / 23

○ **推导过程：**

- 基于特征独立性假设展开条件概率

代入贝叶斯公式得到：

- $$P(Y = c_i | X = x) = \frac{P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}$$

○ **计算优化：**因分母相同，实际比较时只需比较分子部分

● 朴素贝叶斯分类法的应用 15:34

○ 分类步骤：

- 计算每类的先验概率 $P(Y = c_i)$
- 计算各特征的条件概率 $P(X^{(j)} = x^{(j)} | Y = c_i)$
- 对各类别计算分子项 $P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)$
- 选择使分子最大的类别作为分类结果

● 例题 2: 朴素贝叶斯分类法的应用 16:06

朴素贝叶斯分类

Janneil博士

已知：

存在 K 类 $c_1, c_2, \dots, c_K$ ，给定一个新的实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$

问：该实例归属哪一类？

- $$P(Y = c_i | X = x) = \frac{P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}$$
- $$\arg \max_{c_i} P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)$$

Janneil (258497068@qq.com) 统计学习方法 Dec 21, 2020 21 / 23

- **应用要点：**实际分类时只需比较分子部分，即
- 朴素贝叶斯法的核心内容回顾 17:20

朴素贝叶斯

Janneil博士

朴素贝叶斯

Janneil (258497068@qq.com) 统计学习方法 Dec 30, 2020 2 / 19

- **本质：**基于贝叶斯定理与特征条件独立假设的分类方法
 - **方法类型：**生成方法（先学习联合概率分布，再求条件概率分布）
 - 与判别方法区别：
 - 生成方法：学习联合概率分布 → 条件概率分布 → 预测
 - 判别方法：直接从数据学习决策函数或条件概率分布
 - **训练目标：**通过训练数据集学习联合概率分布 $P(X, Y)$
 - **输入输出：**
 - 输入： n 维特征空间
 - 输出： k 个离散类别
- 4) 朴素贝叶斯法 17:32

● 例题：两盒巧克力问题 21:49

○

例子：两盒巧克力			Janneil博士	bilibili
$P(X,Y)$	A盒	B盒		
黑	$\frac{3}{7} \times \frac{7}{16} = \frac{3}{16}$	$\frac{1}{9} \times \frac{9}{16} = \frac{1}{16}$		
白	$\frac{2}{7} \times \frac{7}{16} = \frac{2}{16}$	$\frac{2}{9} \times \frac{9}{16} = \frac{2}{16}$		
棕	$\frac{1}{7} \times \frac{7}{16} = \frac{1}{16}$	$\frac{3}{9} \times \frac{9}{16} = \frac{3}{16}$		
黄	$\frac{1}{7} \times \frac{7}{16} = \frac{1}{16}$	$\frac{1}{9} \times \frac{9}{16} = \frac{1}{16}$		
红	$\frac{0}{7} \times \frac{7}{16} = \frac{0}{16}$	$\frac{2}{9} \times \frac{9}{16} = \frac{2}{16}$		

○ 先验概率计算：

■ A盒概率： $P(Y=A) = \frac{7}{16}$ (A盒7块巧克力/总数16块)

■ B盒概率： $P(Y=B) = \frac{9}{16}$ (B盒9块巧克力/总数16块)

○ 条件概率分布：

■ A盒黑色巧克力： $P(X=黑 \vee Y=A) = \frac{3}{7}$

■ A盒白色巧克力： $P(X=白 \vee Y=A) = \frac{2}{7}$

■ B盒黑色巧克力： $P(X=黑 \vee Y=B) = \frac{1}{9}$

○ 合概率分布：

■ $P(X=黑, Y=A) = P(Y=A) \times P(X=黑 \vee Y=A) = \frac{7}{16} \times \frac{3}{7} = \frac{3}{16}$

■ $P(X=白, Y=A) = \frac{7}{16} \times \frac{2}{7} = \frac{2}{16}$

● 朴素贝叶斯法的朴素 25:56

○ 朴素贝叶斯法的条件独立性假设 26:00

假设：实例特征之间相互独立

$$P(X=x|Y=c_i) = P(X^{(1)}=x^{(1)}, \dots, X^{(n)}=x^{(n)}|Y=c_i) \\ = \prod_{j=1}^n P(X^{(j)}=x^{(j)}|Y=c_i)$$

- **核 心 假 设**：特 征 间 相 互 独 立，即 $P(X=x \vee Y=c_i) = \prod_{j=1}^n P(X^{(j)}=x^{(j)} \vee Y=c_i)$
- **术语来源**：“朴素”源自英文“naive”，表示该假设在现实中往往不成立但依然实用

- 例题 1：通过日漫帅哥看条件独立性假设的必要性 26:49



■ **特征定义：**

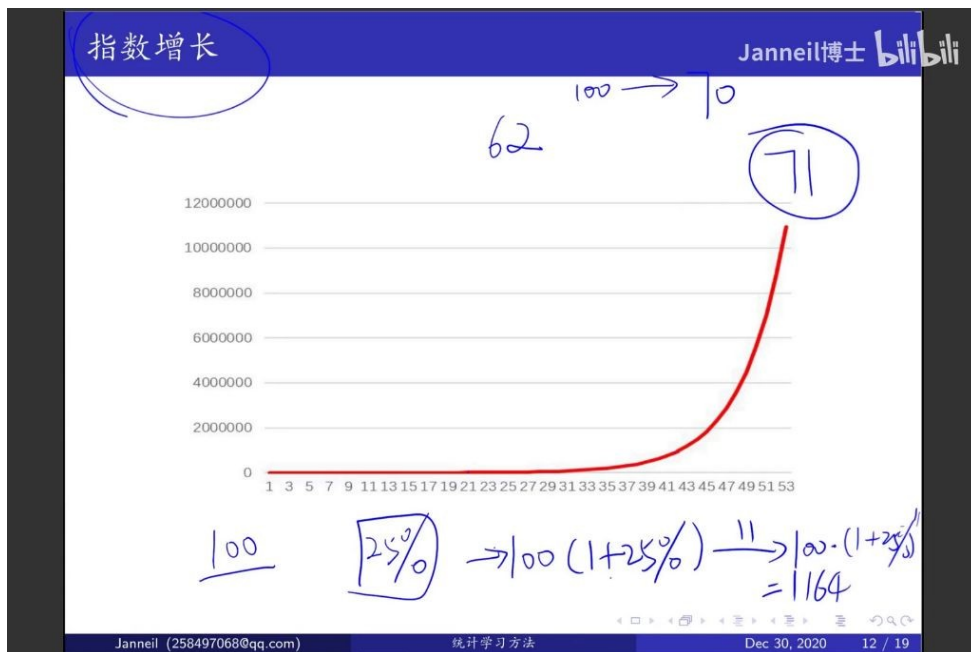
- 身高：>1.7m 为高，≤1.7m 为矮
- 体重：>标准为重，≤标准为轻
- 脸型：圆脸/长脸
- 鼻型：高挺/塌鼻

- 参数个数与组合数计算 27:04

■ **组合数计算：**

- 4 个二值特征： $2^4=16$ 种组合
- 加入类别变量（帅/不帅）： $2^5=32$ 种组合

- 一般化公式： $K \prod_{j=1}^n S_j$ (K 为类别数, S_j 为第 j 个特征取值数)
- 指数增长的恐怖性 29:54



- 疫情传播示例：
 - 初始 100 人感染，日增 25%
 - 第 11 天： $100 \times (1+25\%)^{11} \approx 1164$ 人
 - 第 31 天：超 10 万人
 - 第 42 天：超 100 万人
 - 第 71 天：理论上感染全球人口
- 朴素贝叶斯法的独立性假设与公式 32:28
 - 计算优势：将联合概率分解为各特征条件概率的乘积，避免直接计算指数级组合
 - 实用价值：即使独立性假设不完全成立，模型仍能保持较好效果
- 朴素贝叶斯法与后验概率最大化 33:19
 - 与统计学习三要素的关系：后验概率最大化对应损失最小化准则
 - 逆向思维：与传统条件概率方向相反，实现从特征到类别的推断
- 朴素贝叶斯分类 33:55
 - 朴素贝叶斯分类的基本概念 34:00
 - 问题定义：已知存在 K 个类别 $c_1, c_2, \dots, c_K$ ，给定一个具有 n 个特征的新实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$ ，需要判断该实例所属类别。
 - 核心思想：通过计算后验概率（条件概率）来确定实例所属类别，后验概率由先验概率和条件概率分布组成。
 - 后验概率的计算方法 34:20

已知：

存在 K 类 $c_1, c_2, \dots, c_K$ ，给定一个新的实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$

问：该实例归属哪一类？

- 后验概率

$$P(Y = c_i | X = x) = \frac{P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}$$

- 分类

$$y = \arg \max_{c_i} P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)$$

计算公式：

$$P(Y = c_i | X = x) = \frac{P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}{\sum_{i=1}^K P(Y = c_i) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)}$$

■ 计算原理：

- 分子：等于先验概率 $P(Y = c_i)$ 乘以条件概率 $\prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_i)$
- 分母：对所有类别求和，保证概率归一化
- **独立性假设**：特征之间相互独立，使计算简便可行

■ 分类准则的确定 36:11

■ 决策规则：

■ 化原理：由于分母对所有类别相同，只需比较分子大小即可确定最大后验概率

■ **核心思想**：将实例分配给使其后验概率最大的类别

■ 后验概率最大化准则与损失最小化的关系 37:14

- 0-1 损失函数 (0-1 Loss Function)

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$$

- 期望风险

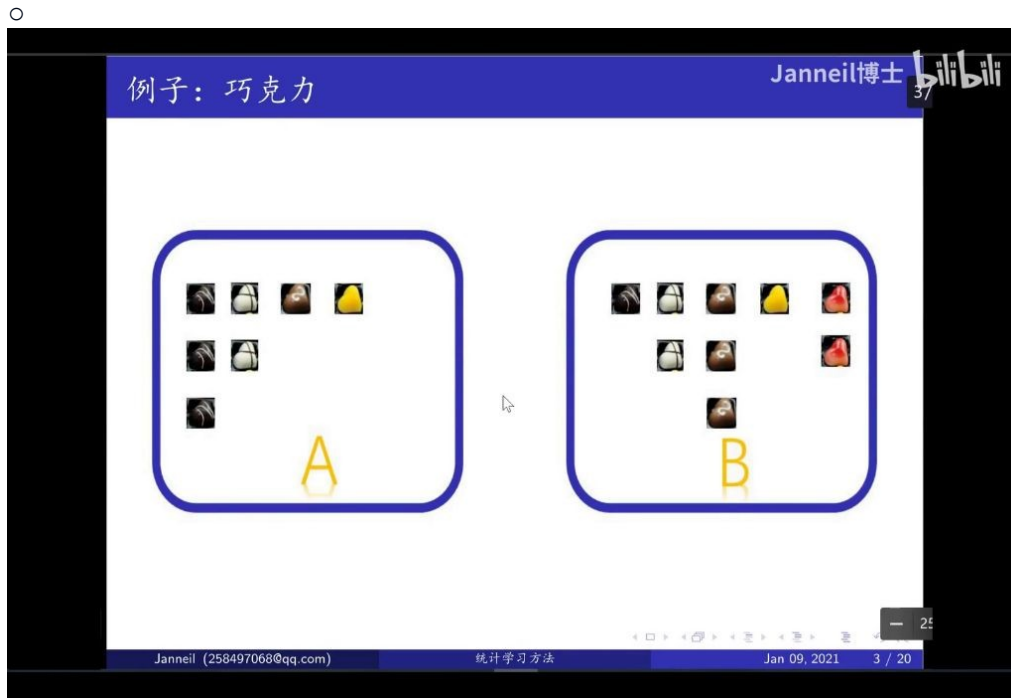
$$R(f) = E[L(Y, f(X))]$$

- 后验概率最大化

$$f(x) = \arg \max_{c_i} P(c_i | X = x)$$

损失函数定义：

- $L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$
- **期望风险**： $R(f) = E[L(Y, f(X))]$
- **决策函数**：
- **关系推导**：
 - 期望风险最小化等价于条件期望最小化
 - 通过数学变换，将最小化问题转化为后验概率最大化问题
 - 最终证明后验概率最大化准则与 0-1 损失下的期望风险最小化等价
- 后验概率最大化 37:23
 - 零一损失函数 37:27
 - **定义特点**：分类错误时损失为 1，正确时为 0
 - **应用场景**：适用于分类问题中的错误计数
 - 期望风险 37:44
 - 数学表达： $R(f) = E[L(Y, f(X))]$
 - **优化目标**：使期望风险最小化
 - **条件期望**：在贝叶斯方法中需要计算条件期望
 - 期望风险最小化与后验概率最大化 42:48
 - **转化过程**：
 - 将期望风险表达式展开为求和形式
 - 通过数学变换将最小化问题转化为最大化问题
 - 最终得到后验概率最大化的等价形式
 - **理论意义**：为朴素贝叶斯分类器提供了统计学习理论基础
- 3. 极大似然估计 43:09
 - 1) 朴素贝叶斯分类 43:29
 - **核心公式**：
 - **计算原理**：通过最大化后验概率实现分类，由于分母相同，只需比较分子部分（先验概率与条件概率的乘积）
 - 朴素贝叶斯分类：先验概率 43:33
 - **计算公式**： $P(Y = c_k) = \frac{\sum_{i=1}^N I(y_i \dot{=} c_k)}{N}$
 - **实际意义**：表示训练集中属于类别 c_k 的样本比例
 - **示例说明**：如巧克力例子中，A 盒概率=7/16，B 盒概率=9/16
 - 朴素贝叶斯分类：条件概率 44:17
 - **计算公式**： $P(X^{(j)} \dot{=} a_{jl} \mid Y \dot{=} c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} \dot{=} a_{jl}, y_i \dot{=} c_k)}{\sum_{i=1}^N I(y_i \dot{=} c_k)}$
 - **特征取值**：第 j 个特征有 S_j 种取值（ a_{j1} 到 a_{jS_j} ）
 - 朴素贝叶斯分类：参数估计与模型训练 45:17
 - **参数本质**：先验概率和条件概率是总体中的 $n+1$ 个参数
 - **估计方法**：通过训练集样本进行参数估计，即模型训练过程
 - **参数数量**：需要估计 $K \times (S_1 + S_2 + \dots + S_n)$ 个条件概率参数
 - 朴素贝叶斯分类：符号含义 45:50
 - **类别空间**： $c_k \in \{c_1, c_2, \dots, c_K\}$ ，共 K 个类别
 - **特征维度**：输入空间为 n 维， $x^{(j)}$ 表示第 j 个特征
 - **特征取值**：第 j 个特征有 S_j 种可能取值（ $l=1, 2, \dots, S_j$ ）
 - 例题 1: 巧克力 47:25



- **题目解析:**
 - 先验概率计算：A 盒概率=7/16，B 盒概率=9/16
 - 条件概率示例：A 盒中白色巧克力概率=2/7
 - **关键点:** 通过数数法计算概率，体现极大似然估计思想
 - 朴素贝叶斯分类：条件概率的估计值计算 48:22
 - **计算步骤:**
 - 分母：统计属于类别 c_k 的样本总数
 - 分子：统计同时满足特征 $x^{(j)}=a_{jl}$ 和 $y=c_k$ 的样本数
 - 示例**验证:** 巧克力例子中白色概率计算（2 块白色/7 块 A 盒）
 - 例题 2:巧克力 49:13
 - **题目解析:**
 - 特征分析：颜色为唯一特征，5 种取值（黑、白、棕、黄、红）
 - 具体计算：A 盒白色概率=2/7
 - **计算原理:** 条件概率=联合事件计数/条件事件计数
 - 2) 极大似然估计的原理 52:16
 - **核心思想**：概率最大化原则，即在所有可能的参数值中，选择使观测数据出现概率最大的那个参数值
 - **生活实例**：以购买苹果为例，当从未知箱子中取出 155 克苹果时，最可能属于中果类别（150 克），因为该类别下出现 155 克苹果的概率最大
- 数学定义**：设随机变量 x 的密度函数为 $f(X, \beta)$ ，对于独立同分布样本 $X_1, X_2, \dots, X_n$ ，其联合密度函数（似然函数）为：
- $L(x_1, \dots, x_n; \beta) = \prod_{i=1}^n f(x_i, \beta)$
- 参数估计**： β 的极大似然估计 $\hat{\beta}$ 满足：
- - **函数性质**：当样本取定值时，似然函数 $L(\beta)$ 是关于参数 β 的函数，而非关于样本的函数
- 3) 应用案例 57:25
 - 例题:极大似然估计计算

极大似然估计：例 1

Janneil博士

例 1: 假定一个盒子里有白球、黑球共三个，但不知白球和黑球分别有几个。如果有放回的从盒子里抽取三个球，发现第一个、第三个球是黑色的，第二个球是白色的。

问：如何估计盒中黑球所占比例 β ?

- 参数空间： $\Theta = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ 。

- 样本：从盒子中有放回的取球，第 i 次取出的结果记为

$$X_i = \begin{cases} 1, \text{Black} \\ 0, \text{White} \end{cases} \quad (i = 1, 2, 3)$$

- 概率函数：盒中黑球所占比例 β ， Y 的概率密度函数为

$$f(x) = \beta^x (1 - \beta)^{1-x}$$

题目解析

- **问题描述**：盒中有 3 个黑白球（数量未知），有放回抽取 3 次，结果为黑、白、黑，估计黑球比例 β

- **参数空间**： $\Theta = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$

变量定义：设 $X_i = 1$ 表示黑球， $X_i = 0$ 表示白球，概率函数为：

- $f(x) = \beta^x (1 - \beta)^{1-x}$

似然函数构建：根据观测结果 ($X_1 = 1, X_2 = 0, X_3 = 1$)，联合概率函数为：

- $L(\beta) = \beta^{x_1 + x_2 + x_3} (1 - \beta)^{3 - (x_1 + x_2 + x_3)} = \beta^2 (1 - \beta)$

- **参数求解**：通过枚举法计算各 β 值对应的似然函数值：

- $\beta = 0: L(0) = 0$

- $\beta = \frac{1}{3}: L(\frac{1}{3}) = \frac{2}{27}$

- $\beta = \frac{2}{3}: L(\frac{2}{3}) = \frac{4}{27}$

- $\beta = 1: L(1) = 0$

- **结论**：当 $\beta = \frac{2}{3}$ 时似然函数取得最大值，因此 $\hat{\beta} = \frac{2}{3}$

- **方法局限**：枚举法仅适用于参数空间有限且较小的情况，对于连续参数空间需采用求导等解析方法

- 联合概率函数： (X_1, X_2, X_3) 的联合概率函数为

$$L(x_1, x_2, x_3; \beta) = \beta^{x_1+x_2+x_3}(1-\beta)^{3-(x_1+x_2+x_3)}$$

- 似然函数： (X_1, X_2, X_3) 对应的结果为“第一个、第三个球是黑色的，第二个球是白色的”，即 $x_1 = x_3 = 1, x_2 = 0$,

$$L(\beta) = \beta^2(1-\beta)$$

- 参数求解：

β	0	$\frac{1}{3}$	$\frac{2}{3}$	1
$L(\beta)$	0	$\frac{2}{27}$	$\frac{4}{27}$	0

- 极大似然估计：

$$\hat{\beta} = \frac{2}{3}$$

4) 极大似然估计的数值计算方法

- 直接通过似然函数 $L(\beta)$ 求解

- 当 $L(\beta)$ 可微时，可通过方程组

$$\frac{\partial L(\beta)}{\partial \beta_1} = 0, \frac{\partial L(\beta)}{\partial \beta_2} = 0, \dots, \frac{\partial L(\beta)}{\partial \beta_m} = 0$$

求得 $L(\beta)$ 的极大值点。

- 当 $L(\beta)$ 不存在偏导数时，需要直接研究 $L(\beta)$ ，寻找最大值点。

- 通过对数似然函数 $\ln L(\beta)$ 求解 $\hat{\beta}$ 也是 $\ln L(\beta)$ 的最大值点，

- 当 $\ln L(\beta)$ 可微时，可通过下列方程组，

$$\frac{\partial \ln L(\beta)}{\partial \beta_1} = 0, \frac{\partial \ln L(\beta)}{\partial \beta_2} = 0, \dots, \frac{\partial \ln L(\beta)}{\partial \beta_m} = 0$$

求解，判断根是不是最大值点。

- 当 $\ln L(\beta)$ 不存在偏导数时，需要研究 $\ln L(\beta)$ ，寻找最大值点。

- **直接求解法**：当似然函数 $L(\beta)$ 可微时，通过解方程组 $\frac{\partial L(\beta)}{\partial \beta_1} = 0, \dots, \frac{\partial L(\beta)}{\partial \beta_m} = 0$ 求极大值点
- **特殊情况处理**：当 $L(\beta)$ 不可微时，需要直接研究函数本身寻找最大值点
- **对数转换法**： $\hat{\beta}$ 也是 $\ln L(\beta)$ 的最大值点，当可微时通过解 $\frac{\partial \ln L(\beta)}{\partial \beta_i} = 0$ 的方程组求解
- **对数不可微情况**：当 $\ln L(\beta)$ 不可微时，需直接研究对数似然函数寻找最大值点
- 5) 例题：正态分布的极大似然估计 01:05:36
- 问题描述

例 2：样本 X 服从正态分布：

$$f(x, \mu, \delta) = \frac{1}{\sqrt{2\pi\delta}} \exp \left\{ -\frac{(x - \mu)^2}{2\delta} \right\}$$

其中， $\delta = \sigma^2 \in (0, \infty), \mu \in (-\infty, \infty)$ 。

- 似然函数

$$L(x_1, \dots, x_N; \mu, \delta) = \left(\frac{1}{\sqrt{2\pi\delta}} \right)^N \exp \left\{ -\frac{\sum_{i=1}^N (x_i - \mu)^2}{2\delta} \right\}$$

- 对数似然函数

$$\ln L(\mu, \delta) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \delta - \frac{1}{2\delta} \sum_{i=1}^N (x_i - \mu)^2$$

- 分布形式：样本 x 服从正态分布 $f(x, \mu, \delta) = \frac{1}{\sqrt{\square}}$
- 参数说明：
 - μ ：均值，度量平均水平，取值范围 $(-\infty, \infty)$
 - $\delta = \sigma^2$ ：方差，度量数据波动情况，取值范围 $(0, \infty)$
- 参数特点：由于参数取值范围无限，无法通过遍历获得参数值
- 似然函数构建
 - 似然函数： $L(x_1, \dots, x_N; \mu, \delta) = \left(\frac{1}{\sqrt{\square}} \right)^N$
 - 对数似然函数： $\ln L(\mu, \delta) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \delta - \frac{1}{2\delta} \sum_{i=1}^N (x_i - \mu)^2$
 - 第一项为常量
 - 第二项关于 δ
 - 第三项关于 μ 和 δ
- 求解过程

- 对数似然方程组：

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = \frac{1}{\delta} \sum_{i=1}^N (x_i - \mu) = 0 \\ \frac{\partial \ln L}{\partial \delta} = -\frac{n}{2\delta} + \frac{1}{2\delta^2} \sum_{i=1}^N (x_i - \mu)^2 = 0 \end{cases} \quad (i = 1, 2, \dots, N)$$

- 极大似然估计：

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad \hat{\delta} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

求导方程组：

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = \frac{1}{\delta} \sum (x_i - \mu) = 0 \\ \frac{\partial \ln L}{\partial \delta} = -\frac{N}{2\delta} + \frac{1}{2\delta^2} \sum (x_i - \mu)^2 = 0 \end{cases}$$

解析解：

- $\hat{\mu} = \frac{1}{N} \sum x_i = \bar{x}$ (样本均值)
- $\hat{\delta} = \frac{1}{N} \sum (x_i - \bar{x})^2$ (样本方差)

6) 数值解的局限性

- 直接通过似然函数 $L(\beta)$ 求解
 - 当 $L(\beta)$ 可微时，可通过方程组

$$\frac{\partial L(\beta)}{\partial \beta_1} = 0, \frac{\partial L(\beta)}{\partial \beta_2} = 0, \dots, \frac{\partial L(\beta)}{\partial \beta_m} = 0$$

求得 $L(\beta)$ 的极大值点。

- 当 $L(\beta)$ 不存在偏导数时，需要直接研究 $L(\beta)$ ，寻找最大值点。

- 通过对数似然函数 $\ln L(\beta)$ 求解 $\hat{\beta}$ 也是 $\ln L(\beta)$ 的最大值点，
 - 当 $\ln L(\beta)$ 可微时，可通过下列方程组，

$$\frac{\partial \ln L(\beta)}{\partial \beta_1} = 0, \frac{\partial \ln L(\beta)}{\partial \beta_2} = 0, \dots, \frac{\partial \ln L(\beta)}{\partial \beta_m} = 0$$

求解，判断根是不是最大值点。

- 当 $\ln L(\beta)$ 不存在偏导数时，需要研究 $\ln L(\beta)$ ，寻找最大值点。

● 解析解不可得的情况：

- 方程组可微但无法解出具体表达式
- 函数不存在偏导数，无法得到具体表达

● 解决方法：需要采用其他数值计算方法（如迭代法、优化算法等）

7) 迭代法 01:09:04

● 迭代法的起源与描述

○

Grevnd of Artes

*Goose at this wecke as kappo doth lead.
By chance to truth he may preceed.
And firste wecke by the question,
Although no truth he therein be den.
Suche falsheho is so good a grounde,
That truth by it will seeme be founde.
From many hets to many more,
From to fewe take to fewe alow.
With to much tryne to fewe againe,
To to fewe odds to manye plaine.
In crossewaies multiplye contrary kinde,
All truth by falsheho feo to fynde.*



Robert Recorde
(1510 ~ 1558)

艺术基础

按照自己的意愿猜一个答案。
运气好的话，你可能会接近真理。
对问题进行初次计算，
尽管真理仍然遥不可及。
这种错误是良好的基础，
你很快就会发现真相。
走过的道路越来越多，
离目标的距离越来越远。
再长的道路也会走到尽头，
再小的水滴也能聚成大海。
不同种类交叉相乘，
错误的方法也可以找到真理。

- **历史背景**：由英国数学家 Robert Recorde(1510-1558)在《艺术基础》中用诗歌形式描述

- **核心思想**：通过重复修正逐步逼近真理，如诗中所述"走过的道路越来越多，离目标的距离越来越远"

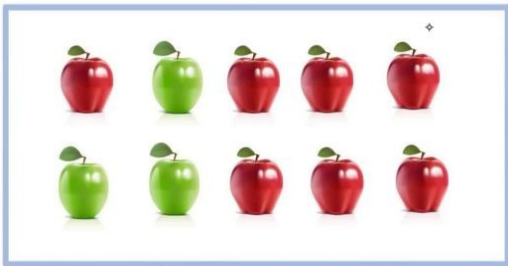
- 初始步骤：先给定参数初始值（诗中"按照自己的意愿猜一个答案"）

● 迭代法的原理与步骤 01:09:27

- **实现过程：**
 - **初始设定：**给定参数初始值 β_0 （对应诗中"初次计算"）
 - **目标函数：**设定目标函数 Q ，将 β_0 代入计算
 - **迭代修正：**根据计算结果修正得到 β_1 ，再代回计算（诗中"错误是良好基础"）
- **收敛特性：**通过不断修正最终接近真实值（诗中"再小的水滴也能聚成大海"）
- 迭代法的应用与解析解 01:11:35
 - **遍历法：**
 - **原理：**遍历参数空间所有参数，选择使似然函数最大的参数
 - **局限：**当参数空间元素过多或无限时不可行
 - **解析解：**利用高等数学知识直接求解方程（最优方法）
 - **数值解：**当解析解不可得时采用迭代法逼近
- 迭代法的数值解 01:12:13
 - **计算优势：**现代计算机可快速完成百万次迭代
 - **实现要点：**
 - 需要设置合理的停止条件（如误差阈值）
 - 初始值选择影响收敛速度
 - **历史对比：**16 世纪手工计算困难，现代计算效率大幅提升
- 8) 应用案例 01:12:40
 - 例题：数苹果概率问题

朴素贝叶斯：苹果

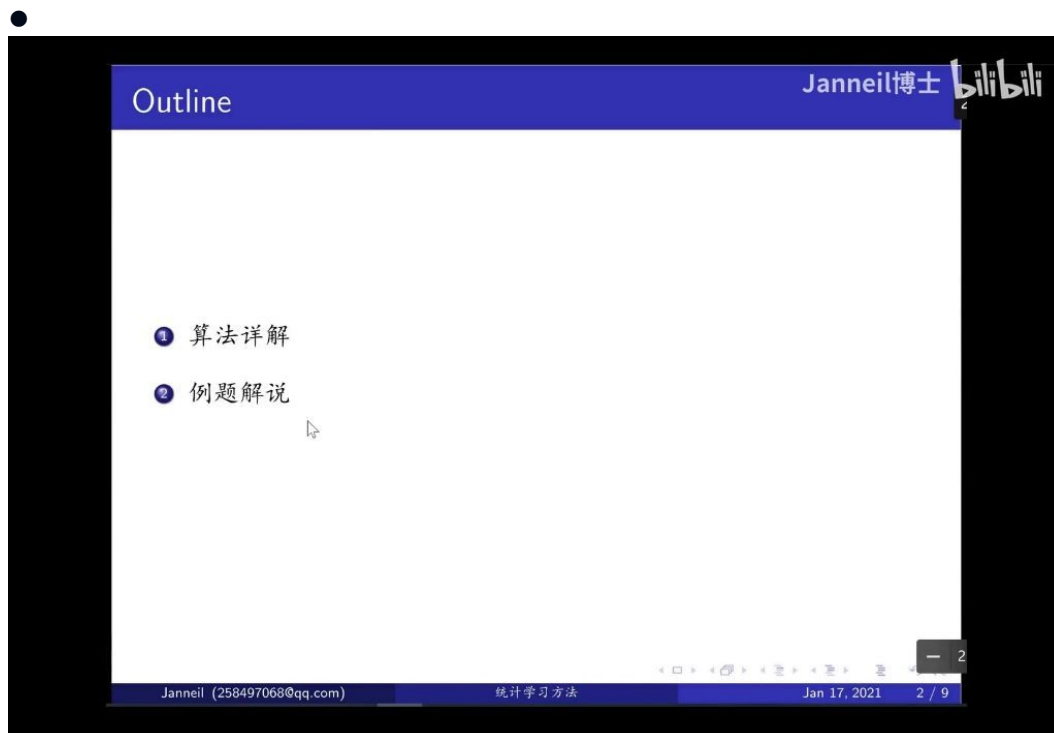
红、绿 |



问：抽出一个苹果，属于红苹果的概率是多少？

Janneil (258497068@qq.com) 统计学习方法 Jan 09, 2021 18 / 20

- **题目解析：**
 - **问题描述：**10 个苹果（7 红 3），求随机抽取苹果概率
 - **直观计算：** $P(\text{红})=7/10$ ， $P(\text{绿})=3/10$
 - **似然函数：** $L(p) = p^7(1-p)^3$ （ p 为红苹果概率）
 - **求导验证：**
 - 对 $L(p)$ 求导得： $7(1-p)-3p=0$
 - 解得 $p=7/10$ ，验证直观计算正确
 - **方法本质：**通过计数得到的概率就是极大似然估计
 - **推广结论：**分类问题中先验概率的极大似然估计就是类别频率
4. 朴素贝叶斯法 01:16:43
- 1) 算法详解 01:17:04



- **算法输入**：训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ，实例 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$
 - **算法输出**：实例 x 所属类别 y
 - **计算步骤**：
 - **先验概率计算**：使用极大似然估计， $P(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k)}{N}$ ，即每个类别的实例数除以总实例数
 - **条件概率计算**： $P(X^{(j)} = a_{jl} \vee Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k)}{\sum_{i=1}^N I(y_i = c_k)}$ ，即在某类别下特征取特定值的比例
 - **后验概率计算**：对于新实例 x ，计算 $P(Y = c_k) \prod_{j=1}^n P(X^{(j)} = x^{(j)} \vee Y = c_k)$ ，比较不同类别的值，取最大者作为预测类别
- 2) 例题：朴素贝叶斯算法演示 01:20:04

输入：训练集：

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

输出：实例点 $x = (2, S)^T$ 的类标记 y .

● 题目解析：

○ 数据集：15 个实例，特征 $x^{(1)}$ 取值 $\{1, 2, 3\}$ ， $x^{(2)}$ 取值 $\{S, M, L\}$ ，类别 $y \in \{-1, 1\}$

○ 待预测实例： $x = (2, S)$

○ 计算过程：

■ 先验概率：

- $P(Y=1) = 9/15$
- $P(Y=-1) = 6/15$

■ 条件概率 ($Y=1$ 时)：

- $P(X^{(1)}=2 \vee Y=1) = 3/9$
- $P(X^{(2)}=S \vee Y=1) = 1/9$

■ 条件概率 ($Y=-1$ 时)：

- $P(X^{(1)}=2 \vee Y=-1) = 2/6$
- $P(X^{(2)}=S \vee Y=-1) = 3/6$

■ 后验概率分子：

- $Y=1 : (9/15) \times (3/9) \times (1/9) = 1/45$
- $Y=-1 : (6/15) \times (2/6) \times (3/6) = 1/15$

○ 结果：比较后验概率分子， $1/45 < 1/15$ ，故预测 $y=-1$

3) 贝叶斯估计

- 先验概率的贝叶斯估计

$$P_{\lambda}(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k) + \lambda}{N + K\lambda}$$

- 条件概率的贝叶斯估计

$$P_{\lambda}(X^{(j)} = a_{jl} | Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k) + \lambda}{\sum_{i=1}^N I(y_i = c_k) + S_j \lambda}$$

注: $\lambda \geq 0$ $\lambda = 0$ 时为极大似然估计, $\lambda = 1$ 时为拉普拉斯平滑(Laplacian Smoothing)。

- 改进原因: 极大似然估计在训练集中未出现的特征组合会导致概率为 0, 影响分类结果

先验概率估计:

- $P_{\lambda}(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k) + \lambda}{N + K\lambda}$

- K 为类别数
- $\lambda \geq 0$, $\lambda = 0$ 时为极大似然估计, $\lambda = 1$ 时为拉普拉斯平滑

条件概率估计:

- $P_{\lambda}(X^{(j)} = a_{jl} | Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k) + \lambda}{\sum_{i=1}^N I(y_i = c_k) + S_j \lambda}$

- S_j 为第 j 个特征的可能取值数

- 平滑思想: 通过加入 λ 避免零概率问题, 使概率分布更"平滑"

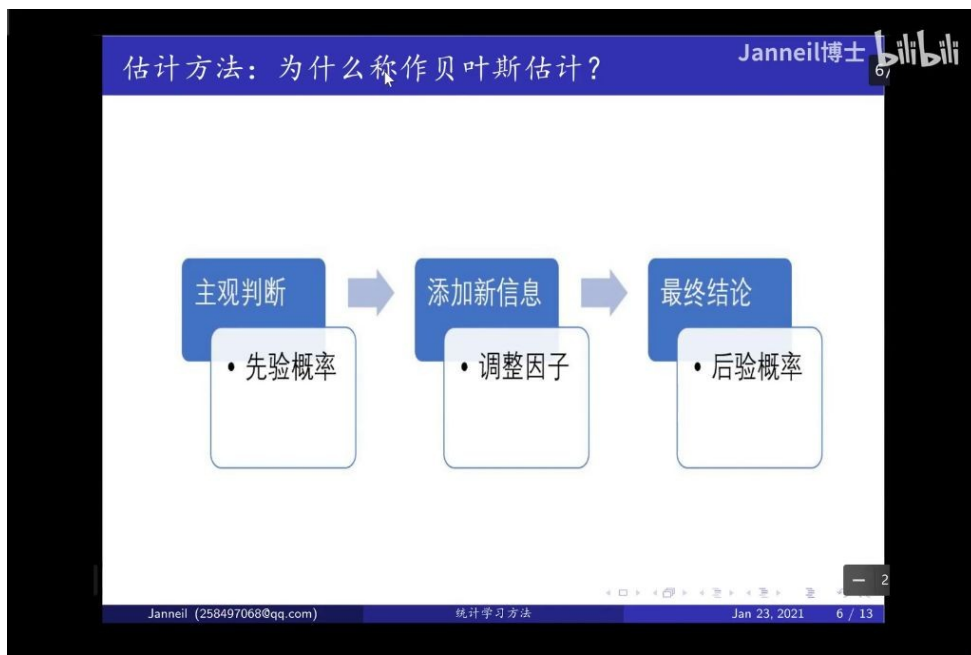
- 贝叶斯本质: 将参数视为随机变量, 通过先验分布和样本信息计算后验分布

5. 贝叶斯估计 01:31:13

1) 估计方法 01:33:03

- 例题: 贝叶斯公式求解 01:42:10

- 贝叶斯估计基础概念



■ **核心思想**：基于贝叶斯思维，通过先验知识和新观测数据相结合来更新概率估计

■ **计算流程**：

- 主判断：初始先概率
- 添加新信息：观测数据
- 调整因子：通过贝叶斯公式更新
- 最终结论：后验概率分布

○ 贝叶斯公式解析

估计方法：贝叶斯估计

Janneil博士 Bilibili

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

例：Y 分为两类 c_1, c_2 ，相应参数为 θ_1 ，假设参数服从 Beta 分布 $Be(\alpha, \beta)$ ，求参数 θ_1 的贝叶斯估计。

- 参数 θ_1 的先验概率

$$f(\theta_1; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta_1^{\alpha-1} (1 - \theta_1)^{\beta-1}$$

- 已知 θ_1 时 Y 的条件概率

$$g(Y|\theta_1) = \begin{cases} \theta_1, & Y = c_1 \\ 1 - \theta_1, & Y = c_2 \end{cases}$$

Janneil (258497068@qq.com) 统计学习方法 Jan 23, 2021 7 / 13

■ 公式表达： $P(\theta \vee Y) = \frac{P(Y \vee \theta)P(\theta)}{P(Y)}$

■ 分量解释：

- **后验概率**： $P(\theta \vee Y)$ 表示观测到数据 Y 后参数 θ 的概率
- **似然函数**： $P(Y \vee \theta)$ 是给定参数时数据的概率
- **先验概率**： $P(\theta)$ 是参数的初始概率分布

- **边缘概率**： $P(Y)$ 作为归一化常数，保证概率总和为1
- **计算技巧**：实际计算时可忽略分母 $P(Y)$ ，只需最大化分子 $P(Y \vee \theta)P(\theta)$
- 二分类贝叶斯估计实例

估计方法：贝叶斯估计 Janneil博士 bilibili

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}, \quad f(\theta_1; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta_1^{\alpha-1} (1-\theta_1)^{\beta-1}, \quad g(Y|\theta_1) = \begin{cases} \theta_1, & Y = c_1 \\ 1-\theta_1, & Y = c_2 \end{cases}$$

$\{y_1, \dots, y_n\} \quad c_1 \quad N_1, \quad c_2 \quad N - N_1$

$$\Rightarrow \theta_1^{N_1 + \alpha - 1} (1 - \theta_1)^{N - N_1 + \beta - 1} \cdot [(N_1 + \alpha - 1) \cdot (1 - \theta_1) - (N - N_1 + \beta - 1) \theta_1] = 0$$

↑↑

Janneil (258497068@qq.com) 统计学习方法 Jan 23, 2021 8 / 13

■ **问题设定：**

- 随机变量 Y 分为两类 c_1, c_2
- 参数 θ_1 表示 $P(Y = c_1)$
- 先验分布： $\theta_1 \sim \text{Beta}(\alpha, \beta)$

■ **概率分布：**

- 先验密度： $f(\theta_1; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta_1^{\alpha-1} (1-\theta_1)^{\beta-1}$
- 条件概率： $g(Y \vee \theta_1) = \begin{cases} \theta_1, & Y \in c_1 \\ 1-\theta_1, & Y \in c_2 \end{cases}$

■ **求解步骤：**

- 计算样本中 c_1 类数量 n_1 和 c_2 类数量 $n - n_1$
- 构建后验概率分子： $\theta_1^{n_1 + \alpha - 1} (1 - \theta_1)^{n - n_1 + \beta - 1}$
- 通过求导最大化后验概率

■ **估计结果：** $\hat{\theta}_1 = \frac{n_1 + \alpha - 1}{n + \alpha + \beta - 2}$

■ **特例分析：**当 $\alpha = \beta$ 且 $\alpha - 1 = \lambda$ 时，简化为 $\frac{n_1 + \lambda}{n + 2\lambda}$

- 均匀先验下的贝叶斯估计

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

例：Y 分为两类 c_1, c_2 ，相应参数为 θ_1 ，假设参数服从均匀分布 $U(0, 1)$ ，求参数 θ_1 的贝叶斯估计。

- 参数 θ_1 的先验概率

$$f(\theta_1) = 1$$

- 已知 θ_1 时 Y 的条件概率

$$g(Y|\theta_1) = \begin{cases} \theta_1, & Y = c_1 \\ 1 - \theta_1, & Y = c_2 \end{cases}$$

■ 先验设定： $\theta_1 \sim U(0, 1)$ ，即 $f(\theta_1) = 1$

■ 求解过程：

- 后验概率分子简化为 $\theta_1^{n_1}(1 - \theta_1)^{n - n_1}$
- 通过求导得极值点 $\hat{\theta}_1 = \frac{n_1}{n}$

■ 理论意义：

- 均匀先验对应无信息先验
- 结果与极大似然估计一致
- 体现了频率学派的思想：完全依赖数据说话

○ 两类学派对比

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}, \quad f(\theta_1) = 1, \quad g(Y|\theta_1) = \begin{cases} \theta_1, & Y = c_1 \\ 1 - \theta_1, & Y = c_2 \end{cases}$$

■ 贝叶斯学派：

- 引入先验分布
- 通过数据更新后验

- 参数视为随机变量
 - 率学派：
 - 不使用先验信息
 - 完全依赖数据
 - 参数视为固定未知量
 - 联系：当采用均匀先验时，贝叶斯估计退化为频率学派的极大似然估计
- 2) 平滑思想是什么 01:58:40
- 平滑思想与贝叶斯估计 01:58:42
 - **核心公式**： $P_{\lambda}(Y=c_k) = \frac{\sum_{i=1}^N I(y_i=c_k) + \lambda}{N + K\lambda}$
 - **数学变换**：通过代数变形可表示为 $\theta_k \times (N + K\lambda)$ $\dot{=} n_k + \lambda$ ，其中 θ_k 是 $Y=c_k$ 的概率估计
 - **组合解释**：贝叶斯估计实质上是极大似然估计 ($\frac{n_k}{N}$) 与先验概率 ($\frac{1}{K}$) 的凸组合
 - **参数意义**： λ 控制平滑强度， K 是类别总数， N 是样本总数
 - 平滑思想与正则化 02:00:57
 -

估计方法：平滑思想是什么？ Janneil博士 Bilibili

- 贝叶斯估计：

$$P_{\lambda}(Y=c_k) = \frac{\sum_{i=1}^N I(y_i=c_k) + \lambda}{N + K\lambda}$$
- 正则化：

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f)$$

Janneil (258497068@qq.com) 统计学习方法 Jan 23, 2021 11 / 13

- **正则化目标**： $\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f)$
 - **双重作用**：
 - 第一部分：经验风险最小化（对应极大似然估计）
 - 第二部分：模型复杂度惩罚项（对应先验概率约束）
 - **防过拟合机制**：通过 λ 平衡拟合优度与模型复杂度，避免参数估计波动过大
 - 拉普拉斯平滑 02:02:00
 - **特殊形式**：当 $\lambda=1$ 时的贝叶斯估计称为拉普拉斯平滑
 - **历史背景**：为解决零概率问题，在分子加 1，分母加类别数 K
 - **大样本特性**：当 $N \rightarrow \infty$ 时，平滑效果减弱，估计值收敛于 $\frac{n_k}{N}$
 - **实际意义**：防止极端估计（如女儿国例子中男性比例估计为 0 的情况）
- 3) 应用案例 02:03:01
- 例题：贝叶斯估计计算

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x^{(1)}$	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3
$x^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

	$X^{(1)}$			$X^{(2)}$		
	1	2	3	S	M	L
$Y = 1$						
$Y = -1$						

- 类别数 $K=2$ ($Y \in \{-1, 1\}$)
- 特征 1 取值空间 $S_1=3$ ($\{1, 2, 3\}$)
- 特征 2 取值空间 $S_2=3$ ($\{M, L\}$)

- $P(Y=1)=\frac{9+1}{15+2}=\frac{10}{17}$
- $P(Y=-1)=\frac{6+1}{15+2}=\frac{7}{17}$

- $P(X_1=1 \vee Y=1) = \frac{2+1}{9+3} = \frac{3}{12}$
- $P(X_1=2 \vee Y=1) = \frac{3+1}{9+3} = \frac{4}{12}$
- $P(X_1=3 \vee Y=1) = \frac{4+1}{9+3} = \frac{5}{12}$

- $P(X_2=S \vee Y=1) = \frac{1+1}{9+3} = \frac{2}{12}$
- $P(X_2=M \vee Y=1) = \frac{4+1}{9+3} = \frac{5}{12}$

- $P(Y=1 \vee X) \propto \frac{10}{17} \times \frac{4}{12} \times \frac{2}{12} = \frac{5}{153}$
- $P(Y=-1 \vee X) \propto \frac{7}{17} \times \frac{3}{9} \times \frac{4}{9} = \frac{28}{459}$

○ **实际应用：**

- 女儿国人口比例估计：当 $N=50$
 - 女性比例 $p=\frac{50+1}{50+2}\approx 0.98$
 - 男性比例 $q=\frac{0+1}{50+2}\approx 0.02$
- 避免极端结论：即使样本中无男性，估计值仍保持合理非零值

二、知识小结

知识点	核心内容	考试重点/易混淆点	难度系数
贝叶斯定理	逆概率思维，通过先验概率和条件概率计算后验概率	先验概率与后验概率的区别	☆☆☆
朴素贝叶斯法	基于特征条件独立假设的贝叶斯分类方法	独立性假设在实际应用中的局限性	☆☆☆☆
极大似然估计	通过概率最大化思想进行参数估计	样本偏差导致的估计偏差问题	☆☆☆☆
贝叶斯估计	引入先验分布的参数估计方法，避免零概率问题	拉普拉斯平滑的数学原理	☆☆☆☆ ☆
条件概率	在已知条件下事件发生的概率计算	联合概率与条件概率的关系	☆☆
分类决策准则	后验概率最大化与期望风险最小化的等价性	损失函数的选择对分类结果的影响	☆☆☆☆
特征独立性	朴素贝叶斯法的核心假设	特征相关的理 方法	☆☆☆
参数估计方法对比	极大似然估计与贝叶斯估计的优缺点比较	先验信息在估计中的作用	☆☆☆☆