Bayesian Theorem 总结

Bayesian Theorem（贝叶斯定理）是一种概率推理方法，用于更新事件的概率，基于新的证据或信息。

- $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$
  - **P(A|B)**: 在事件 B 发生的条件下，事件 A 发生的后验概率（Posterior Probability）。
  - **P(B|A)**: 在事件 A 发生的条件下，事件 B 发生的似然（Likelihood）。
  - **P(A)**: 事件 A 的先验概率（Prior Probability）。
  - **P(B)**: 事件 B 的边际概率（Marginal Probability）。

Bayesian Theorem 的核心思想是通过结合先验知识和新证据，动态调整对事件的概率估计。

条件概率（Conditional Probability）

条件概率是指在已知某一事件发生的条件下，另一个事件发生的概率。其公式为：

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

其中：

- **P(A|B)**: 在事件 B 发生的条件下，事件 A 发生的概率。
- **P(A ∩ B)**: 事件 A 和事件 B 同时发生的概率。
- **P(B)**: 事件 B 发生的概率。

条件概率是 Bayes' Theorem 的基础，用于描述事件之间的依赖关系。

联合概率

a、b 两个事件同时发生的概率，称为它们的**联合概率**，记作 $P(A \cap B)$ 或 $P(A, B)$。

- 其定义为：$P(A \cap B)$ 表示事件 A 和事件 B 同时发生的概率。
- 如果已知条件概率 $P(A|B)$ 和 $P(B)$，则有： $P(A \cap B) = P(A|B) \cdot P(B)$
- 同理，也有：$P(A \cap B) = P(B|A) \cdot P(A)$

全概率公式（Law of Total Probability）

设 $B_1, B_2, \ldots, B_n$ 是一组两两不相交且覆盖全集的事件（即一个完备事件组），则对于任意事件 A，有：

- $P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$
- 全概率公式用于将事件 A 的概率分解为在不同条件 B_i 下的条件概率与 B_i 的概率的加权和。常用于贝叶斯推理和概率分布的分解。

朴素贝叶斯分类推理步骤及公式

1. **计算先验概率（Prior Probability）**
   每个类别 C_k 的先验概率：$P(C_k) = \frac{\text{类别 } C_k \text{ 的样本数}}{\text{总样本数}}$

2. **计算条件概率（Likelihood）**
   对于每个特征 x_i，计算其在类别 C_k 下的条件概率： $P(x_i|C_k) = \frac{\text{类别 } C_k \text{ 中 } x_i \text{ 的样本数}}{\text{类别 } C_k \text{ 的样本数}}$

3. **计算联合概率（Naïve Bayes 假设条件独立）**
   假设各特征条件独立，联合概率为： $P(x_1, x_2, \ldots, x_n|C_k) = \prod_{i=1}^{n} P(x_i|C_k)$

4. **计算后验概率（Posterior Probability）**
   对每个类别，利用贝叶斯公式计算后验概率：

   - $P(C_k|x_1, x_2, \ldots, x_n) = \frac{P(x_1,x_2,\ldots,x_n|C_k)\cdot P(C_k)}{P(x_1,x_2,\ldots,x_n)}$
   - 由于分母对所有类别相同，实际分类时只需比较分子部分大小。

5. **分类决策**
   选择后验概率最大的类别作为预测结果：$\hat{C} = \arg\max_{C_k} \left[P(C_k) \prod_{i=1}^{n} P(x_i|C_k)\right]$

6. **分类决策**
   比较后验概率，选择概率较大的类别作为预测结果。

Bayesian Classification 总结

应用场景

1. **文本分类**: 如垃圾邮件检测。
2. **情感分析**: 基于文本内容预测情感类别。
3. **医学诊断**: 基于症状预测疾病。

朴素贝叶斯分类的合理性

朴素贝叶斯分类的核心假设是特征之间的条件独立性，即在给定类别的条件下，每个特征对类别的贡献是独立的。优点是计算效率高，适用于高维数据，但其独立性假设在某些场景下可能不成立.

1. **简化计算**
   条件独立性假设**将联合概率的计算从指数级复杂度简化为线性复杂度**，使得朴素贝叶斯分类在高维数据中依然高效。

2. **鲁棒性**
   即使独立性假设不完全成立，朴素贝叶斯分类在许多实际场景中仍能提供良好的分类性能。这是因为它在概率估计中对误差具有一定的容忍度。

Bayesian Belief Networks (BBN)

Bayesian Belief Networks (BBN) 是一种扩展的贝叶斯推理方法，用于显式捕获变量之间的依赖关系。它通过图形结构（Directed Acyclic Graph, DAG）和 Conditional Probability Tables (CPTs) 表示随机变量及其条件依赖性。

1. **Directed Acyclic Graph (DAG)**

   - **Nodes**: 表示随机变量。
   - **Edges**: 表示条件依赖性。

- 如果存在边 $A \to B$，则表示 B 条件依赖于 A。
- **注意**: DAG 是基于领域知识或数据分析构建的，用于表示变量之间的依赖关系，而不是直接从 CPT 表格中推导出来的。CPT 则用于量化这些依赖关系。

2. **Conditional Probability Tables (CPTs)**

CPT 示例及 BBN 使用方法

以下是一个简单的 CPT 示例，假设节点 A 的父节点为 B 和 C：

B	C	$P(A = 0/B, C)$	$P(A = 1/B, C)$
0	0	0.7	0.3
0	1	0.6	0.4
1	0	0.2	0.8
1	1	0.9	0.1

解释：

1. 每一行表示父节点 B 和 C 的一个可能组合。
2. 每一列表示目标节点 A 在该条件下的概率分布。例如：
   - 当 $B = 1$ 且 $C = 0$ 时：
     - $P(A = 0|B = 1, C = 0) = 0.2$
     - $P(A = 1|B = 1, C = 0) = 0.8$

BBN 步骤

1. **明确目标变量和已知变量**
   确定需要分类的目标变量（如 D）以及已知的条件变量（如 A, B, C）。

2. **构建依赖关系**
   根据 DAG 和 CPTs，明确目标变量与条件变量之间的依赖关系。例如：

   - $P(D|A, B, C)$ 表示在已知 A, B, C 的条件下，D 的后验概率。

3. **应用贝叶斯公式**
   使用贝叶斯公式计算后验概率：

   - $P(D|A, B, C) = \frac{P(A,B,C|D) \cdot P(D)}{P(A,B,C)}$

4. **分解联合概率**
   根据 BBN 的依赖关系，分解联合概率。例如：

   - 在 Naïve Bayes Classifier 中：$P(A, B, C|D) = P(A|D) \cdot P(B|D) \cdot P(C|D)$
   - 在 BBN 中：$P(A, B, C|D) = P(A|D) \cdot P(B|A, D) \cdot P(C|A, B, D)$

5. **简化计算**
   根据条件独立性假设，进一步简化计算。例如：

- $P(A, B, C|D) = P(A|D) \cdot P(B) \cdot P(C|A, B)$

6. **计算后验概率**
   使用链式规则和 CPTs 计算后验概率。例如：
   - $P(A_1, A_2, \ldots, A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1, A_2) \cdot \ldots \cdot P(A_n|A_1, A_2, \ldots, A_{n-1})$