姓名和学号：李胜志 2210180232

章节名称：朴素贝叶斯算法的原理与流程

知识目标：理解朴素贝叶斯的原理与流程。

能力目标：能够使用朴素贝叶斯算法计算样本属于每个类别的概率。

素质目标：养成分析问题、事前规划的良好习惯。

知识重点：朴素贝叶斯算法的原理与流程。

知识难点：对贝叶斯定理的理解。

朴素贝叶斯算法是一种基于贝叶斯定理的简单概率分类器，它在处理分类问题上尤其是在文本分类（如垃圾邮件和情感分析）中得到了广泛的应用。下面是朴素贝叶斯算法的原理与流程：

朴素贝叶斯算法原理：

朴素贝叶斯的原理是根据数据集中已有数据得到先验概率，然后求解待测样本属于每个类别的后验概率，哪个类别概率高就将新样本判定为哪个类别。

朴素贝叶斯算法基于贝叶斯定理，该定理可以表示为：

其中，

是后验概率，即在发生的条件下发生的概率。

是条件概率，即在发生的条件下发生的概率。

和是先验概率，即和独立发生的概率。

假设用表示特征，和表示两个类别，则样本数据属于和的后验概率分别为：

若数据集的特征个数和类别有多个，分别用和表示，则样本数据属于的后验概率为：

对于用一个数据集而言，每个类别的后验概率（即分母）是一样的，只需计算出分子即可。

朴素贝叶斯算法假设特征之间相互独立，特征直接不存在依赖关系，则以下式子成立：

根据数据集里的样本即可算出 的值，经过比较即可得到新样本属于某个类别的概率。

朴素贝叶斯算法的流程：

1. 假设为一个分类样本，为样本的特征。
2. 有类别集合
3. 分别计算每个类别的后验概率：，等价于求,,…,的值。
4. 如果,则

朴素贝叶斯的优缺点：

优点：

1. 素贝叶斯算法基于贝叶斯定理，逻辑简单，概念直观，易于理解和实现。
2. 算法稳定，具有较好的健壮性。
3. 对小规模的数据表现很好，能处理多分类任务。
4. 计算效率高。特征条件独立的假设，算法在计算后验概率时只需要考虑单个特征的概率分布，计算速度快，适合处理大规模数据集。
5. 朴素贝叶斯算法提供的概率解释有助于理解分类决策，这对于需要解释模型决策过程的场景非常有用。

缺点

1. 朴素贝叶斯算法假设所有特征在给定类别的条件下相互独立，这在现实世界的许多问题中并不成立，因此可能会影响分类的准确性。
2. 算法的性能很大程度上取决于特征表示方式，如果特征之间相关性较高，算法效果可能不佳。
3. 不适合处理特征相关性高的数据，当特征之间存在较强的依赖关系时，朴素贝叶斯算法的性能会受到影响。