姓名和学号：李胜志 2210180232

章节名称：朴素贝叶斯算法常见问题及其解决方法

知识目标：了解朴素贝叶斯算法常见问题及其解决方法。

能力目标：能够使用拉普拉斯平滑等方法解决朴素贝叶斯算法常见问题。

素质目标：养成分析问题、事前规划的良好习惯。

知识重点：拉普拉斯平滑等解决方法。

知识难点：朴素贝叶斯算法存在问题的原因及解决方法。

朴素贝叶斯算法常见问题及其解决方法：

1. 零概率问题：

当数据集中包含离散属性时，如果某些类别在训练集中没有出现，会导致概率计算为零，影响分类结果。

解决方法：使用拉普拉斯平滑方法，即为所有概率值分子加1，分母加特征个数。如果样本量很大的情况下，每个分量的计数加1造成的估计概率变化可以忽略不计，这样可以有效地避免零概率问题。

1. 溢出问题

在文本分类中，特征空间可能非常大，导致数据变得非常稀疏。

解决方法：结合使用朴素贝叶斯和其他算法，例如使用SVM或深度学习模型。

1. 特征独立性无法满足的问题

朴素贝叶斯算法假设特征之间相互独立，但在现实世界中，这个假设往往不成立。

解决方法：选择相互独立或者近似独立的特征。使用半朴素贝叶斯分类模型来表示特征之间的依赖关系，该模型的基本思想是假设样本的每个特征都可以关联一个对其产生影响的特征，基于此再进行计算。

1. 大规模数据集上的计算效率较差

在处理大规模数据集时，朴素贝叶斯算法的计算效率可能成为瓶颈。

解决方法：利用并行计算资源来加速概率的计算；对于新数据，使用增量学习更新模型，而不是重新训练。