1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：P77拉普拉斯修正的作用是什么，是否有可能对原先的预测结果造成影响？

讨论后的理解：拉普拉斯修正能够有效避免因训练集样本不充分而导致估值概率为零的问题，而且因为n的取值是c的类别数，修正是比较平滑的。并且在训练集变大的时候，修正过程所引入的先验的影响也会变得可忽略，使得估值渐趋向于实际概率值。

1. 提出的问题2： “一个在测试数据中出现的属性值可能不在训练数据中出现”时使用拉普拉斯和“丢失的数据可以忽略”有冲突吗？

讨论后的理解：是没有冲突的。丢失数据可以忽略的意思是在大量的训练数据中，丢失的数据比较少，对结果影响不大，就可以选择忽略。而拉普拉斯修正是为了避免零概率的发生，如果这个数据丢失了后使某一属性值因此不在训练数据中出现，那么说明这个数属性出现的概率很小，接近于零，可以用拉普拉斯修正。

1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题2：条件独立假设是什么意思？对于朴素贝叶斯分类有什么意义？为什么实际操作中做出这样的假设之后分类效果依旧很好？

自己的理解：朴素贝叶斯算法假设了数据集属性之间是相互独立的，因此算法的逻辑性十分简单，并且算法较为稳定，当数据呈现不同的特点时，朴素贝叶斯的分类性能不会有太大的差异。换句话说就是朴素贝叶斯算法的健壮性比较好，对于不同类型的数据集不会呈现出太大的差异性。当数据集属性之间的关系相对比较独立时，朴素贝叶斯分类算法会有较好的效果。属性独立性的条件同时也是朴素贝叶斯分类器的不足之处。数据集属性的独立性在很多情况下是很难满足的，因为数据集的属性之间往往都存在着相互关联，如果在分类过程中出现这种问题，会导致分类的效果大大降低。

1. 问题3：先验概率和后验概率有什么联系？

自己的理解：先验概率是指根据以往经验和分析得到的概率，如全概率公式，它往往作为"由因求果"问题中的"因"出现的概率。在贝叶斯统计推断中，不确定数量的先验概率分布是在考虑一些因素之前表达对这一数量的置信程度的概率分布。例如，先验概率分布可能代表在将来的选举中投票给特定政治家的选民相对比例的概率分布。未知的数量可以是模型的参数或者是潜在变量。后验概率是信息理论的基本概念之一。在一个通信系统中，在收到某个消息之后，接收端所了解到的该消息发送的概率称为后验概率。

后验概率的计算要以先验概率为基础。后验概率可以根据通过贝叶斯公式，用先验概率和似然函数计算出来。

1. （必填）读书计划

1、本周完成的内容章节：3.6-3.7

2、下周计划：3.8

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

1、读书摘要及理解（选做）

3.6朴素贝叶斯分类是一种十分简单的分类算法，叫它朴素贝叶斯分类是因为这种方法的思想真的很朴素。朴素贝叶斯的思想基础是这样的：对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。为避免因训练集样本不充分而导致概率估值为零，可以使用拉普拉斯修正。

3.7朴素贝叶斯文本分类把每篇文档看作一袋词，每个文档可以看成多项式分布生成。然后进行参数估计，这是基于经验的计数得到的。为了解决不常出现的单词的零概率估计，要使用拉普拉斯光滑。确定估计的参数后计算测试文档属于每个类别的概率。如果最后的分类器是将每个文档放到一个类别中，那么有最大后验概率的那个类别就是最后选择。