方骏-2020年3月29 日-读书报告

1. 自己提出的问题的理解：
2. 提出的问题1：请问如何用最大似然估计证明两个公式？

讨论后的理解：本公式的证明比较复杂，这里就不再具体详细的列举出证明过程，但是对于最大似然的理解可以说明一下，即最大似然估计的目的就是：利用已知的样本结果，反推最有可能（最大概率）导致这样结果的参数值。就是说，现在得到的样本结果是未知的参数导致的，从结果推参数，认为出现了这种结果就是因为这种结果出现的可能性大，所以未知的参数就是推出有最大概率的参数。

1. 别人提出的问题的理解：
2. 问题2：朴素贝叶斯方法效率较高，体现在哪些方面呢？

自己的理解：首先，朴素贝叶斯方法就是简单的计算一系列概率，不需要过于复杂的训练过程，测试过程也很简单，因此所花时间短，而由于这是基于古典概率模型，有理有据，而且应用条件独立假设，使得对于问题方法更加简单，可以保证准确率和速度的兼容，这种朴素的概率思想才是效率高的原因。

1. 问题3： 怎么理解条件假设等于是说用于分类的特征在类确定的条件下都是条件独立的，这一假设使朴素贝叶斯法变得简单，但有时也会牺牲一定的分类准确率。

自己的理解：朴素贝叶斯法关键有一个条件独立假设，但是真实情况下并不一定就是绝对的条件独立，输入变量中的每一维的属性可能是相互关联的，所以直接用条件独立是有所欠妥的，得到的分类结果肯定会有错误。

1. 问题4： 朴素贝叶斯能否解决属性不独立同分布的情况？

自己的理解：正如上一个问题所说，这个问题的解决并不容易，首先朴素贝叶斯这种属性独立同分布一定会带来误差，我认为如果属性安全是绝对相关的，那么用朴素贝叶斯应该还是可以分类的，只不过可能误差较大，应该视数据而定，误差在可考虑范围内仍然可以朴素贝叶斯法，因为简单而且有一定的效率。

1. 读书计划

1、本周完成的内容章节：《统计机器学习》第四章

2、下周计划：《统计机器学习》第五章

四、读书摘要及理解

1、朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。对于给定的训练数据集，首先基于特征条件独立假设学习输入/输出的联合概率分布，然后基于此模型，对给定的输入x，利用贝叶斯定理求出后验概率最大的输出y。朴素贝叶斯法实现简单，学习与预测的效率都很高，是一种常用的方法。

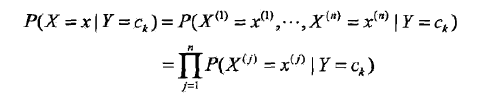
2、训练数据集由独立同分布产生。朴素贝叶斯法通过训练数据集学习联合概率分布。学习的是先验概率分布及条件概率分布。先验概率分布如下：



条件概率分布如下：

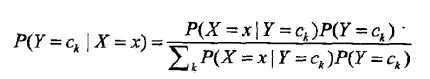


这样就可以学习到联合概率分布。条件概率分布有指数级数量的参数，估计实际上没有那么准确，朴素贝叶斯法对条件概率分布作了条件独立行的假设，假设为：



条件独立假设等于是说用于分类的特征在类确定的条件下都是条件独立的。这一假设使朴素贝叶斯法变得简单，但有时会牺牲一定的分类准确率。

分类时，对给定的输入x，通过学习到的模型计算后验概率分布，将后验概率最大的类作为x的类输出：

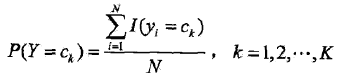


经过换算后可以得到分类结果就是：

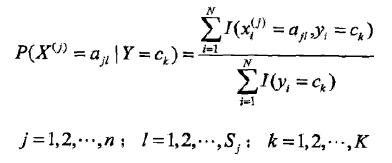


根据经验风险最小化可以得到该后验概率最大化准则。

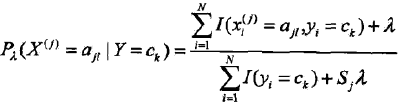
3、通过极大似然估计可以得到上述所说的训练过程中的概率分布，先验概率的极大似然估计是：



条件概率的极大似然估计是：



4、用极大似然估计可能会出现所要估计的概率值为0的情况，这时会影响到后验概率的计算结果，使分类产生偏差。解决这一问题的方法是采用贝叶斯估计。具体地，条件概率的贝叶斯估计是：



通常取，这时称为拉普拉斯平滑。同样，先验概率的贝叶斯估计是：

