读书报告

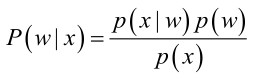
71117203-袁佳怡

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：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。

讨论后的理解：因为存在一些情况并不满足条件独立假设，条件之间无法做到完全独立，可能在某种情况下相互间还是会存在一定的联系。

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：习题4.1中请问如何用最大似然估计证明两个公式？最大似然估计有什么比较直观易懂的理解吗？

自己的理解：首先来理解贝叶斯分布



P(w)为先验概率，表示每种类别分布的概率；P(x|w)：类条件概率，表示在某种类别前提下，某事发生的概率；而P(w|x)为后验概率，表示某事发生了，并且它属于某一类别的概率，有了这个厚颜概率，我们可以对样本进行归类。后验概率越大，说明某事物属于这个类别的可能性越大，我们越有理由把它归到这个类别下。 但在现实问题中，我们需要利用既有的样本数据进行分类，一种可行的方法是需要对先验概率和类条件概率进行估计，然后再套用贝叶斯分类器。

其中，先验概率的估计比较简单，每个样本所属的自然状态是已知的（利用监督学习），或者是依靠经验或用训练样本中各类出现的频率估计。但是，类条件概率的估计（非常难），因为概率密度函数包含了一个随机变量的全部信息，样本数据可能不多，特征向量x的维度可能很大。所以这时候需要把完全未知的概率密度p(x|)转为估计参数。将概率密度估计转化为参数估计，极大似然估计就是一种参数估计方法。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。极大似然估计是建立在极大似然估计原理的基础上的一个统计方法，是概率论在统计学中的应用。利用试验结果得到某个参数值能够使样本出现的概率为最大，则称为极大似然估计。进行极大似然估计呢有一个未知参数和多个未知参数的情况，分别采用不同的方法去求解。方法的一般步骤为：写出似然函数，对似然函数取对数，并整理，求导数，解似然函数。

1. 问题3：朴素贝叶斯方法效率较高，体现在哪些方面呢？

自己的理解：因为使用朴素贝叶斯方法进行了条件独立的假设，这会使得参数个数会减少，所以朴素贝叶斯的效率会较高。

1. （必填）读书计划

1、本周完成的内容章节：《统计学习方法》第三章

2、《统计学习方法》第四章