*"""  
Created on 2017/8/13  
  
machine-learning-course  
  
@author: DSG  
"""*

1. **贝叶斯定理**

条件概率: 设与是样本空间中的两事件，若则称为在发生下的条件概率，同样有。结合这两个方程式，我们可以得到：



两边同除以就可以得到贝叶斯定理：



**二、贝叶斯公式**

**条件概率推论:**

性质：若中的互不相容，则



为事件域。

乘法公式：若，则， 若，则



全概率公式：设为样本空间上的一个分割，即互不相容，且，如果，则对任一事件有：



这里完整的写法，解释：在样本空间上发生的概率，就等于在所有分割上发生的概率总和。

贝叶斯公式：设为样本空间上的一个分割，即互不相容，且如果，则：



推导：对分子用乘法公式，对分母用全概率公式，同常叫做导致事件发生的概率。如果只对的分子用乘法公式，分母不动：



这个公式就是贝叶斯算法的核心公式。



**二、朴素贝叶斯分类器的公式**

假设某个体有项特征（Feature），分别为。现有个类别（Category），分别为。贝叶斯分类器就是计算某个数据在自己特征下概率最大的那个分类，也就是求下面这个算式的最大值：



对分母使用全概率公式：



假设如果项特征彼此独立则上公式又可以变为：



因为对于同一条数据来讲是就是整个特征向量咋数据集上的概率，对于这个复杂的概率可以用全概率公式和乘法公式转化求得。在求的过程中，对所有的都是相同的，所以问题可以转化为：



因为和有相同的单调性和增减性所以为了避免数据下越结，对上公式右边取自然对数那么问题就会变成比较下公式的大小：



其中表示特征向量的地维，表是类别。表示在训练集中的出现概率，表示在类别中特征出现的概率。

对于特征为布尔型的算法应用，首先要算出每个类别的概率，还要知道每个类别中所有特征出现的概率（其实就每个特征出现的次数占总特征数总和的比份，在文档分类器其中，就是每个词汇出现的次数占这个类别词汇总数的比份），然后找到测试数据的所有特征在每个类别中对应特征的概率，然后把他们想乘想加，在加上类别概率。

如果特征为数值型的数据，可以计算下式子的大小，不同的是所有的就是每个特征在总特征总数下的比分，下式子和有相同的单调性：值得注意的是期望和分布的峰值并不是在同一点取值。

