实验二

朴素贝叶斯分类器

实验过程：

生成向量：遍历数据集，将每个文档中的单词进行统计，生成词典。然后根据词典生成每个文档中向量。这一过程，实验一中已经实现。

训练数据：将20个文件夹划分为测试集和训练集，比例是8:2。首先计算先验概率，然后对每个单词在每一类下的概率进行计算。期间进行平滑处理。

测试数据：利用贝叶斯公式对每个测试集文档属于哪一类的概率进行计算，概率最大的作为这个文档的分类结果。

实验结果：0.816