SVM实验报告

191250111 裴为东

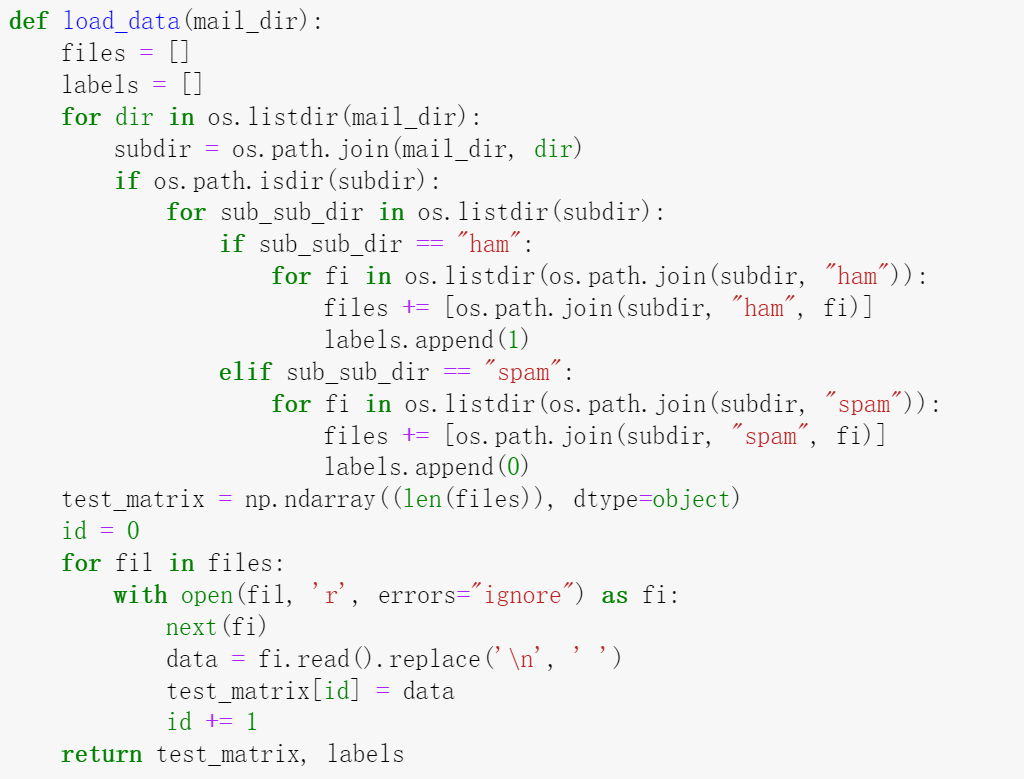
1.算法原理：

普遍意义上的SVM是一个二分类线性分类器，通过在分类超平面的正负两边各找到一个离分类超平面最近的点，使得这两个点距离分类器超平面的距离和最大

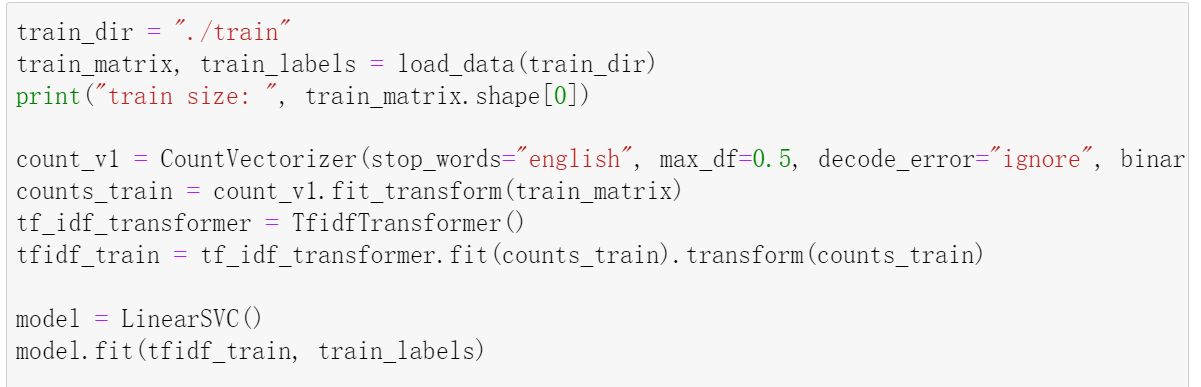
2.实验过程：

1）数据选用了已经过预处理的数据，故不需要再进行预处理

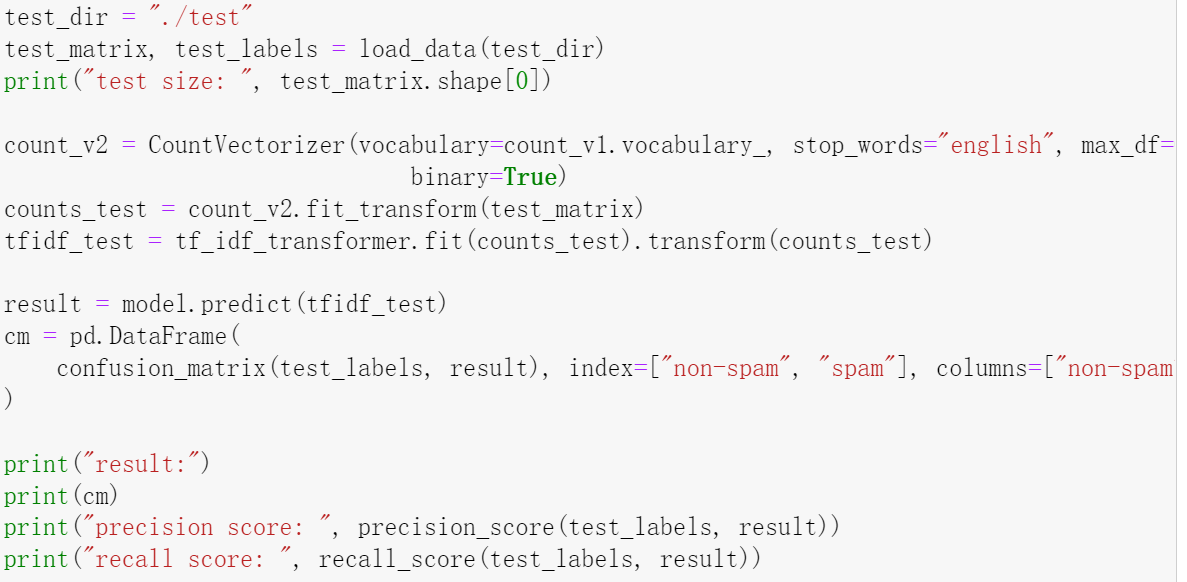
2）读取数据的内容，将数据集中enron5、enron6作为测试集，其余作为训练集



3）提取特征并训练模型，CountVectorize函数会将文本中的词语转换为词频矩阵，通过fit\_transform函数计算各个词语出现的次数，再经过TfidfTransformer函数输出词频逆反文档频率，最后通过sklearn库中的线性核SVM模型进行训练



4）评估模型



3.实验结果

