**程序报告**

1. **问题重述**

利用所学的监督学习的知识，完成对垃圾短息的分类

1. **设计思想**

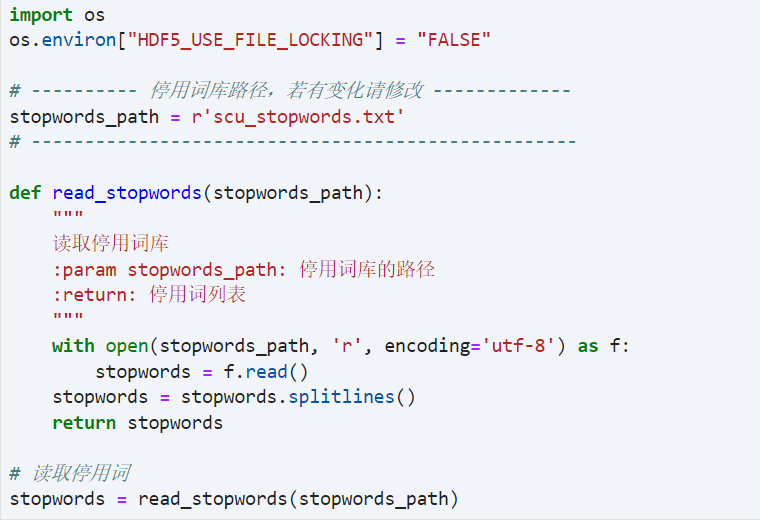
采取朴素贝叶斯的方法，将判断所用到的概率利用条件公式进行转化。同时应用停词表完成优化。为了进一步提高准确率，采取了MaxAbsScalar归一化的方法，同时也调节了ngram—\_range参数。

1. **代码内容**

调节pipeline中处理数据方法的参数：



读取停词表：



1. **总结**

通过调节参数，可以使已经十分强大的监督学习方法进一步提高准确率。