【背景】

在接触深度学习的第二周（第一周自学CNN在图像处理中的运用），孙霖老师布置的任务，在老师提供的86万条亚马逊评论数据中，通过CNN对自然语言进行处理，监督学习，将86万条评论数据划分为五类，通过课余时间自学完成。

【过程】

在刚开始的时候错误的选择了直接硬啃代码，通过类似的文本分类代码直接上手学习，结果研究了一周后，发现还是不会，第二周换了条路，选择先学习基础知识，了解整个过程，在听了莫烦的教程后，理解了很多，也开始自己编写代码。

在编写代码的过程中，主要遇到的问题有：

1. 在将词向量和源数据进行匹配的时候出现MemoryError的错误，先是找了list中元素形状不一致的原因，后来发现在进行大量数据匹配的时候仍然会出现该错误，我觉得是因为我把匹配的环节放到了整个过程前面将所有词向量和源数据进行匹配出现内存溢出，所以后来决定将匹配的过程放在训练的模型中，每次只匹配需要学习的数据。
2. 在转换为list和np.array中发生MemoryError的错误，在只进行八，九万条数据学习的时候没有报错，在进行十多万至七八十万的数据时就报错了，后来发现因为每条评论都贼长，最长的有六千多个单词，写个商品评论搞得跟写小作文一样，因为我是将所有评论长度统一至最长的那一条评论，所以6000\*860000的数据量是非常的庞大的，后来将评论长度进行优化，才成功让八十多万条的数据成功转化

【备注】

词向量用的是训练好的glove.6B.50d

评论数据只挑选了五个中一个小的