Lab 4 and Exercises Multi-Layer Perception

Step1 下载FTP中的数据集 *cnews*

把解压后的cnews文件夹放在项目TextClassification中即可，本次实验需要安装有一下python包文件：

```
tensorflow
numpy
collections #用来统计数据集中所有单词的频率
```

Step2 阅读理解数据处理的文件 *data_preprocess*

主要包括7个处理数据文件的函数：

```
# 读取文件数据
def read_file(filename)
# 根据训练集构建词汇表，存储
def build_vocab(train_dir, vocab_dir, vocab_size=5000)
# 读取词汇表
def read_vocab(vocab_dir)
# 读取分类目录(爬虫新闻固定的类别)
def read_category()
# 将id表示的内容转换为文字
def to_words(content, words)
# 读取新闻文件  将新闻的内容和标签都转为数字
def process_file(filename, word_to_id, cat_to_id, max_length=600)
# 生成批次数据
def batch_iter(x, y, batch_size=64)
```

Step3 阅读多层感知器相关参数的配置文件 *config*

主要包括词向量、单词字典、感知机中随机失活(dropout)、学习率和文件读取的配置

Step4 完成MLP作业 *mlp*

填写mlp.py文件中等式右边缺失的部分，实现多层感知机模型的训练，调用的相关函数说明参考tensorflow的相关API英文官方文档或者中文官方文档，也可以在pycharm中写出相关函数后按control键进入函数所在包文件的位置，仔细阅读其函数声明。
为了降低大家训练模型的难度，仅使用验证集作为训练数据。
但是在实际中应该使用训练集合验证集交叉训练，最后用测试集测试