首先从锐思数据库中下载数据：

使用锐思数据库中的模糊查询方法，下载与我的学号后2位相同的股票的数据。

接下来对这些数据进行合并，选择其中的八个股票作为研究对象。

在这里遇到了一个问题，我一开始创建了新的后缀名为xlsx的文件来存储各个股票数据，然后直接通过修改后缀名的方式来将其转化为csv文件，结果发现出现了很多意想不到的错误：read\_csv()函数的usecols没法读取对应的列，访问越界等等。

解决办法：最后我采用另存为的方法来将xlsx文件转化为csv文件，这样子就不会出现问题了。

还有就是编码格式的问题，我这里另存为之后使用的是utf-8编码的方式，需要将老师代码中的encoding=’GB2312’改为encoding=’utf-8’即可。

xls和xlsx两个文件另存为产生的csv文件中的时间类型，导入python中是不一样的，一个是‘\’分隔另一个是‘-’分隔，这就会导致在pd.merge时merge失败。（并不是这样的，实际上应该看excel中的单元格格式的设置）

蒋志强第一次实验的收益率求法？？？？

对于日期这种数据，注意要进行数据格式的变换，鬼知道它里面的日期格式长什么样。

就好比没有转换前的stockdata1和rf，一个是2000/01/04，2000/1/4。人傻了！！！

risk\_free['date'] = pd.to\_datetime(risk\_free['date'], format='%Y/%m/%d')

注意：merge\_data=pd.merge(left=merge\_data[['date','close','rfreturn','return']],right=index[['date','indexreturn']],on='date',how='inner')里面放的是一个列表。为什么是列表？因为他这里merge后得到的数据可以根据自己的需要设定列的先后（通过列表）

直接将risk\_free和stockdata进行内连接，但是risk\_free之前已经将date列转化为datetime64对象了，而stockdata的date列是string对象，出错。

解决方法：和之前一样

Dataframe和list的append方法是不一样的。List的append中不可以加参数，而dataframe的append中是可以加参数的

stock\_dataall.sort\_values(by=['code'],inplace=True)

按照所以其实之前不用那么辛苦自己手动处理数据，可以直接用这个办法

注意pandas的dataframe这个对象，假设stockdata为dataframe对象，stockdata[’date’]的类型并不是numpy.ndarray而是别的（名字很长），它包含了numpy.ndarray和其他很多东西。stockdata[’date’].values的类型是numpy.ndarray，只剩下纯粹的数据，索引也都没了

ind=(stock\_dataclass[count]['return']<=0.1)&(stock\_dataclass[count]['return']>=-0.1)这东西只能对dataframe或是numpy.ndarray使用，对于list是不适用的

注意在使用列表套dataframe的时候搞清楚哪个是list哪个是dataframe

python中的列表不能使用元组进行切片

y = ret\_stocks[:, 1]

这里在进行切片时会把[:, 1]解释为元组，这样就报了这个错。

y = ret\_stocks[:, 1]这个的意思是，对于ret\_stocks的所有元素，取序号为1的列。只有对于numpy.ndarray才适用，不能对于list使用