**量化投资问题**

量化投资是指通过数量化方式及计算机程序化发出买卖指令，以获取稳定收益为目的

的交易方式。投资者通过数据分析探索市场运行规律，并预测市场走势，从而进行决策交

易。随着大数据技术的发展，量化投资在全球金融交易市场上的地位愈加重要。但是由于

市场信息十分庞杂，同时产品的价格也受到其他诸多因素的影响，如何从海量的市场信息

中提取出有效指标，制订交易策略，是一个具有挑战性的工作。

本题附表提供的数据指标，主要包括：

①宏观市场指标（采购经理指数、社会消费品

零售总额、居民消费价格指数、国内生产总值、人民币存款利率、人民币贷款利率）；

②国内股票市场指标（上证综合指数成交量、上证综合指数成交金额、股票市场总值、股票

市场流通值（股票流通市值）、沪深 300 指数、上证综合指数、中证 500 指数、创业板指

数、上证 50 指数、上证 A 股指数、深证成份指数、深证综合指数科创 50 指数）；

③技术指标（VMA、VMACD、ARBR、OBV、BBI、DMA、MA、EXPMA、MTM、MACD、BIAS、KDJ、RSI、BOLL）；

④国际股票市场指标（道琼斯工业指数、纳斯达克综合指数、标准普尔 500 指数、美国证交所指数、美元/人民币汇率、香港恒生指数、东京日经 225 指数、伦敦金融时报100 指数、法国巴黎 CAC40 指数、荷兰 AEX 指数、俄罗斯 RTS 指数、意大利 MIB 指数、欧元/美元汇率）；

⑤“数字经济”板块信息（每 5 分钟开盘价、每 5 分钟收盘价、每 5 分钟最高价、每 5 分钟最低价、每 5 分钟成交量，每 5 分钟金额）；

⑥其他板块信息（“数字媒体”板块指数、“数字孪生”板块指数、“快手概念”板块指数、“互联网电商”板块指数、“互联网”板块指数）。

请参赛团队基于 2021 年 7 月 14 日至 2022 年 1 月 28 日每 5 分钟的“数字经济”板块给出的数据信息，完成以下任务：

（1）对所提供的各项指标进行分析，从中提取出与“数字经济”板块有关的主要指标。

[*https://zhuanlan.zhihu.com/p/507472728*](https://zhuanlan.zhihu.com/p/507472728)

*想法：这个题目看的不是特别懂，我的理解是对所有数据进行分析，然后找出和数字经济呈现出最大相关性的指标，然后后续的题目都是根据这个找出的指标来进行分析。*

*（一）*

[*https://www.zhihu.com/question/31992750/answer/2238103559*](https://www.zhihu.com/question/31992750/answer/2238103559) *数据相关性分析。*

*从这个回答里可以先绘制一些散点图，然后进行初步筛选，然后计算相关系数。*

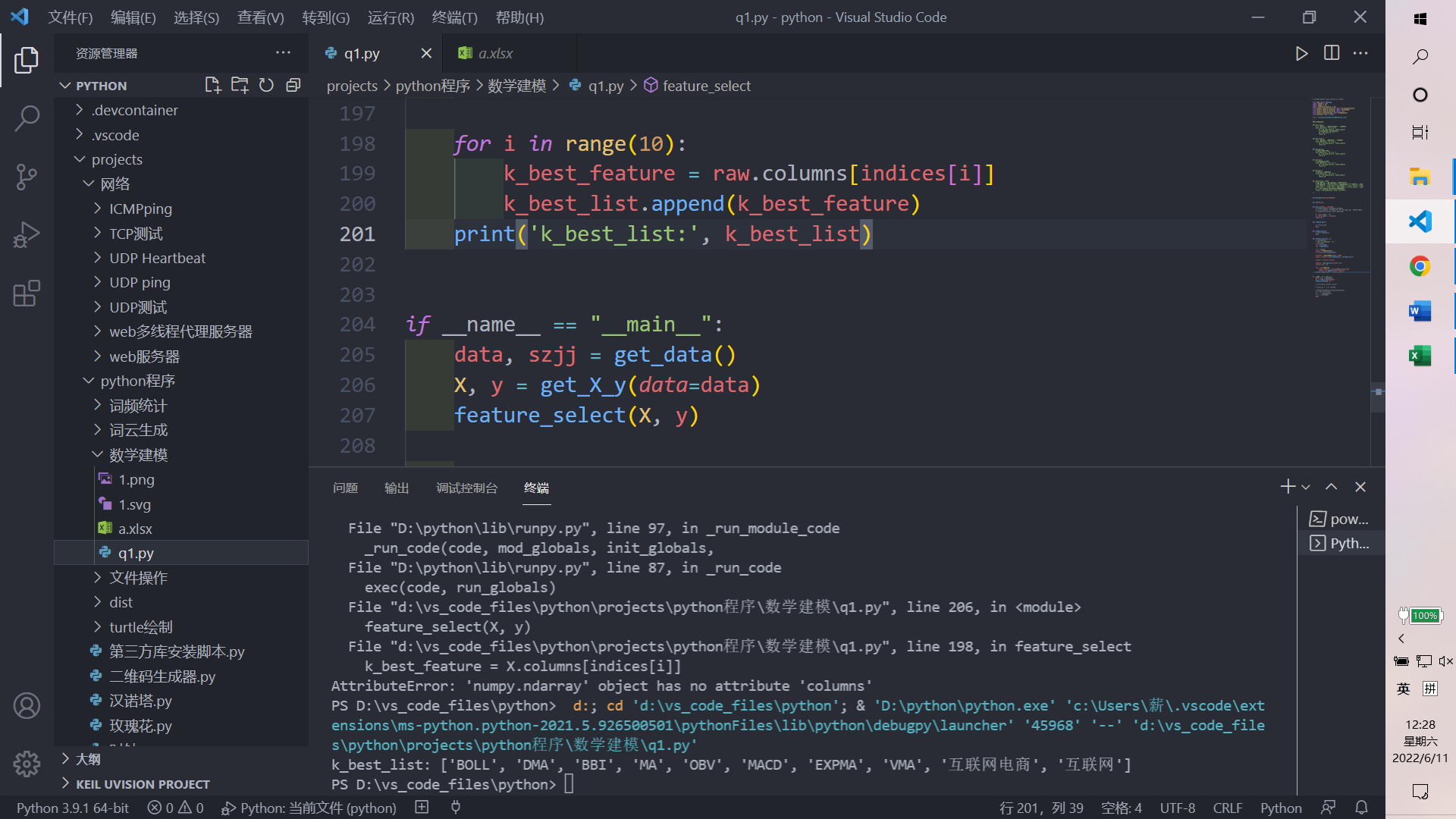
*（二）*

*其实这个还同时让我想到了机器学习里面的特征提取方法，可以通过计算信息熵和信息增益的方式来进行筛选。*

*①****（1）、卡方(Chi2)检验***

***原理参考：***[*https://mp.csdn.net/postedit/101861846*](https://mp.csdn.net/postedit/101861846)

*经典的卡方检验是****检验定性自变量对定性因变量的相关性****。比如，我们可以对样本进行一次chi2 测试来****选择最佳的10项特征：***



*②*

*（三）*

*同样是机器学习的内容，不过是采用降维的方法，这类方法有很多，sklearn中有很多现成的实现，可以直接使用。*

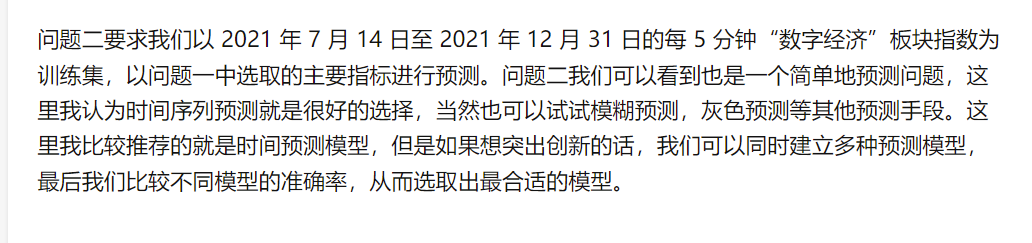
[*https://blog.csdn.net/s294878304/article/details/101851990*](https://blog.csdn.net/s294878304/article/details/101851990)

*我们可以三种方法或者还有更多的方法都使用一下，然后多面比较*

（2）以 2021 年 7 月 14 日至 2021 年 12 月 31 日的每 5 分钟“数字经济”板块指数为训练集，以 2022 年 1 月 4 日至 2022 年 1 月 28 日的每 5 分钟“数字经济”板块指数为测试集。根据问题（1）提取出来的各项指标对“数字经济”板块指数每 5 分钟成交量进行预测。

想法：这个就是回归问题了，首先要建立出数学模型

<https://zhuanlan.zhihu.com/p/507617817>



都是些没用过的模型。。。如果要使用上面的这些模型就只能从网上抄现成的了。

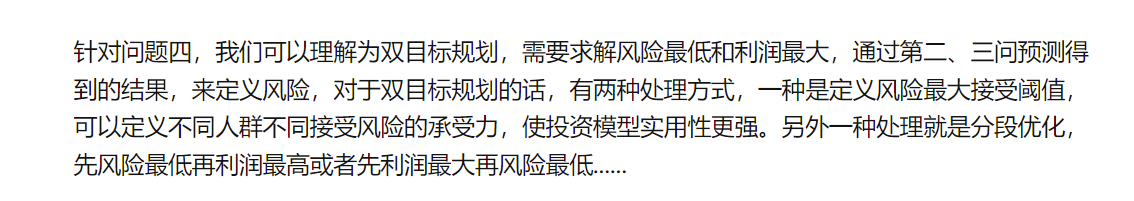
我想到的方法就是直接使用深度神经网络进行拟合，但是这点数据肯定会过拟合….想不到别的什么模型了。

（3）以 2021 年 7 月 14 日至 2021 年 12 月 31 日的每 5 分钟“数字经济”板块指数为训练集，以 2022 年 1 月 4 日至 2022 年 1 月 28 日的每 5 分钟“数字经济”板块指数为测试集。根据（1）和（2）建立模型对每 5 分钟的“数字经济”板块指数（收盘价）进行预测。

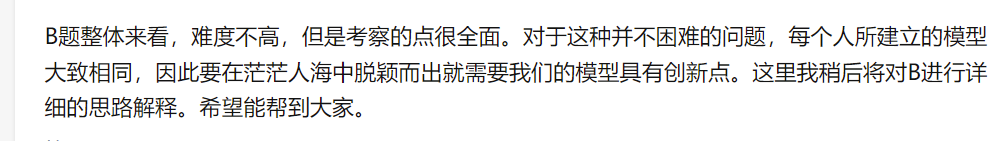
思路：这个就是求泛化误差了，可以根据这个泛化误差来验证前面的模型的效果。

（4）假设以“数字经济”板块指数为交易对象（在实际交易中指数无法交易，只能交易其中的个股），给定初始资金 100 万元，交易佣金为 0.3％，根据（3）得到的结果对“数字经济”板块每 5 分钟频率价格进行买卖交易，计算在 2022 年 1 月 4 日至 2022 年 1月 28 日期间交易的总收益率、信息比率、最大回撤率。

思路：这个是使用预测的模型进行预测，来判断结果。是一个优化问题？



如果按照上面的这个表述来看，可以参考黄皮书上的一道题。



说明：

1）总收益率=[1×（1＋第 1 日当日收益率）×（1＋第 2 日当日收益率）×（1＋第 3

日当日收益率）×……×（1＋第 N 日当日收益率）－1]×100%，

当日收益率=（当日收盘总资产市值－前一日收盘总资产市值）/前一日收盘总资产市值。

2）信息比率=日均超额收益率/超额收益率标准差，

日均超额收益率=第 i 天超额收益率/T，其中 T 表示按交易日计算的天数，

超额收益率=当日收益率－当日中证 500 指数收益率×90%，

超额收益率标准差：T（按交易日计算的天数）日内的每日超额收益率序列求得的标准差。

3）设 为第 i 天的产品净值， 是 后面第 j 天的净值，那么最大回撤率