

第三届“大湾区杯”粤港澳金融数学建模竞赛

2022 A题 高频交易的策略设计和异常交易识别

华南理工大学数学学院

刘深泉教授

目录

- 题目，高频交易的策略设计和异常交易识别
- 问题1，评分标准和模型方法
- 问题2，评分标准和模型方法
- 问题3，评分标准和模型方法
- 问题4，评分标准和模型方法
- 总结，高频交易策略和识别

高频交易的策略设计和异常交易识别

- 高频交易（HFT）是金融投资市场上最热门的话题之一，它是一种电子化交易，是指利用极为短暂的市场变化中寻求获利的计算机程序化交易。它采用复杂的算法来评估众多市场，并根据市场条件执行交易订单。
- 高频交易的主要特征有：
 - （1）由计算机自动完成的程序化交易；
 - （2）交易量大；
 - （3）持仓时间很短，日内交易次数很多；
 - （4）每笔收益率很低，但追求总体收益稳定。

高频交易的策略设计和异常交易识别

- 高频交易的交易标的往往跨越了股票、固定收益、衍生品以及外汇，有着换手频率高，市场中性等特点。进行高频交易前，需完成交易策略设计和搭建交易硬件环境两部分工作。其中，股票的高频交易的手段有：
 - (1) 自动做市商高频交易系统。自动做市商系统运用自有资金持续买卖股票，并且从买卖价差中获利；
 - (2) 量化因子高频交易系统。通过市场的特征因子，采用统计套利、因子分析和特征深度学习等量化策略，在极短时间内，执行大量交易指令，寻求获利机会。

高频交易的策略设计和异常交易识别

- 在金融投资市场上，典型的高频交易策略有：
- (1) 詹姆斯-西蒙斯网格交易策略。将资金分成若干份，在交易之前设定每份资金的买入、卖出价。确定初始基准价，价格区间和网格步长，每一份的买入价格，都是上一份资金的卖出价格，捕获网格之间的利润；
- (2) 高频交易拆单策略。拆单买入低于面前价格的股票，拆单卖出低于面前价格的股票；
- (3) 股票多因子深度学习策略。是通过选择股票的alpha 因子和风险因子，并利用历史数据进行深度学习而形成的交易策略。

高频交易的策略设计和异常交易识别

- 本题目主要是讨论股票高频交易的策略设计，请你运用数学建模方法，完成如下任务：
- 1、针对中国股票市场，请建立分类模型，将附录2中湾区指数的30只股票进行分类，选出最适合进行高频交易的10只股票。
- 2、针对不同股票的特点，例如大盘蓝筹股票或者波动显著的股票，请建立数学模型，设计高频交易策略。

高频交易的策略设计和异常交易识别

- 3、运用任务2设计的交易策略，对以下股票组合进行高频交易：
- 假若某粤港澳基金公司同时持有中国平安、美的集团，以及从上面任务1中选出来的2只湾区指数的股票，共4只股票，每只股票市值均为150万元（按10月31日收盘价算），同时账号持有流动资金400万元，合计1000万市值。请为该粤港澳基金公司设计高频交易策略，计算股票的收益率，给出收益率曲线。
- 注1：中国平安是国内金融牌照最齐全、业务范围最广泛的金融集团之一，中国平安（601318）是粤港澳大湾区优质蓝筹股票，受众多金融投资机构吹捧并重仓。
- 注2：美的集团是一家家用电器、暖通空调、机器人与自动化系统、智能供应链（物流）的科技集团，美的集团（000333）股票同样受到众多基金机构的追捧。。

高频交易的策略设计和异常交易识别

- 4、异常事件情况下，高频交易策略的改进：影响股票波动的因素有很多，其中突发事件和股东增减持是常见的异常事件。请针对以下事件，分别提出高频交易的改进策略，分析改进后的效果。
 - （1）光大证券乌龙指事件。
 - 2013年8月16日，光大证券(601788)股票在交易的时候，由于交易员的操作失误，带动了整个股指和其它股票的上涨，极短时间内，光大证券的操作失误给A市场的股票价格带来了巨大的影响（参见附录3）。
 - （2）大股东巴菲特减持比亚迪股票事件。
 - 比亚迪汽车产品紧跟国际潮流，其股票备受投资者青睐，从2008年开始，世界著名的投资家、股神巴菲特陆续购进比亚迪股份H股(01211)，其持股比例最高达20.49%，成为了比亚迪的大股东之一。而2022年8月30日，据港交所股权披露数据显示，股神巴菲特旗下伯克希尔公司在2022年8月24日减持了133.1万股比亚迪股份H股，其持股比例降至19.92%，受H股消息影响，比亚迪(002594)A股股价也随即持续下挫。。

高频交易的策略设计和异常交易识别

- 5、请你提供一个时间段高频交易策略数据和实测数据的比较的报表，说明你采用高频交易策略获取的超额投资利润。
- 注意，粤港澳基金公司高频交易的基本要求：
 - (1) 粤港澳基金公司账户总市值1000万，其中持仓股票市值为600万，持有现金400万。
 - (2) 账户日内交易，即每个交易日盘后持仓股票的数量保持不变。
 - (3) 股票交易佣金按交易金额双向0.15%收取单笔交易佣金，佣金不足5元按5元收取；印花税按卖出时交易金额的1.0%收取。
- 账户对单支股票全天高频交易数量不得超过该股票全天成交量的10%。

高频交易的策略设计和异常交易识别

- 附录1、股票市场的数据
 - (1) 中国平安 (601318) , 20221001-20221031期间股票交易数据。
 - (2) 美的集团 (000333) , 20221001-20221031期间股票交易数据。
 - (3) 光大证券 (601788) , 20130801-20130831期间股票交易数据。
 - (4) 比亚迪 (002594) , 20220801-20220931期间股票交易数据。
 - (5) 湾区股票历史数据, 20221001-20221031期间股票交易数据。
 - (6) 湾区股票实测数据, 20221101-20221131期间股票交易数据。

高频交易的策略设计和异常交易识别

- 附录2、湾区指数的30只股票。

公司名称	股票代码	公司名称	股票代码	公司名称	股票代码
分众传媒	002027	华侨城 A	000069	塔牌集团	002233
亿纬锂能	300014	金地集团	600383	粤水电	002060
立讯精密	002475	保利地产	600048	顺丰控股	002352
风华高科	000636	招商积余	001914	中顺洁柔	002511
国星光电	002449	中国平安	601318	美盈森	002303
生益科技	600183	瀚蓝环境	600323	珠江啤酒	002461
德赛电池	000049	广电运通	002152	中炬高新	600872
顺络电子	002138	海信家电	000921	白云山	600332
长盈精密	300115	华帝股份	002035	丽珠集团	000513
华发股份	600325	格力电器	000651	国药一致	000028

高频交易的策略设计和异常交易识别

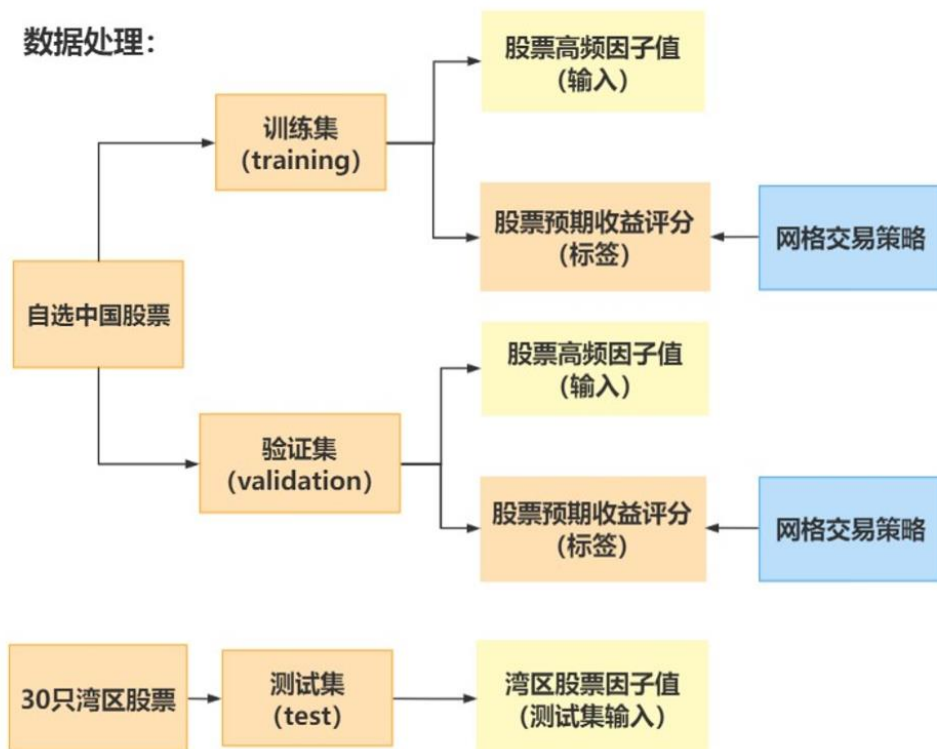
- 附录3、光大证券乌龙指事件
- 光大证券乌龙指事件，交易员操作失误是造成问题产生的重要原因。2013年8月16日，光大证券按计划开展ETF套利交易，开盘之前核定交易额度为8000万元；9点41分，策略交易员分析判断180ETF有套利空间，并发出第一组买入180ETF成分股177笔委托订单，委托金额不超过200万元；10点13分，交易员发出第二组买入180ETF成分股102笔委托订单，委托金额在150万元内。11点02分，交易员发出第三组买入180ETF成分股172笔委托，委托金额不超过200万元，但由于交易系统故障，2秒内系统生成26082笔预期外委托订单并发至交易所。由于操作错误，11点05分，上证综合指数突然上涨5.96%，中石油、中石化、工商银行和中国银行等权重股均触及涨停，大盘被直线拉升，沪指涨幅超过 5%，市场投资者猜测到系统出现问题。11点07分，光大证券交易员发现成交金额异常，并接到上海证券交易所电话询问，交易员迅速批量撤单，并终止套利政策订单生成系统的运行。11点30分收盘，光大证券股票成交金额72.7亿元，累计用于对冲而卖出的股指期货IF1309空头合约共253张。11点32分，21世纪网发布新闻“A股暴涨：光大证券自营盘70亿乌龙指”，将事件矛头对准了光大证券自营部门。11点59分，光大证券董事会秘书梅键否认该传闻，但该误导信息被各大门户网站转载。下午1点开盘后，策略投资部门开始对冲上午买入股票的风险。下午14时25分，光大证券发布公告，承认公司套利系统出现问题，内部正在进行核查和相应的处置工作。。

问题1（20分）：评分标准

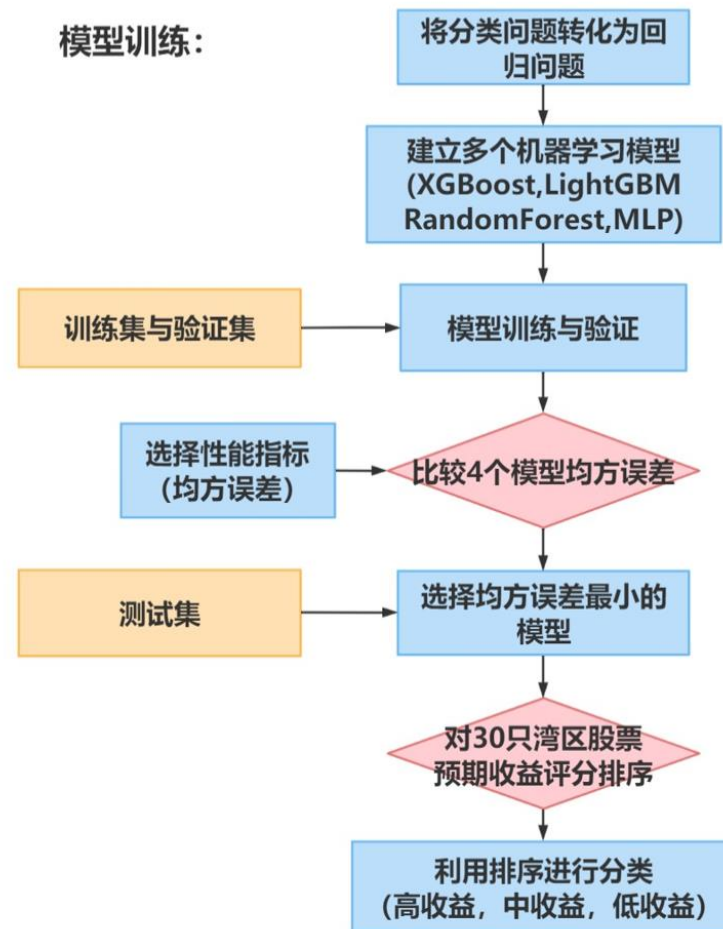
- （10分） 摘要和写作：评阅专家掌握
- 针对中国股票市场，请建立分类模型，将附录 2 中湾区指数的 30 只股票进行分类，选出最适合进行高频交易的 10 只股票。
- （1） 股票的分类方法。
- （2） 湾区股票的分类结果。

30只湾区股票分类，并择取最适合高频交易的10支股票

2.1 任务流程图



模型训练：



高频因子的选择

2.2 高频因子的构建

选择如下7种高频因子作为模型的输入，所选高频因子及其描述如下表：

因子名称（表达式）	因子描述与说明
收益率偏度 $real_skew_{D,i}$	度量收益率分布的不对称性，当偏度 <0 ，表示收益率分布左偏，当偏度 >0 ，表示收益率分布右偏
趋势占比 $trendratio_{D,i}$	衡量日内股票价格的趋势强度
日内收益率 $ret_intraday_{D,i}$	日内收益率仅考虑日内的股价变化
基于成交量的高频资金流因子 $flowInRatio$	资金流向来自交易时产生的委托信息，反映了微观层面的供求关系
下行波动占比	历史收益率低于平均收益率的占比
早/尾盘成交量占比 $ratio_volumeH1(H8)$	开盘后第1（8）个半小时成交量占全天成交量之比
高频量价相关性 ρ	每一交易日内1分钟频率的价格 p_t 和成交量 V_t 序列，计算Pearson相关系数



分类模型的选择

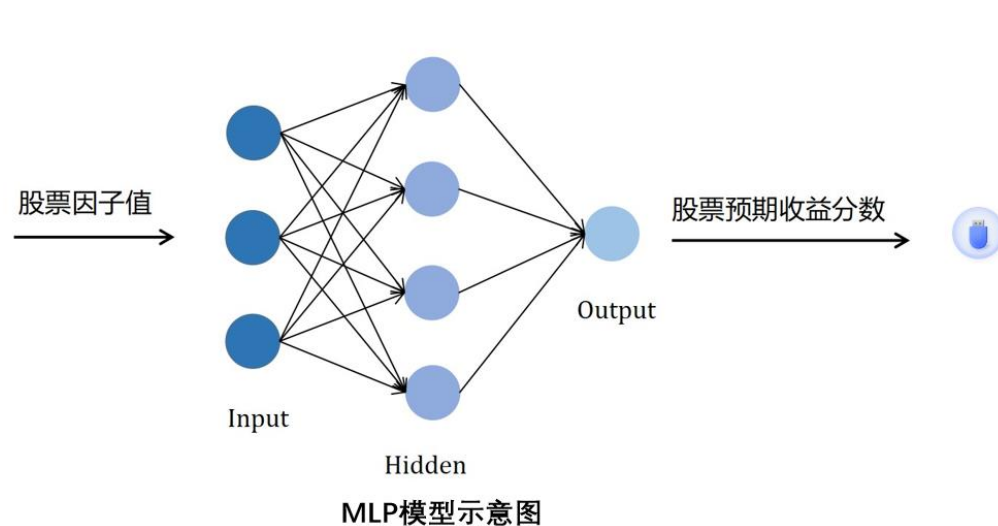
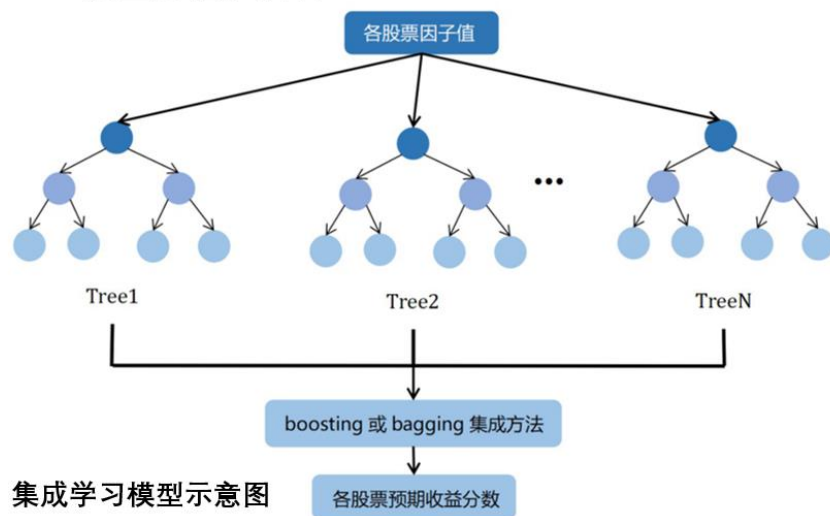
2.3 模型的选择与参数设定

本文选择了对决策树基学习器分别使用*boosting*以及*bagging*集成学习方法的模型：

XGBoost、Light GBM (*boosting*方法) 和RandomForest (*bagging*方法) 以及多层感知机 (MLP)

模型作为拟合高频因子与预测收益评分的候选模型，并使用**网格搜索方法**对其参数进行设定和调整。

模型示意图如下：



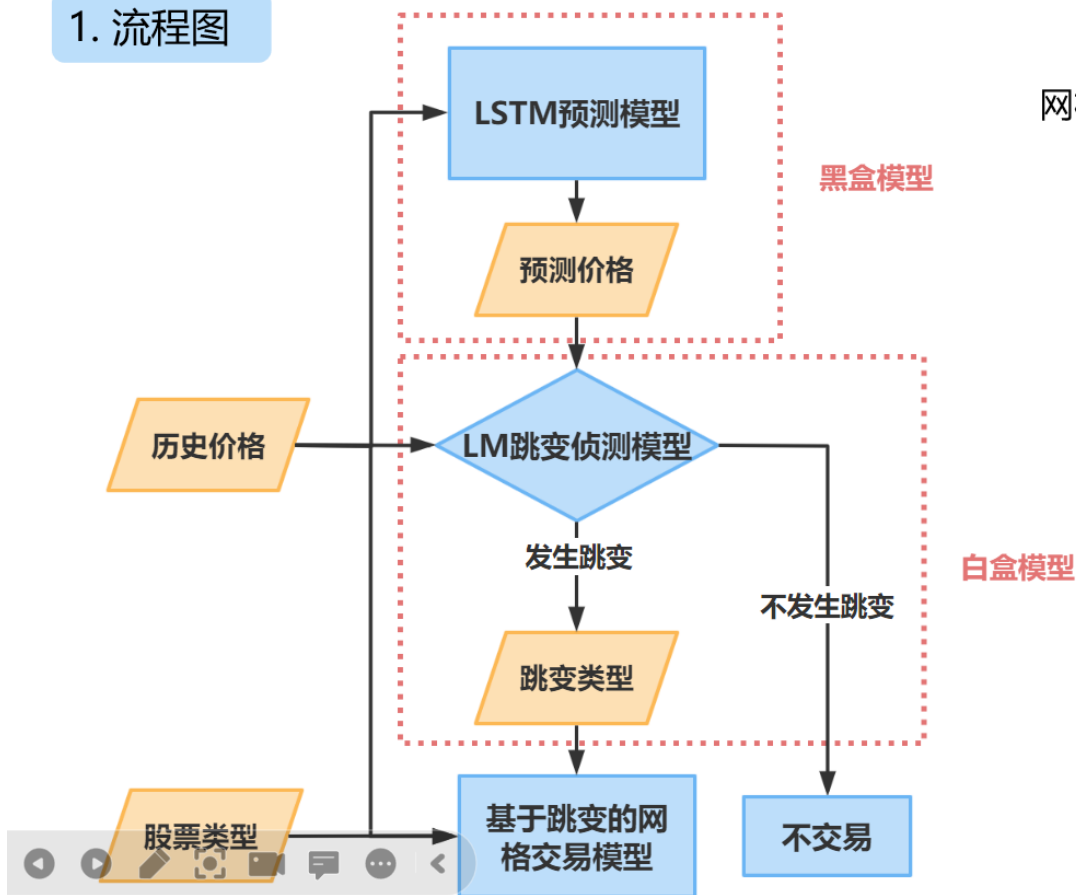
问题 2（25分）：评分标准

- 针对不同股票的特点，例如大盘蓝筹股票或者波动显著的股票，请建立数学模型，设计高频交易策略。
- （1）大盘蓝筹股票-给出高频交易策略。
- （2）波动显著股票-给出高频交易策略。
- （3）其它类型股票-高频交易策略
- （至少有两类股票的高频交易策略）。

问题2 思路

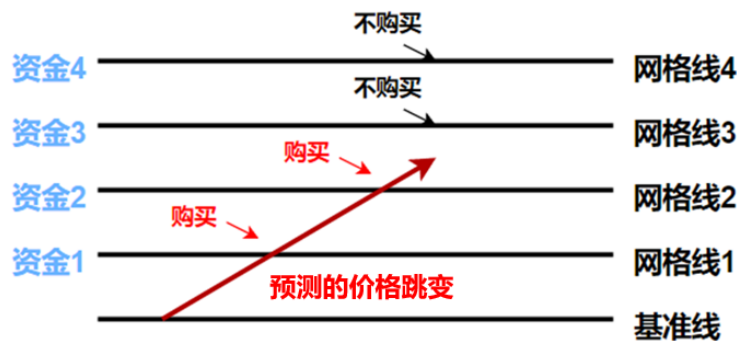
3.1 任务二思路阐述

1. 流程图



2. 基于跳变的网格交易模型

每次跳变的幅度有差异，受到传统网格交易思路的启发，建立网格对跳变的幅度进行区间判断，每个区间对应一定量的交易额。



★ 传统网格交易法：

利用真实价格，低买高卖，赚取中间差价

★ 基于跳变的网格交易模型：

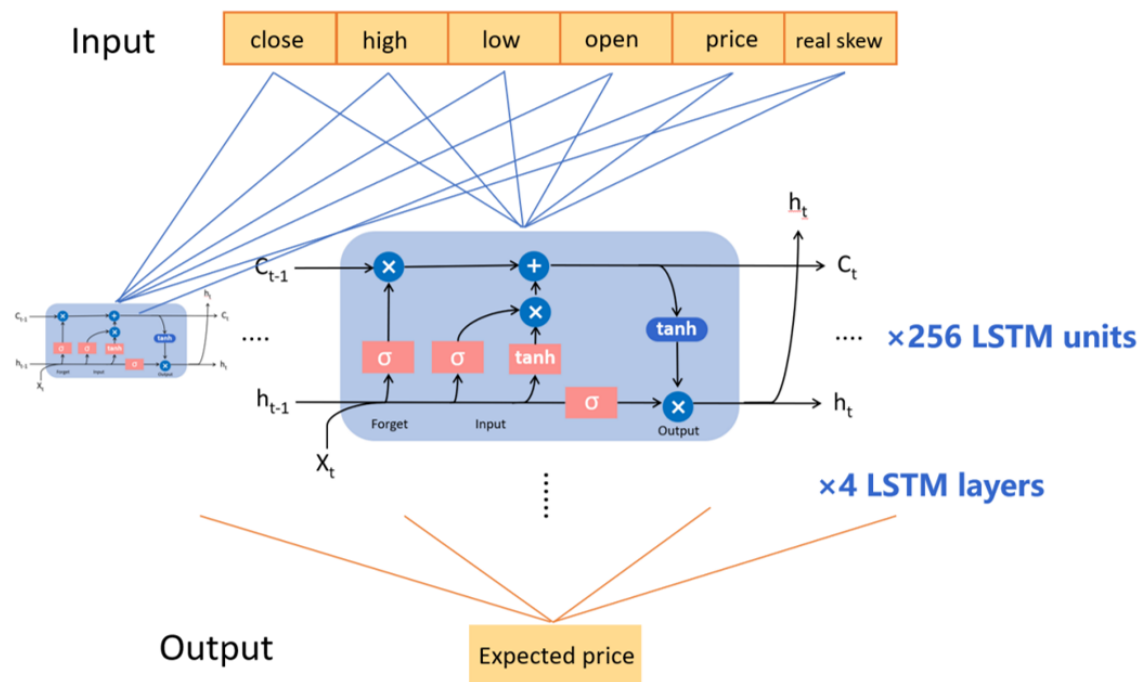
利用预测价格，**涨前买，跌前卖**，获得较高总资产

问题2 LSTM 价格预测方法

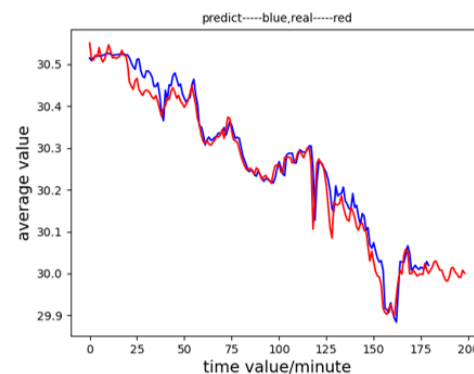
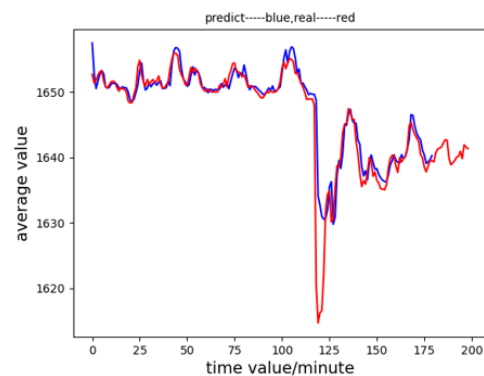
3.2 LSTM价格预测模型

股票价格是与时间、股票特点高度相关的序列数据，LSTM预测模型可以挖掘数据内在的高维特征。

1. 模型结构图



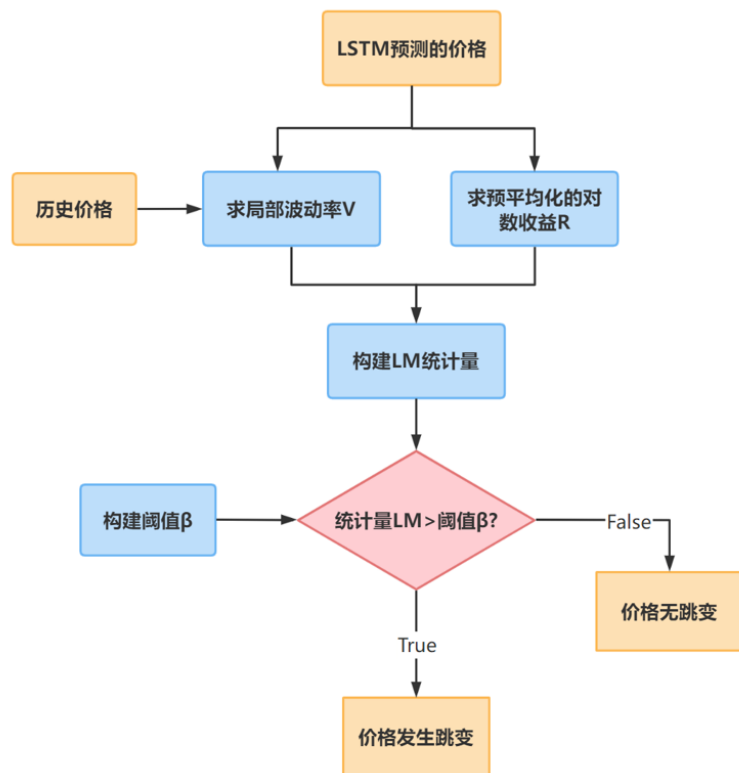
2. 预测效果展示



LM 跳变侦测模型

3.3 LM跳变侦测模型

1. 模型流程图



2. 关键参数公式

(1) 计算R, V

$$R_t = \frac{1}{K1} \sum_{i=1}^{K1} \log Y_{t-k+i} - \frac{1}{k} \sum_{i=1}^{K1} Y_{t-2k+i}$$

$$V_j = \frac{1}{K2-2} \sum_{j=i-K2+2}^{i=1} |R_{j-1}| |R_j|$$

(2) 构建LM

$$L_T = \frac{|R_T|}{\sqrt{V_j}} \sim N(0,1)$$

$$\frac{\max(L_j) - C_n}{S_n} \rightarrow \zeta$$
$$P(\zeta) = \exp(-e^{-\zeta})$$
$$Cn = \frac{(2 \times \log n)^{1/2}}{c} - \frac{\log \pi + \log(\log n)}{2 \times c(2 \log n)^{1/2}}$$
$$Sn = \frac{1}{c(2 \log n)^{1/2}}$$

(3) 比较LM, β

$$\frac{|L_i| - Cn}{Sn} > \beta$$

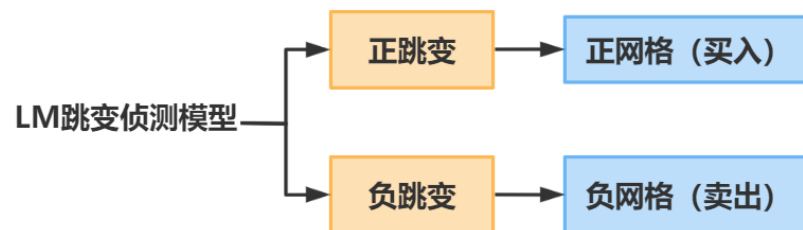
参数	Cn	Sn	β
取值	1.6825	0.6985	2.40



跳变侦测模型

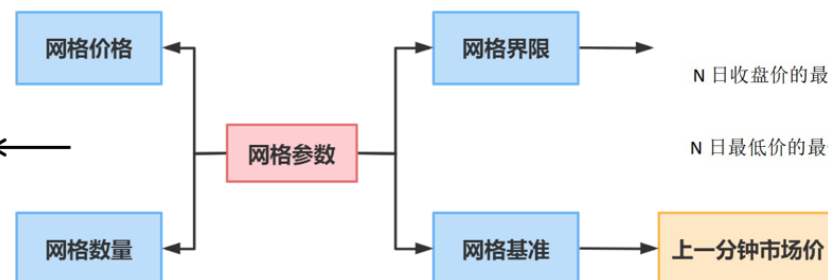
3.4 基于跳变的网格交易模型

1. 网格类型

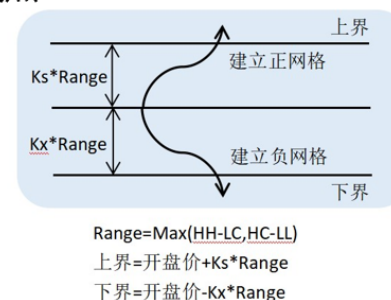
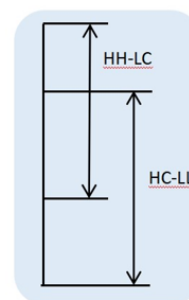


2. 网格参数

	网格数量	网格价格
蓝筹股	疏	高
稳定股	密	低
波动股	中	中



Dual Thrust模型计算上下触发点



实现针对不同股票特点的策略特异化

问题3（30分）：评分标准

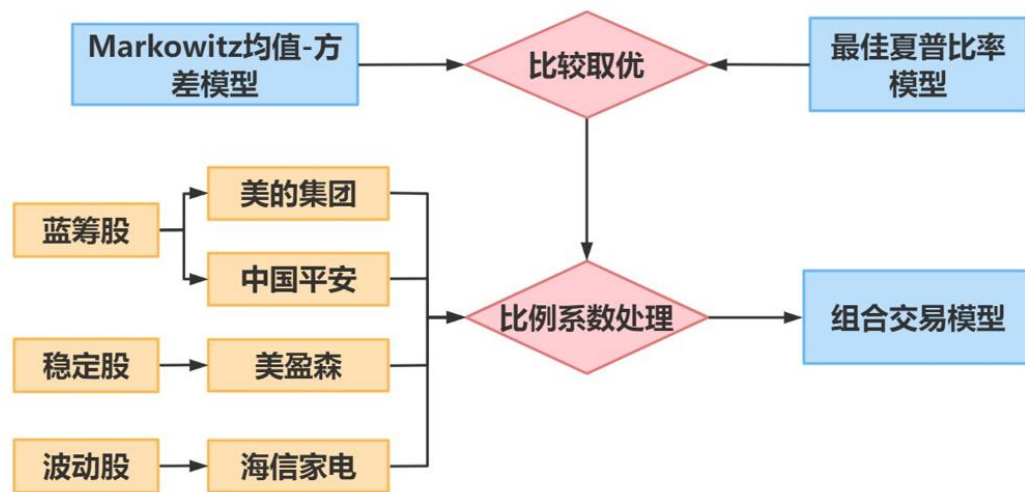
- 运用任务 2 设计的交易策略，对以下股票组合进行高频交易。
- 假若某粤港澳基金公司同时持有中国平安、美的集团，以及从上面任务 1 中选出来的 2
- 只湾区指数的股票，共 4 只股票，每只股票市值均为 150 万元（按 2022 年 10 月 31 日收盘价算），同时账号持有流动资金 400 万元，合计 1000 万市值。请为该粤港澳基金公司设计 2 高频交易策略，计算股票的收益率，给出收益率曲线。
- （1）中国平安：高频交易策略，计算股票的收益率，给出收益率曲线。
- （2）美的集团：高频交易策略，计算股票的收益率，给出收益率曲线。
- （3）2只湾区指数的股票：高频交易策略，计算股票的收益率，给出收益率曲线。

问题3 思路流程

4.1 任务三流程图

由于题目所给的两只股票都属于蓝筹股，故从以下方面考虑选股组合：

- 根据任务一选择高频高收益的湾区股票
- 选择不同行业的股票，规避行业风险
- 选择不同类型的股票，**稳定股可降低投资风险，波动股可利用震荡行情提高收益**



问题3 模型分析

4.2 资金分配占比的确定

构建投资组合的目的是在低风险的前提下实现高收益。

1.最佳夏普比率模型

🎯 模型目标：找出对应最大夏普比率的组合

$$\text{SharpeRatio} = \frac{E(R) - R_0}{\sigma}$$

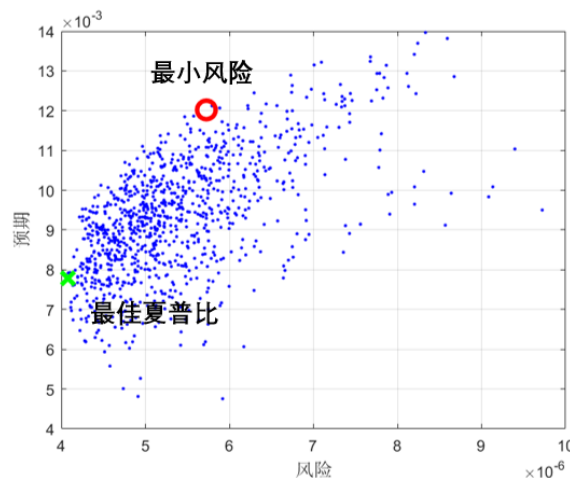
股票	中国平安	美的集团	海信家电	美盈森
分配比例	29.03%	10.21%	38.75%	22.02%
组合风险	4.08E-6		组合收益	2.22%

2.Markowitz均值-方差模型

🎯 模型目标：固定预期收益，找出最小风险组合

(1) 预期收益率E, 波动率V

$$E(R_p) = \sum_{i=1}^n x_i E[R_i] \quad V[R_p] = \sum_{i=1}^n \sum_{j=1}^i x_i x_j \text{cov}[R_i, R_j]$$



(2) 构建二次规划问题

$$\min V[R_p]$$

$$\begin{cases} s.t. E[R_p] \geq retE \\ \sum_{i=1}^n x_i = 1 \end{cases}$$

问题4 （15分） 评分标准

- 影响股票波动的因素有很多，其中突发事件和股东增减持是常见的异常事件。请针对以下突发异常事件，分别提出高频交易的改进策略，分析改进后的效果。
- （1）光大证券乌龙指事件：高频交易的改进策略，分析改进后的效果。
- （2）大股东巴菲特减持比亚迪股票事件：高频交易的改进策略，分析改进后的效果。
- （3）异常事件总结：高频交易的改进策略，分析改进后的效果。

问题4 光大证券乌龙指

5.1 任务四 (1)

1.光大证券乌龙指事件

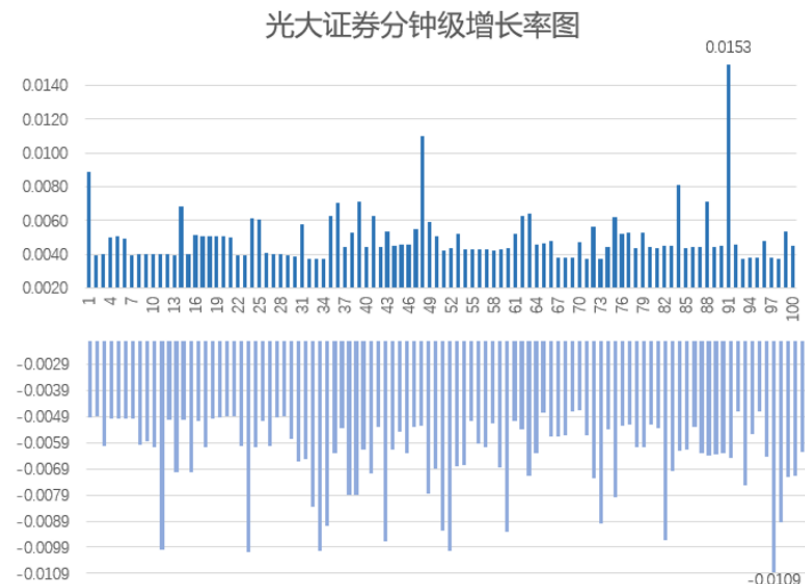
🧠 分析：由于操作不当导致股指在短时间内急剧上涨。

📊 改进措施：将阈值作为新约束条件添加到LM跳变侦测模型中，当LM统计量满足原发生跳变条件，且股票实时的**涨跌幅度处于阈值范围内**，才判定为发生跳变。

2. 阈值确定

📌 上涨幅度 $< 3\%$ ，下跌幅度 $> -3\%$

光大证券（2013年7月）分钟级增长率最大最小的100个数据



问题4 巴菲特减持比亚迪

5.2 任务四 (2)

1. 巴菲特减持比亚迪股票事件

🧠 问题分析：大股东减持股票，导致股价大幅下跌，且H股下跌早于A股。

📊 改进措施：（1）选取判定阈值，对比亚迪H股进行异常情况进行监听（2）当H股发生异常时，则比亚迪A股股票开启为期**7天的窗口期**。在此期间，LM模型**关闭正向跳变**，即发生正向跳变时不采取买卖操作。

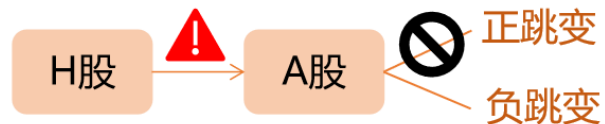
2. 阈值确定

✦ 交易量 > 10M股

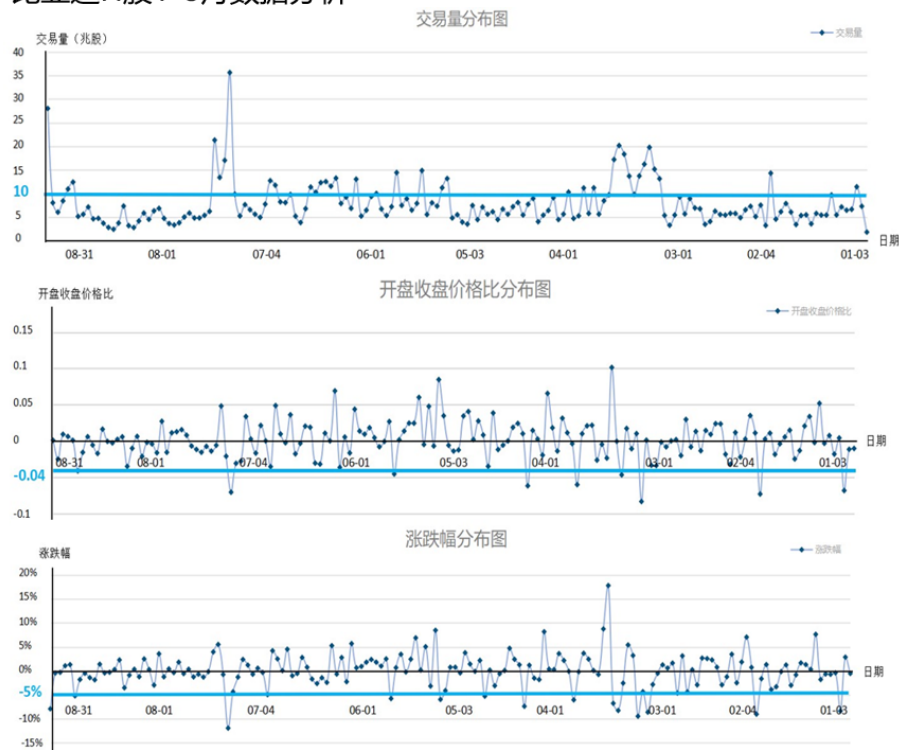
✦ 收盘开盘价格比 < -0.04

✦ 涨跌幅为 < -5%

} 同时满足即异常



比亚迪H股1-8月数据分析



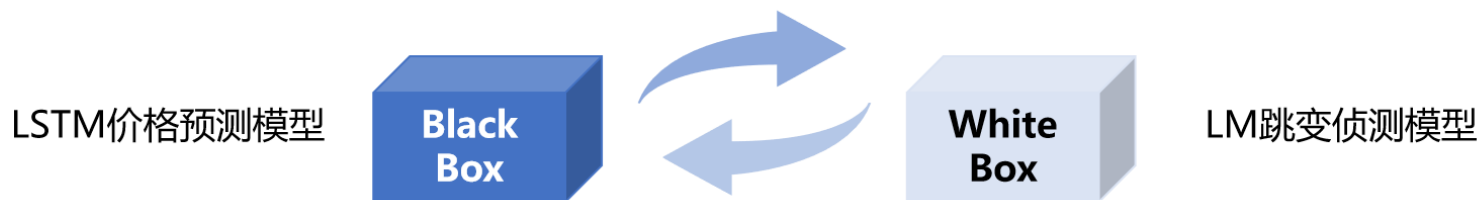
实测（单独计30分）评分标准

- 高频策略进行实测（分7档）
- （1）第一阶段：收益率曲线和回撤等重要评价指标。
- （2）第二阶段：文字实测报告。

总结，策略设计和异常识别

6.1 模型及策略的优点

- 1. 任务一使用了**多种算法**对股票模型进行拟合，并从中选优作为分类依据，分类结果具有较高参考价值
- 2. 所提出的交易模型是**“黑白结合”**的优异算法，同时具有黑盒模型LSTM**强大的学习能力**，及白盒模型LM**优异的扩展能力**，可依据不同股票、应用情境调整，避免了参数更新和重新学习，具有一定的应用前景。
- 3. 所提出的网格交易模型能充分利用价格信息，以LM模型输出的跳变作为输入，消除了噪声、异常值及波动率变化影响，具有**更强的择时能力**；依据股票价格在自身市场中的位置确定网格数，更具有**特异性**。



总结，高频交易

- 1、市场做市
- 2、统计套利
- 3、趋势追踪
- 4、事件驱动
- 5、高频对冲
- 第三届“大湾区杯”粤港澳金融数学建模竞赛
- 2022 A题 高频交易的策略设计和异常交易识别