

《Python 金融大数据》大作业题目

1. 运用相关库获取上市公司 A 的股票基本数据、计算相应的衍生变量数据、通过相关性分析选取合适的衍生变量、可视化呈现数据、并生成 Excel 工作簿。

具体要求如下：(30%)

- 1) 运用 Tushare 库获取上市公司 A 从 2020-02-03 至 2020-04-01 的股价涨跌幅和前 20 分钟成交量数据。
- 2) 根据以下两种方法计算从 2020-02-03 至 2020-04-01 的前 20 分钟成交量涨跌幅：
 - a) 采用当日交易量和昨日交易量的方法
 - b) 采用当日交易量和多日交易量的平均值的方法
- 3) 通过皮尔逊相关系数分别进行两种前 20 分钟成交量涨跌幅与股价涨跌幅的相关性分析，并选择出更优的前 20 分钟成交量涨跌幅。
- 4) 在进行简单数据优化后，绘制股价涨跌幅与前 20 分钟成交量涨跌幅的可视化图表（2 个 Y 轴的形式）。
- 5) 运用 xlwings 库将优化后的数据以及可视化图表导出至 Excel 工作簿。
(除了最终的大作业报告，还需提交相应源代码以及 Excel 文件)

2. 运用相关库从新浪财经网获取上市公司 A 的网页源代码，提取相应的标题、网址、来源和发布日期信息，进行简单的数据清洗，并生成文本文件。具体要求如下：(30%)

- 1) 运用 requests 库从新浪财经中获取上市公司 A 的网页源代码。
- 2) 编写正则表达式获取新闻标题、网址、来源和发布日期信息。
- 3) 对获取的新闻信息进行简单的数据清洗。
- 4) 尝试从新浪财经中获取上市公司 A 的前 5 页网页源代码，同时完成 2) 和 3) 中的内容（新闻信息的提取和简单数据清洗）。
- 5) 将 4) 中的数据信息保存到一个文本文件中。
(除了最终的大作业报告，还需提交相应的源代码以及 txt 文件)

3. 运用相关库从东方财富个股吧中获取 A 公司的网页源代码,提取相应的标题、网址、来源和发布日期等信息,进行简单的数据清洗,并生成文本文件。具体要求如下:(20%)

- 1) 运用 selenium 库从东方财富个股吧中获取上市公司 A 的网页源代码。
- 2) 编写正则表达式获取新闻标题、网址、来源和发布日期等信息。
- 3) 对获取的新闻信息进行简单的数据清洗。
- 4) 将数据信息保存到一个文本文件中。

(除了最终的大作业报告,还需提交相应的源代码以及 txt 文件)

4. 运用相关库获取上市公司 A 和上证指数的股票价格数据,分析其基本特征,进行正态性检验,并实施贝叶斯回归。具体要求如下:(20%)

- 1) 运用 Tushare 库获取 A 公司和上证指数(000001)从 2018-01-01 至 2019-12-31 的股票价格数据(收盘价)。
- 2) 将 A 公司和上证指数的股票价格数据规范化,并绘制股票规范化价格数的线图和散点图。
- 3) 计算 A 公司和上证指数的对数收益率,并进行正态性检验(3 种方法选其中 2 种)。
- 4) 运用 pymc3 库对 A 公司和上证指数的股票价格实施贝叶斯回归(A 公司的股票规范化价格为 x 值,上证指数的股票规范化价格为 y 值),输出统计结果,并绘制后验分布及轨迹图,和贝叶斯回归线。

(除了最终的大作业报告,还需提交相应的源代码)