

# 《Python 数据分析与机器学习》

## 作业及格式要求

### 1、总体要求：

- 总体原则：以本课程案例基金数据，进行数据分析分析
- 数据集

链接: <https://pan.baidu.com/s/1TDlDONubCf9pw7VnKblA2A>

提取码: nhrh

### 2、内容要求：

#### **A 类题：基础题（必做）：**

- [1] 编写函数，输出基金名字中包含有“混合”两字的基金名称及其代码（从总简表中）；
- [2] 编写函数，输出所有基金名称的云图（从总简表中）；
- [3] 编写函数，输出跌天数比率最大的 10 个基金的代码（从价格文件中）；
- [4] 编写函数，可以画出多个基金的价格变化曲线（从价格文件中）；
- [5] 编写函数，输出近 1 个月来跌幅最大的 10 个基金（从总简表中）；
- [6] 编写函数，可实现对于指定的一些基金 codes，输出近一个月来涨的天数比率最大的 10 个基金的代码；
- [7] 编写函数，输出近 1 周来和近 1 个月来跌幅都排名在前 20 的基金（从总简表中）；

#### **B 类题：中等题（5 选 2）：**

- [8] 编写函数 `rising_days_distribution(code)`，可以统计某一个基金 code 的连续涨跌天数，并以直方图的形式画出其连续涨跌天数分布直方图（从价格文件中）；
- [9] 编写函数 `gain_loss(code, from_date, end_date)`，计算如果从 `from_date` 天开始投入 10 万元购买代码为 code 的基金，到 `end_date` 这一天，可以赢利或亏损多少？（从价格文件中）；

- [10] 编写函数 `gain_loss_max(codes, money, from_date, end_date)`, 计算如果分别都投入资金数额为 `money` 的资金给代码为 `codes` 的多个基金 (`codes` 为基金代码序列), 投资时间范围是从 `from_date` 到 `end_date`, 计算投资哪支基金的收益将会最大, 输出该基金代码、名称, 及相应的收益额 (从价格文件中);
- [11] 编写函数, 输出所有基金中, 近一个月排名和近一周排名综合起来最好的基金 (可以用两个排名的名次加起来, 名次和越小, 就称排名最好) (从总简表中);
- [12] 编写函数 `search_sim_fund(target_fund, codes)`, 对于某个基金 `target_code`, 从多个候选基金 `codes` 中查找, 近一个月来涨跌与其最相似的基金 (从价格文件中);

### C 类题: 开放题 (5 选 1):

- [13] 抓取主题基金: 编写程序抓取各个主题或行业的基金代码。天天基金网有各种基金主题和行业, 如何获取某个主题和行业的基金代码? 也可不从天天基金上抓取。天天基金网上的基金主题网址:

[http://fund.eastmoney.com/ztjj/#!c/0/curr/zt-%E4%B8%BB%E9%A2%98%E7%B4%A2%E5%BC%95/fs/SON\\_1N/fst/DESC](http://fund.eastmoney.com/ztjj/#!c/0/curr/zt-%E4%B8%BB%E9%A2%98%E7%B4%A2%E5%BC%95/fs/SON_1N/fst/DESC)

- [14] 基金涨势: 如何认定一个基金在某一时刻是总体涨势还是跌势? 根据你自己的研究和思考, 编写程序, 判断一个基金在某一天到底是涨势还是跌势?
- [15] 基金推荐: 如果要你给我推荐 10 个基金, 你将推荐哪 10 个, 为什么? 编写程序给出推荐结果。
- [16] 基金聚类: 如果把所有基金聚类成  $N$  类, 编写程序可以实现基金聚类
- [17] 其他的利用本案例数据集进行的任何机器学习预测性分析或量化交易等方面分析

## 3、报告模板及格式要求

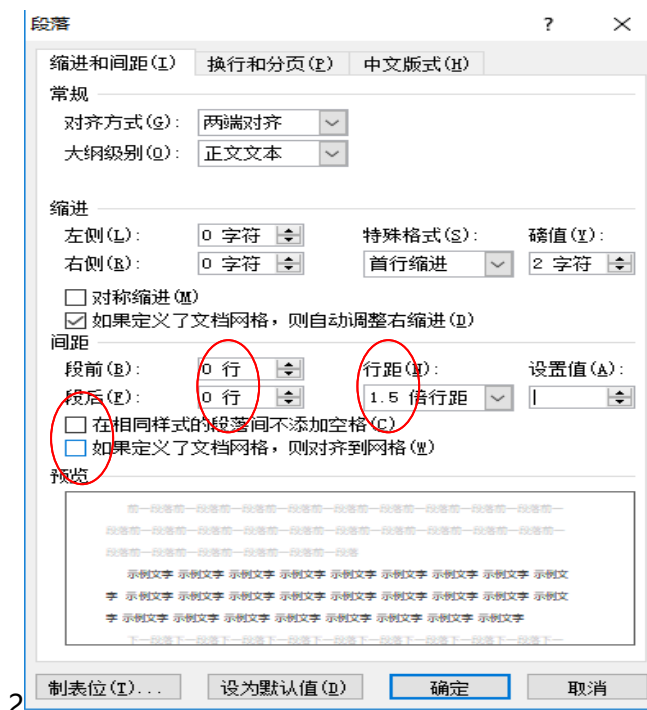
### (1) 需要提交的文档:

- **WORD 报告文档:** 把整个作业, 把上述题目把题目要求, 源程序及运行结果写到一个总 WORD 文档中, 作为课程总报告, **【运行结果截图到 WORD 中】**, 评分主要看这个文档。

- **程序文件**: .py 格式的源文件文档，按我的案例文档书写，文档中要有必须的注释；

## (2) WORD 报告文档模板：

- 整个文档要大纲级别层次分别，每个一级标题然后下面有二级标题、三级标题
- 上述所有报告文档书写要规范，层次清晰，在表达形式上一定要注意图文结合。整个报告大家统一按我以下格式：**全文正文字体微软雅黑（有些需要重点强调的可以用其他字体），1.5 倍行距；字号：一级标题用四号字，二级标题小四；正文 11 号字体。**
- 设置段落间距的时候一定要注意，如下图，段落间距的段前、段后都是 0 行，行距为 1.5 倍，然后最重要的是，下图段后下面有两个小框，一个是**“在相同样式的段落间不添加空格”**，一个是**“如果定义了文档网格，由对齐对网格”**这两个小框**都不要打勾!!!!**。



- 报告中如果是插入源代码的，在 WORD 中按如下样式插入代码：微软雅黑字号、小五、1.15 倍行距、灰背景

```
class RNN(nn.Module):
    def __init__(self, input_size, embedding_size, hidden_size, output_size):
        super(RNN, self).__init__()
        self.embedding_size = embedding_size
        self.input_size = input_size
        self.hidden_size = hidden_size
        self.output_size = output_size
        self.embedding = nn.Embedding(num_embeddings=input_size, embedding_dim)
        self.gru = nn.GRU(input_size=self.embedding_size, hidden_size=self.hidden_size)
        self.out = nn.Linear(self.hidden_size, self.output_size)
```

### (3) 文件归类及命名要求：

- 总体，交上来的作业包含三方面：文档、代码
- 总压缩文件命名：金融科技 PythonDA&ML\_学号\_姓名
- WORD 总文档文件命名：金融科技 PythonDA&ML\_学号\_姓名.docx

## 4、提交要求

### (1) 提交方式：

- 发送到邮箱：buaayuli@ruc.edu.cn
- 邮件标题：金融科技 PythonDA&ML\_学号\_姓名

### (2) 提交截止：

- 截止时间：2022 年 3 月 30 号

附 1：封面样式



《Python 数据分析与机器学习》

课程作业报告

姓 名：\_\_\_\_\_ 张 三

学 号：\_\_\_\_\_ 2020200581

学 院：\_\_\_\_\_ 商学院

专 业：\_\_\_\_\_ 工商管理类

授课教师：\_\_\_\_\_ 余 力

2022 年 1 月

## 附 2：WORD 报告文档样式

### 封面页

#### 目录

自动生成一个目录

#### 摘要【单独一页】

用两三百字左右摘要介绍一下本报告。

(下面的目录仅供参考，不是要求就做下面的题目，大家可以在此框架自行设计，但要保证层次清晰)

### 报告正文

#### 1. A 类题目

##### (1) 题目

分析大盘的历史变化

##### (2) 题目

分析大盘的历史变化

#### 2. B 类题目

主要展示你作了哪些数据分析，比如下面的，下面只是提示

##### (1) 题目 1

分析大盘的历史变化

##### (2) 题目 2

#### 3. C 类题目

##### (1) 题目

分析大盘的历史变化