

In [1]:

```
import pandas as pd
import numpy as np
```

# 预处理

注释掉的部分仅为了调试使用，生产环境不用考虑

In [2]:

```
#show excel
```

In [80]:

```
#hyperparameter , max weight of allocation , default value should be 0.15
HYP_MAXWEIGHT = 0.0401
```

In [4]:

```
pd.read_excel("dlqz.xlsx")
```

Out[4]:

|   | trade_date | 大盘股票     | 中小盘股票 | 美国股票 | 香港股票 | 利率债券     | 信用债券 | 货币 | 黄金  |
|---|------------|----------|-------|------|------|----------|------|----|-----|
| 0 | 2019-02-28 | 0.200000 | 0.5   | 0.05 | 0.05 | 0.090000 | 0.01 | 0  | 0.1 |
| 1 | 2019-05-31 | 0.199769 | 0.5   | 0.05 | 0.05 | 0.090231 | 0.01 | 0  | 0.1 |
| 2 | 2019-08-30 | 0.200000 | 0.5   | 0.05 | 0.00 | 0.100000 | 0.05 | 0  | 0.1 |

In [5]:

```
pd.read_excel("fund.xlsx")
```

Out[5]:

|   | fund_code | fund_sname    | fund_score | asset_type | trade_date |
|---|-----------|---------------|------------|------------|------------|
| 0 | 110011    | 易方达中小盘混合      | 100.000000 | 大盘股票       | 2019-02-28 |
| 1 | 162605    | 景顺长城鼎益混合(LOF) | 99.823322  | 大盘股票       | 2019-02-28 |
| 2 | 260108    | 景顺长城新兴成长混合    | 99.646643  | 大盘股票       | 2019-02-28 |
| 3 | 519736    | 交银新成长混合       | 99.293286  | 大盘股票       | 2019-02-28 |
| 4 | 40008     | 华安策略优选混合      | 98.586572  | 大盘股票       | 2019-02-28 |
| 5 | 519732    | 交银定期支付双息平衡混合  | 100.000000 | 中小盘股票      | 2019-02-28 |
| 6 | 1938      | 中欧时代先锋股票A     | 99.333333  | 中小盘股票      | 2019-02-28 |
| 7 | 519712    | 交银阿尔法核心混合     | 98.939929  | 中小盘股票      | 2019-02-28 |
| 8 | 519697    | 交银优势行业混合      | 98.333333  | 中小盘股票      | 2019-02-28 |
| 9 | 11111     | 易方达沪深300ETF联接 | 97.999999  | 大盘股票       | 2019-02-28 |

In [6]:

```
#load data  
df1=pd.read_excel("dlqz.xlsx")  
df2=pd.read_excel("fund.xlsx")
```

In [7]:

```
# for row in df1.iterrows():  
#     print(row[1].大盘股票)
```

In [8]:

```
# get the name of col to list for iter weights  
col=[column for column in df1]  
col_name=col[1:]
```

In [9]:

```
# col_name
```

In [70]:

In [11]:

```
# for index, row in df1.iterrows():  
#     print(row['trade_date'])
```

## 核心算法

### 均值分配

In [71]:

```

#用迭代器获取大类权重表的值
df_res = pd.DataFrame() #for save the result
for index, row in df1.iterrows():

    dateS=row['trade_date'] #时间戳

    for fundtype in col_name:

        total_weights=row[fundtype] #通过迭代器返回各类资产权值

#         print(fundtype)
#         print(df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund

#         dp=row[1]. 大盘股票 #大盘股票的权值
#         zxp=row[1]. 中小盘股票 #中小盘股票的权值
#         mg=row[1]. 美国股票 #美国股票的权值
#         xg=row[1]. 香港股票 #香港股票的权值
#         ll=row[1]. 利率债券 #利率债券的权值
#         xy=row[1]. 信用债券 #信用债券的权值
#         hb=row[1]. 货币 #货币权值
#         hj=row[1]. 黄金 #黄金权值

#         print(row[fundtype])

        """获取该时间戳中的大盘股票到dftmp, 降序排列"""
        dftmp=df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund
        mean_weight=total_weights/dftmp.shape[0] #计算均值

        if mean_weight > HYP_MAXWEIGHT:
            print(f"Error: 在 {dateS} 日期下的 {fundtype} 分配异常, 可能因阈值太低分配溢出导致")
            dftmp['weights'] = -1

        else:
            dftmp['weights'] = mean_weight
#         df2=pd.concat([df2, dftmp], axis=1)
        df_res=df_res.append(dftmp)
# print(df_res)

#         print('col sum of dftmp:', dftmp.shape[0]) #获取列数用以均分
#         print('mean_weight_of_dp:', mean_weight_of_dp)
#         print(df2[df2['trade_date']==dateS])
#         print(df2[(df2['asset_type']=='大盘股票') & (df2['trade_date']==dateS)].sort_values(by='fund_s
#         break

```

Error: 在 2019-02-28 日期下的 大盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-02-28 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-05-31 日期下的 大盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-05-31 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-08-30 日期下的 大盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-08-30 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致

In [72]:

df\_res

Out[72]:

|   | fund_code | fund_sname    | fund_score | asset_type | trade_date | weights |
|---|-----------|---------------|------------|------------|------------|---------|
| 0 | 110011    | 易方达中小盘混合      | 100.000000 | 大盘股票       | 2019-02-28 | -1.000  |
| 1 | 162605    | 景顺长城鼎益混合(LOF) | 99.823322  | 大盘股票       | 2019-02-28 | -1.000  |
| 2 | 260108    | 景顺长城新兴成长混合    | 99.646643  | 大盘股票       | 2019-02-28 | -1.000  |
| 3 | 519736    | 交银新成长混合       | 99.293286  | 大盘股票       | 2019-02-28 | -1.000  |
| 4 | 40008     | 华安策略优选混合      | 98.586572  | 大盘股票       | 2019-02-28 | -1.000  |
| 5 | 519732    | 交银定期支付双息平衡混合  | 100.000000 | 中小盘股票      | 2019-02-28 | -1.000  |
| 6 | 1938      | 中欧时代先锋股票A     | 99.333333  | 中小盘股票      | 2019-02-28 | -1.000  |

In [130]:

fundtype

Out[130]:

'大盘股票'

In [107]:

dftmp

Out[107]:

|    | fund_code | fund_sname         | fund_score | asset_type | trade_date |
|----|-----------|--------------------|------------|------------|------------|
| 35 | 64        | 大摩纯债稳定增利18个月定期开放债券 | 100.0      | 黄金         | 2019-02-28 |
| 36 | 519062    | 海富通阿尔法对冲混合A        | 100.0      | 黄金         | 2019-02-28 |
| 37 | 110011    | 易方达中小盘混合           | 100.0      | 黄金         | 2019-02-28 |
| 38 | 345       | 鹏华丰融定期开放债券         | 100.0      | 黄金         | 2019-02-28 |
| 39 | 511220    | 海富通上证城投债ETF        | 100.0      | 黄金         | 2019-02-28 |

测试报错模块的截图，此时分配的权重是0.04，阈值0.005，低于权重

```
#获取该时间戳中的大盘股票到dfdp, 降序排列
print(df2[(df2['asset_type']=='大盘股票') & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False))
dfdp=df2[(df2['asset_type']=='大盘股票') & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False)
print('col sum of dfdp:', dfdp.shape[0]) #获取列数用以均分
mean_weight_of_dp=dfdp.shape[0] #计算均值用以分配
print('mean_weight_of_dp:', mean_weight_of_dp)

if mean_weight_of_dp > HYP_MAXWEIGHT:
    print(f"Error: 在 {dateS} 日期下的 大盘股票 分配异常, 可能因权重超出阈值导致")
    dfdp['weights'] = -1

break
```

```
fund_code    fund_sname fund_score asset_type trade_date
0      110011      易方达中小盘混合  100.000000    大盘股票  2019-02-28
1      162605  景顺长城鼎益混合(LOF)  99.823322    大盘股票  2019-02-28
2      260108  景顺长城新兴成长混合  99.646643    大盘股票  2019-02-28
3      519736      交银新成长混合  99.293286    大盘股票  2019-02-28
4       40008      华安策略优选混合  98.586572    大盘股票  2019-02-28
col sum of dfdp: 5
mean_weight_of_dp: 0.04
Error: 在 2019-02-28 日期下的 大盘股票 分配异常, 可能因权重超出阈值导致
```

下表展示报错时分配的权重，可见权重此时为-1，表示分配异常

In [71]:

dfdp *#when the error is raised , the result of dfdp :*

Out[71]:

|   | fund_code | fund_sname    | fund_score | asset_type | trade_date | weights |
|---|-----------|---------------|------------|------------|------------|---------|
| 0 | 110011    | 易方达中小盘混合      | 100.000000 | 大盘股票       | 2019-02-28 | -1      |
| 1 | 162605    | 景顺长城鼎益混合(LOF) | 99.823322  | 大盘股票       | 2019-02-28 | -1      |
| 2 | 260108    | 景顺长城新兴成长混合    | 99.646643  | 大盘股票       | 2019-02-28 | -1      |
| 3 | 519736    | 交银新成长混合       | 99.293286  | 大盘股票       | 2019-02-28 | -1      |
| 4 | 40008     | 华安策略优选混合      | 98.586572  | 大盘股票       | 2019-02-28 | -1      |

获取列名，修改assert type为灵活的形式

```

#用迭代器获取大类权重表的值
for row in df1.iterrows():
    for fundtype in col_name:
        # print(fundtype)
        # print(df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False))
        dftmp=df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False)
        mean_weight_of_dp=dp/dftmp.shape[0] #计算均值用以分配
        print('col sum of dftmp:', dftmp.shape[0]) #获取列数用以均分
        print('mean_weight_of_dp:', mean_weight_of_dp)

        if mean_weight_of_dp > HYP_MAXWEIGHT:
            print(f'Error: 在 {dateS} 日期下的 {fundtype} 分配异常, 可能因权重超出阈值导致')
            dftmp['weights'] = -1
            print(dftmp)

# dateS=row[1].trade_date #时间戳
# dp=row[1].大盘股票 #大盘股票的权值
# xsp=row[1].中小盘股票 #中小盘股票的权值
# mg=row[1].美国股票 #美国股票的权值
# xg=row[1].香港股票 #香港股票的权值
# li=row[1].利率债券 #利率债券的权值
# xy=row[1].信用债券 #信用债券的权值
# hh=row[1].货币 #货币权值
# hj=row[1].黄金 #黄金权值

# print(df2[(df2['trade_date']==dateS)])

#获取该时间戳中的大盘股票到dfdp, 降序排列
# print(df2[(df2['asset_type']=='大盘股票') & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False))

break

```

```

col sum of dftmp: 5
mean_weight_of_dp: 0.04
Error: 在 2019-02-28 日期下的 大盘股票 分配异常, 可能因权重超出阈值导致
fund_code fund_sname fund_score asset_type trade_date weights
0 110011 易方达中小盘混合 100.000000 大盘股票 2019-02-28 -1
1 162605 景顺长城鼎益混合(LOF) 99.823322 大盘股票 2019-02-28 -1
2 260108 景顺长城新兴成长混合 99.646643 大盘股票 2019-02-28 -1
3 519736 交银新成长混合 99.293286 大盘股票 2019-02-28 -1
4 40008 华安策略优选混合 98.586572 大盘股票 2019-02-28 -1
col sum of dftmp: 5
mean_weight_of_dp: 0.04
Error: 在 2019-02-28 日期下的 中小盘股票 分配异常, 可能因权重超出阈值导致
fund_code fund_sname fund_score asset_type trade_date weights
5 519732 交银定期支付双息平衡混合 100.000000 中小盘股票 2019-02-28 -1
6 1938 中欧时代先锋股票A 99.333333 中小盘股票 2019-02-28 -1
7 519712 交银阿尔法核心混合 98.939929 中小盘股票 2019-02-28 -1
8 519697 交银优势行业混合 98.333333 中小盘股票 2019-02-28 -1
9 1410 信达澳银新能源产业股票 97.333333 中小盘股票 2019-02-28 -1
col sum of dftmp: 5

```

In [113]:

dftmp

Out[113]:

|    | fund_code | fund_sname         | fund_score | asset_type | trade_date | weights |
|----|-----------|--------------------|------------|------------|------------|---------|
| 35 | 64        | 大摩纯债稳定增利18个月定期开放债券 | 100.0      | 黄金         | 2019-02-28 | -1      |
| 36 | 519062    | 海富通阿尔法对冲混合A        | 100.0      | 黄金         | 2019-02-28 | -1      |
| 37 | 110011    | 易方达中小盘混合           | 100.0      | 黄金         | 2019-02-28 | -1      |
| 38 | 345       | 鹏华丰融定期开放债券         | 100.0      | 黄金         | 2019-02-28 | -1      |
| 39 | 511220    | 海富通上证城投债ETF        | 100.0      | 黄金         | 2019-02-28 | -1      |

## 按权分配

In [93]:

```

#用迭代器获取大类权重表的值
df_res = pd.DataFrame() #for save the result
for index, row in df1.iterrows():

    dateS=row['trade_date'] #时间戳

    for fundtype in col_name:

        total_weights=row[fundtype] #通过迭代器返回各类资产权值

        #获取该时间戳中的大盘股票到dftmp, 降序排列
        dftmp=df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund_
#         print(dftmp)
        score=dftmp['fund_score'].tolist() #score从df向list转换用于MinMax
        res_score=[]
        sum_score=sum(score)

        if total_weights > HYP_MAXWEIGHT*dftmp.shape[0]:
            print(f"Error: 在 {dateS} 日期下的 {fundtype} 分配异常, 可能因阈值太低分配溢出导致")
            dftmp['weights'] = -1
        else:
            """进行(x/sum(score))*total_weights的权重分配, 然后将每次迭代中超过阈值的overValue向后转
            overValue=0
            for x in score:
                y=(x/sum(score))*total_weights+overValue
                if y > HYP_MAXWEIGHT:
                    overValue=y-HYP_MAXWEIGHT
                    y=HYP_MAXWEIGHT
                else:
                    overValue=0
                res_score.append(y)
            dftmp['weights'] = res_score
        df_res=df_res.append(dftmp)

#         df2=pd.concat([df2, dftmp], axis=1)
#         df_res=df_res.append(dftmp)
# print(df_res)

```

Error: 在 2019-02-28 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致

Error: 在 2019-05-31 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致

Error: 在 2019-08-30 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致

### 按权求和始终低于阈值时报警

|   | fund_code | fund_sname    | fund_score | asset_type | trade_date |
|---|-----------|---------------|------------|------------|------------|
| 0 | 110011    | 易方达中小盘混合      | 100.000000 | 大盘股票       | 2019-02-28 |
| 1 | 162605    | 景顺长城鼎益混合(LOF) | 99.823322  | 大盘股票       | 2019-02-28 |
| 2 | 260108    | 景顺长城新兴成长混合    | 99.646643  | 大盘股票       | 2019-02-28 |
| 3 | 519736    | 交银新成长混合       | 99.293286  | 大盘股票       | 2019-02-28 |
| 4 | 40008     | 华安策略优选混合      | 98.586572  | 大盘股票       | 2019-02-28 |

Error: 在 2019-02-28 日期下的 大盘股票 分配异常, 可能因权重超出阈值导致

阈值设置0.0401的时候, 可见算法该部分起效, 下面的list中有多个单元限制在了0.0401, 后续的按数值one-by-one来补充

```

fund_code    fund_sname    fund_score    asset_type    trade_date
0      110011      易方达中小盘混合    100.000000      大盘股票    2019-02-28
1      162605      景顺长城鼎益混合(LOF)    99.823322      大盘股票    2019-02-28
2      260108      景顺长城新兴成长混合    99.646643      大盘股票    2019-02-28
3      519736      交银新成长混合    99.293286      大盘股票    2019-02-28
4      40008      华安策略优选混合    98.586572      大盘股票    2019-02-28
[0.0401, 0.0401, 0.0401, 0.04005523993742331, 0.03964476006257672]

```

In [94]:

df\_res

Out[94]:

|   | fund_code | fund_sname    | fund_score | asset_type | trade_date | weights   |
|---|-----------|---------------|------------|------------|------------|-----------|
| 0 | 110011    | 易方达中小盘混合      | 100.000000 | 大盘股票       | 2019-02-28 | 0.040100  |
| 1 | 162605    | 景顺长城鼎益混合(LOF) | 99.823322  | 大盘股票       | 2019-02-28 | 0.040100  |
| 2 | 260108    | 景顺长城新兴成长混合    | 99.646643  | 大盘股票       | 2019-02-28 | 0.040100  |
| 3 | 519736    | 交银新成长混合       | 99.293286  | 大盘股票       | 2019-02-28 | 0.040055  |
| 4 | 40008     | 华安策略优选混合      | 98.586572  | 大盘股票       | 2019-02-28 | 0.039645  |
| 5 | 519732    | 交银定期支付双息平衡混合  | 100.000000 | 中小盘股票      | 2019-02-28 | -1.000000 |
| 6 | 1938      | 中欧时代先锋股票A     | 99.333333  | 中小盘股票      | 2019-02-28 | -1.000000 |

## baseline

In [95]:

```

def allocateByMeanWeight(df1, df2, HYP_MAXWEIGHT = 0.15):
    #hyperparameter , max weight of allocation , default value should be 0.15
    col=[column for column in df1]
    col_name=col[1:]
    #用迭代器获取大类权重表的值
    df_res = pd.DataFrame() #for save the result
    for index, row in df1.iterrows():
        dateS=row['trade_date'] #时间戳
        for fundtype in col_name:
            total_weights=row[fundtype] #通过迭代器返回各类资产权值
            """获取该时间戳中的大盘股票到dftmp, 降序排列"""
            dftmp=df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund_score', ascending=False)
            mean_weight=total_weights/dftmp.shape[0] #计算均值
            if mean_weight > HYP_MAXWEIGHT:
                print(f"Error: 在 {dateS} 日期下的 {fundtype} 分配异常, 可能因阈值太低分配溢出导致")
                dftmp['weights'] = -1
            else:
                dftmp['weights'] = mean_weight
            df_res=df_res.append(dftmp)
    return df_res

```



In [96]:

```
allocateByMeanWeight(df1, df2) #测试baseline均值分配
```

Error: 在 2019-02-28 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-05-31 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致  
 Error: 在 2019-08-30 日期下的 中小盘股票 分配异常, 可能因阈值太低分配溢出导致

In [97]:

```
def allocateByScalerWeight(df1, df2, HYP_MAXWEIGHT = 0.15):
    #hyperparameter , max weight of allocation , default value should be 0.15
    col=[column for column in df1]
    col_name=col[1:]

    #用迭代器获取大类权重表的值
    df_res = pd.DataFrame() #for save the result
    for index, row in df1.iterrows():
        dateS=row['trade_date'] #时间戳
        for fundtype in col_name:
            total_weights=row[fundtype] #通过迭代器返回各类资产权值
            #获取该时间戳中的大盘股票到dftmp, 降序排列
            dftmp=df2[(df2['asset_type']==fundtype) & (df2['trade_date']==dateS)].sort_values(by='fund_score')
            score=dftmp['fund_score'].tolist() #score从df向list转换用于MinMax
            res_score=[]
            sum_score=sum(score)
            if total_weights > HYP_MAXWEIGHT*dftmp.shape[0]:
                print(f"Error: 在 {dateS} 日期下的 {fundtype} 分配异常, 可能因阈值太低分配溢出导致")
                dftmp['weights'] = -1
            else:
                """进行(x/sum(score))*total_weights的权重分配, 然后将每次迭代中超过阈值的overValue向
                overValue=0
                for x in score:
                    y=(x/sum(score))*total_weights+overValue
                    if y > HYP_MAXWEIGHT:
                        overValue=y-HYP_MAXWEIGHT
                        y=HYP_MAXWEIGHT
                    else:
                        overValue=0
                        res_score.append(y)
                dftmp['weights'] = res_score
            df_res=df_res.append(dftmp)
    return df_res
```

In [98]:

```
allocateByScalerWeight(df1, df2) #测试baseline按权分配
```

|     |        |                     |            |      |            |          |
|-----|--------|---------------------|------------|------|------------|----------|
| 94  | 270023 | 广发全球精选股票(QDII)      | 100.000000 | 美国股票 | 2019-08-30 | 0.010000 |
| 95  | 1149   | 汇丰晋信恒生A股行业龙头C       | 50.000000  | 香港股票 | 2019-08-30 | 0.000000 |
| 96  | 540012 | 汇丰晋信恒生龙头指数A         | 50.000000  | 香港股票 | 2019-08-30 | 0.000000 |
| 97  | 513600 | 南方恒生ETF(QDII)       | 2.102102   | 香港股票 | 2019-08-30 | 0.000000 |
| 98  | 948    | 华夏沪港通恒生ETF联接(QDII)A | 1.801802   | 香港股票 | 2019-08-30 | 0.000000 |
| 99  | 513660 | 华夏沪港通恒生ETF(QDII)    | 1.501502   | 香港股票 | 2019-08-30 | 0.000000 |
| 100 | 2600   | 易方达裕景添利6个月定开债券      | 98.169717  | 利率债券 | 2019-08-30 | 0.021132 |
| 101 | 2486   | 上银慧添利债券             | 93.178037  | 利率债券 | 2019-08-30 | 0.020057 |

In [ ]: