分时走势分类

# 目录

## 1. 单日K线图的分类

## 2.1利用马氏链分析分时图中的状态转移情况。

## 2.2 齐次马氏链的探索

## 3.基于马氏性对于日线级别的转移性质的分析

## 单日K线图的分类

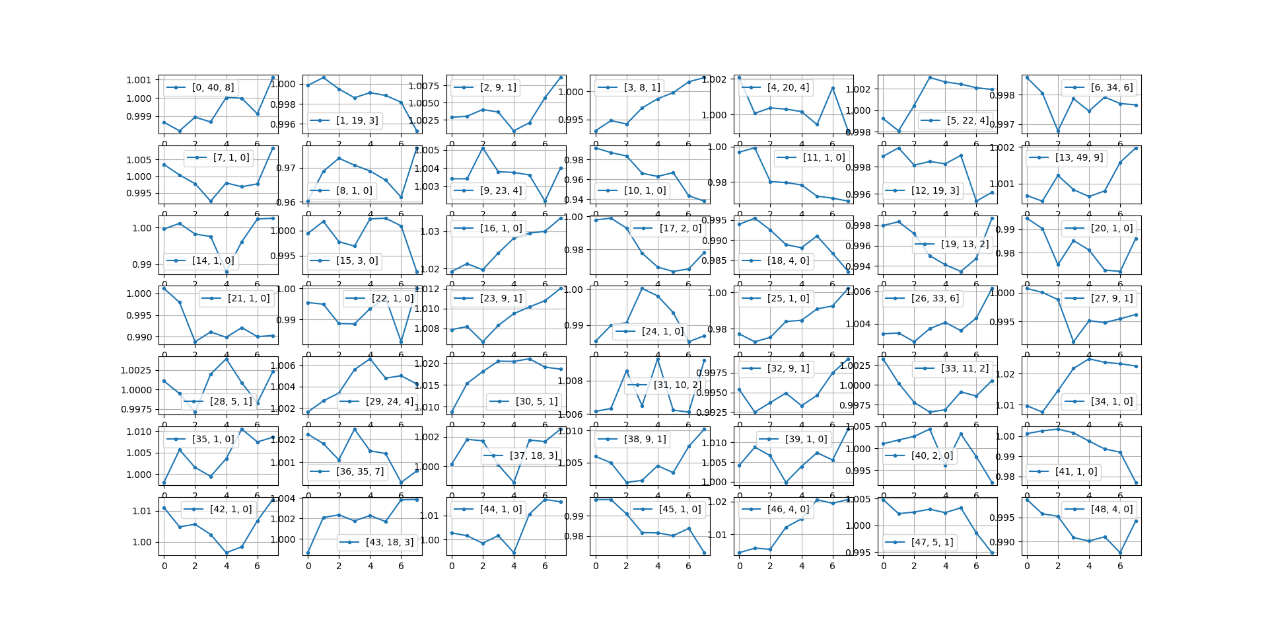
1. 思路：

**综合：就是选定某只股票（参数2），一段时期内（参数3）的某一周期（参数1）的不同类型的（参数5）交易数据，将其分类成N个类型（参数4）。这里的分类方法采用的是K-means来做的。**

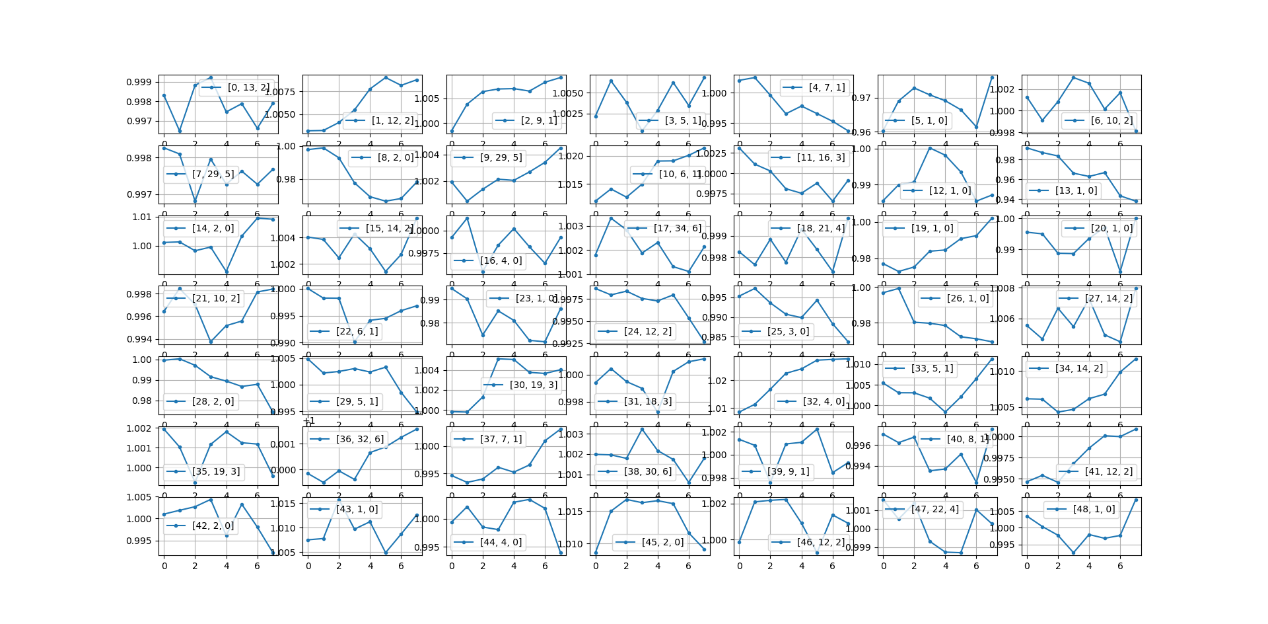
具体参数有

1. 参数1：选择时间周期：60min/30min/15min/5min
2. 参数2：选择股票代码
3. 参数3：选择数据的时间区间：这里的数据最早可以获得2016-02-05后的493个交易日数据
4. 参数4：分类的个数：默认49个
5. 参数5：分类类型：类型1：只包含收盘收益率；类型2：包含收盘涨跌幅和最高最低涨跌幅；类型3：包含类型2和成交量和成交价格特征
6. **例子：这里采用了上证指数，最近493个交易日内30分钟的价格和成交量的数据，将其聚类成了49个类型。**
7. 具体分类的均值情况：如下图：（图中的[0,40,8]分别代表第0个分类，发生了40次，发生概率为8%）

图需要放大来看，图中代表同类型的交易日的中心的均值情况，也就是从第一个收盘到第8个收盘的情况的分类。经过比对，有一定的代表意义。



类似的，如果我们把参数5改成类型1，只用收盘价分类，得到下面结果



## 2.1利用马氏链分析分时图中的状态转移情况。

1. 思路：

**总结：首先获得上证50所有成分股数据，这50只股票所有可以获得数据（参数1），然后对于所有的股票的30分钟k线（参数3）利用k-means进行分类，分类的类别数为（参数5），而这每一类就是一种K线的状态，有了状态后，我们将所有数据分为两个部分（比例为参数4,默认7:3，第一部分也就是前70%的股票，第二部分为后30%股票，前后两个部分都是不同的股票），最后估计两个部分的状态转移矩阵，进行比较。**

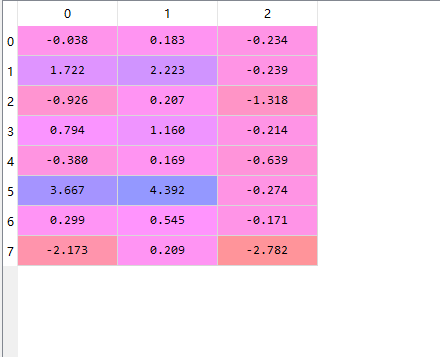
1. 参数1：数据:上证50指数所有提取到的数据。（数据代码和区间都可以选择）
2. 参数2：类型，选择收盘+最高+最低收盘价，（这里的类型可以和之前参数5一样去三种）
3. 参数3：单日k线周期，这里选择30min的。
4. 参数4：将所有取得的数据分为两个部分，用来对比彼此的状态转移矩阵，这里比例默认为0.3,也就是0.7和0.3
5. 参数5：对于单个30分钟k线的分类的个数
6. 案例和结果

首先获得上证50所有成分股，从2016-2-5日后可以获得的30分钟k线数据，分成两个部分第一部分数据由16221个交易日的30分钟K线构成，第二部分由7328个交易日的30分钟K线构成，将30K线数据的分别分类其8个状态状态。然后统计出这些状态的转移矩阵。（通过分类我们得到了30分钟k线只有8种状态，而这8种状态相互转化，下面就是讨论状态彼此转化的概率的问题）

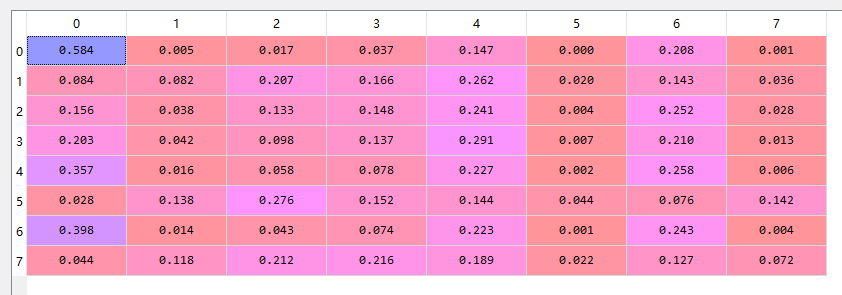
1. 我们将30分钟K线的分类列出来：

下图是一个8\*3的矩阵，第i行为第i个分类，第一列是30分钟k线收盘收益率，第二列为k线最高价对应的收益率，第3列是最低价的收益率。这里的表里的8类是kmeans处理后最后的每一类的均值。

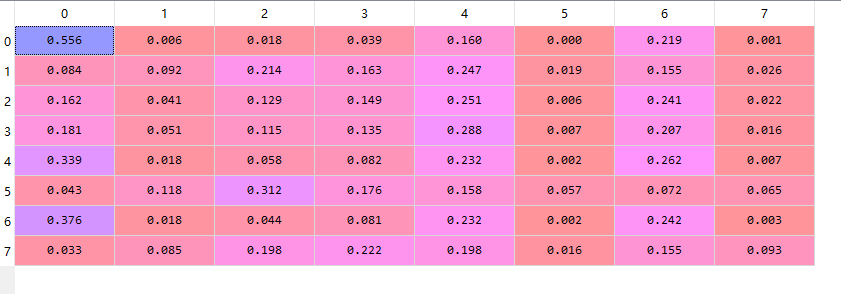
可以看到，这里的分类基本代表了股票30分钟上涨的各种情况。



2.接下来是第一部分的数据得到的状态转移矩阵（第I,j个元素，表示从第i个状态为出发，到达第j个状态的概率）



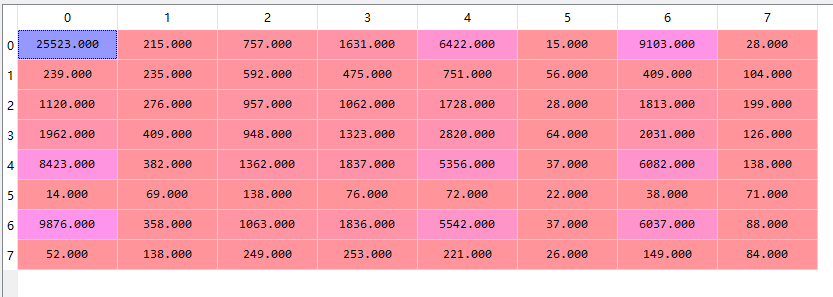
下面是第二部分的转移概率矩阵



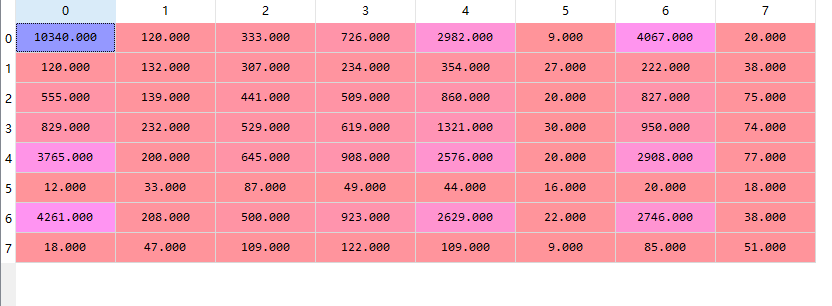
通过对比发现，状态转移矩阵的误差非常小，也就是不同股票的状态转移具有非常接近的特性。

更详细的，下面是发生次数，

第一部分的次数



第二部分的次数

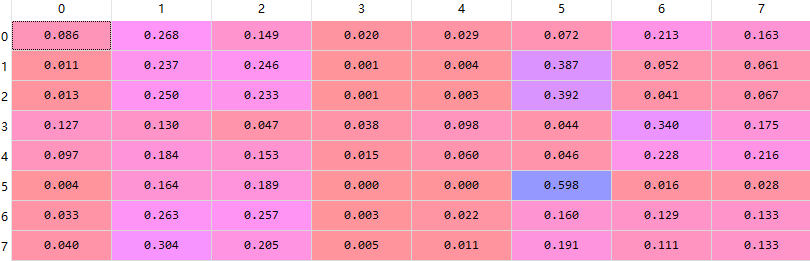


## 2.2 齐次马氏链的探索

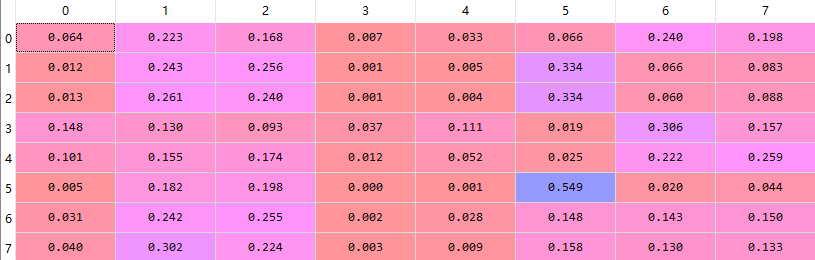
这里以上证50的成分股为例，分析30分钟k线状态转移的情况，分别得到前70%的时间的概率转移矩阵和后30%的时间概率转移矩阵是否一致

这里我们的时间、类型、与上面一致。

第一部分，前70%时间（15092个交易日的30分数据，这里之所以不是16221，是因为在数据提取是，有些数据提取失败，因此会少一些）



第二部分，后30%时间（6429个交易日的数据）



这里可以看到前后两段时间的转移阵接近是一样的，结果也比较理想。因此基本可以认为这里的转移阵不会随时间而改变，因此30分钟K线状态的马氏链是齐次马氏链。

## 3.基于马氏性对于日线级别的转移性质的分析

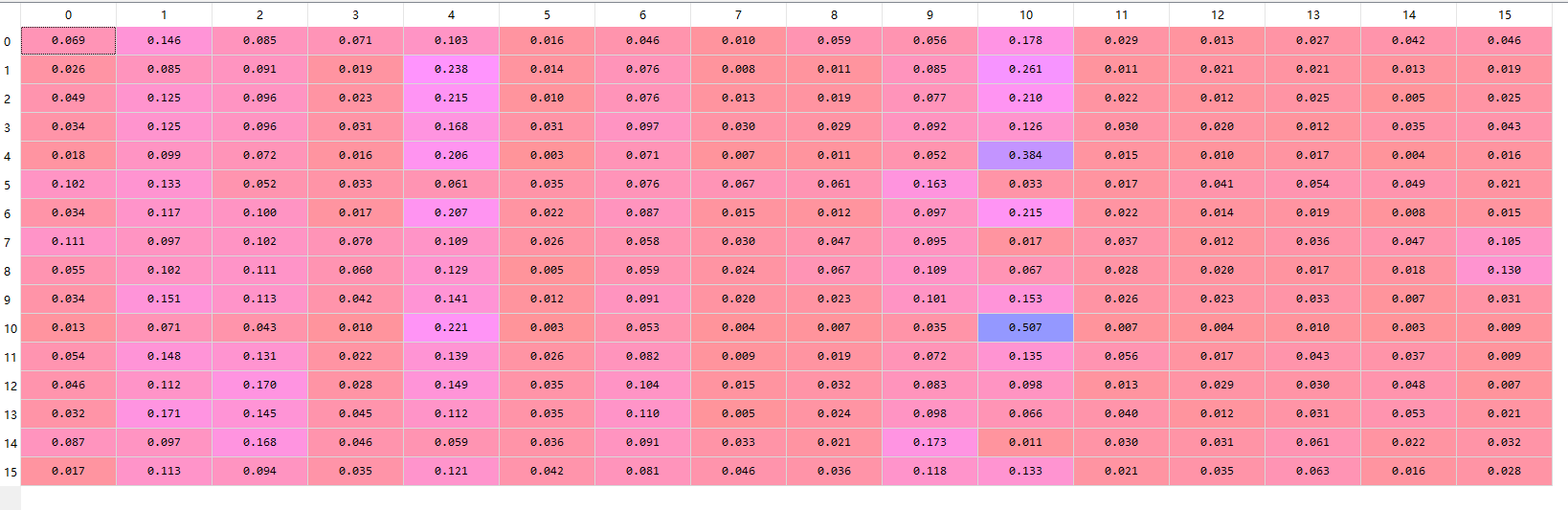
（1）思路，基于30分钟k线数据将日线数据分为n类，每一类代表一个状态，研究不同状态之间转移的概率

这里的数据与2.1案例的数据一致，而这里是研究日的状态的转移问题。（这里将数据分成了两个部分去研究）

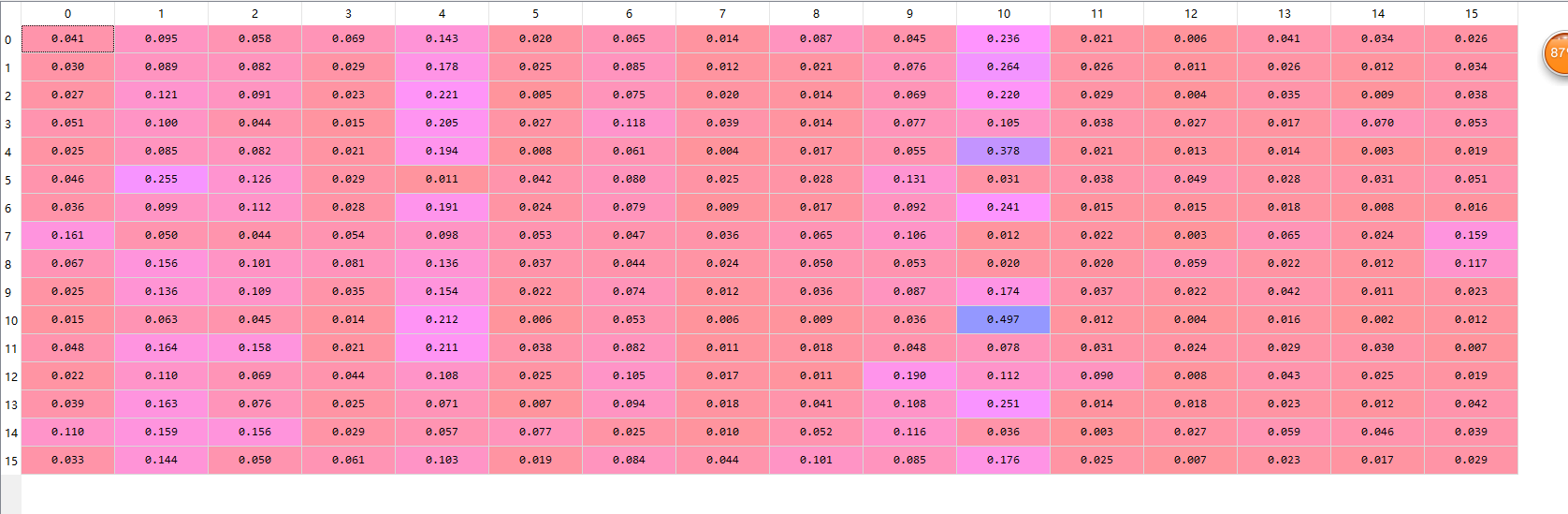
（2）具体案例

N=16,也就是将日线数据分为16种状态，经过数据分析得到的转移概率矩阵如下。（需要放大）

第一部分数据的转移矩阵



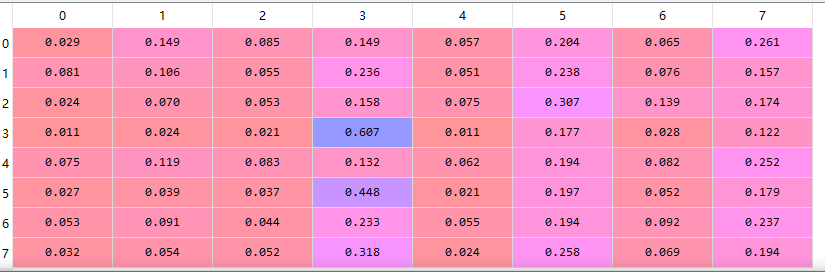
第二部分的数据转移矩阵



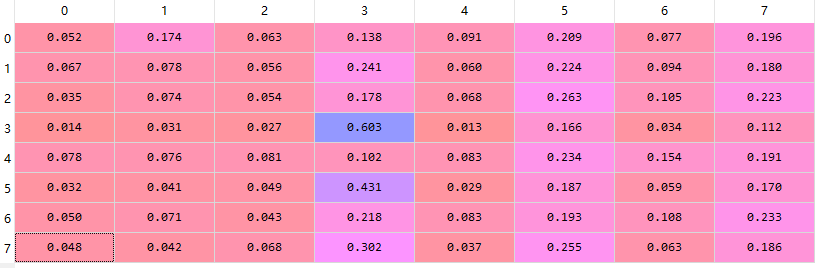
这里可以看到两个部分数据差别比较大，也就是日线的马氏性在目前来看不明显。

下面尝试减小分类情况n=8

第一部分：



第二部分：



可以看到转移概率比较接近，但是和之前30分钟的转移矩阵接近程度依然有一定差距，估计时由于样本数据不足的原因的。不过这里因为电脑限制，只能先处理的这么多数据了。

这里应该可以利用状态转化的马氏性，去构建最大效用函数，优化投资策略，后面还需要尝试。上述内容在参数选择上还有很多可以尝试的参数，暂时还没有去进一步调整。