

题目 5&6：隐马尔可夫模型

2018211316班 2018211568号 杜明欣

任务定义

本题目考查一个股票量化交易任务。假设状态个数 $|Q|=3$ ，可以简单地认为是“牛市”、“熊市”、“稳定市场”。状态到观测的发射模型是一个多项式分布。假设输出的观测个数 $|O|=5$ ，考虑序列长度 $T=100$ ，对应为 100 个连续的交易日。假设 HMM 的模型参数已知（已经通过 EM 算法在一个较大的数据集上完成了参数的估计），本题目的任务是根据模型参数，完成 100 个交易日上的推断，并预测接下来的 28 天（4 个礼拜）的观测值。

实验环境

windows 系统、spyder 软件

方法描述

• 任务一

使用 Forward-Backward 算法推断其背后的隐状态

利用前向迭代逐步求解不同时间对应隐状态概率

核心算法

```
def forward(a, tran, emis, O, t):  
    for i in range(N):  
        sum1=0  
        for j in range(N):  
            sum1+=a[t-1][j]*tran[j][i]*emis[i][O[t]-1]#利用放射概率和转移概率计算  
        下一时间步概率  
        a[t][i]=sum1  
    return a
```

• 任务二

viterbi 算法核心思想动态规划：每一时间步计算最大概率对应隐状态

再反向回溯记录整个隐状态序列

公式推导

Viterbi 动态规划

$$\text{设 } V_t(i) = \max_{1 \leq k \leq N} P(o_1, o_2, \dots, o_t, q_1, \dots, q_{t-1} = s_k, q_t = s_i | \lambda).$$

$$V_{t+1}(j) = \max_{1 \leq k \leq N} P(o_{1-t+1}, q_{1-t+1}, q_t = s_k, q_{t+1} = s_j | \lambda)$$

$$= \max_{1 \leq k \leq N} [P(o_{t+1} | o_{1-t}, q_{1-t+1}, \lambda) P(o_{1-t}, q_{1-t+1}, \lambda)]$$

$$= \max_{1 \leq k \leq N} [P(o_{t+1} | q_{t+1}) \cdot P(q_{t+1} | o_{1-t}, q_{1-t+1}, \lambda) P(o_{1-t}, q_{1-t+1} | \lambda)]$$

$$= \max_{1 \leq k \leq N} [P(o_{t+1} | q_{t+1}) \cdot P(q_{t+1} | q_t) V_t(k)]$$

核心算法

```
for i in range(1,102):
    for n in range(N):
        p = dp[i-1,:] * tran[:,n] * emis[n,o[i-1]-1] #利用发射概率和转移概率
        计算下一时间步概率
        pre[i,n] = np.argmax(p) #选择最大概率对应隐状态
        dp[i][n] = np.max(p)

    while i>1:
        path.append(pre[i][maxid]) #回溯记录隐状态序列
        maxid=pre[i][maxid]
        i=i-1
```

• 任务三

预测任务：

使用前向算法预测对应时刻隐状态概率，采用随机抽样方法依次抽取隐状态和对应的观测值

核心算法

```
for i in range(0,28):
    p=forward() #前向算法获取概率
    q=sample1(p) #随机抽样隐状态
    o=sample2(q) #随机抽样观测值
```

结果分析

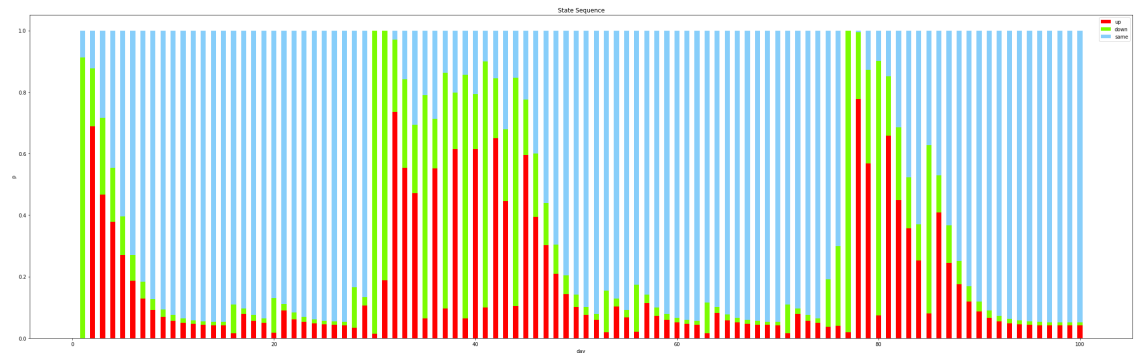
• 任务一

绘制 $t = 1, \dots, 100$ 上三个不同状态对应的状态序列

蓝色代表稳定、绿色代表熊市、红色代表牛市

由下图可见状态序列中平稳市场的概率更高

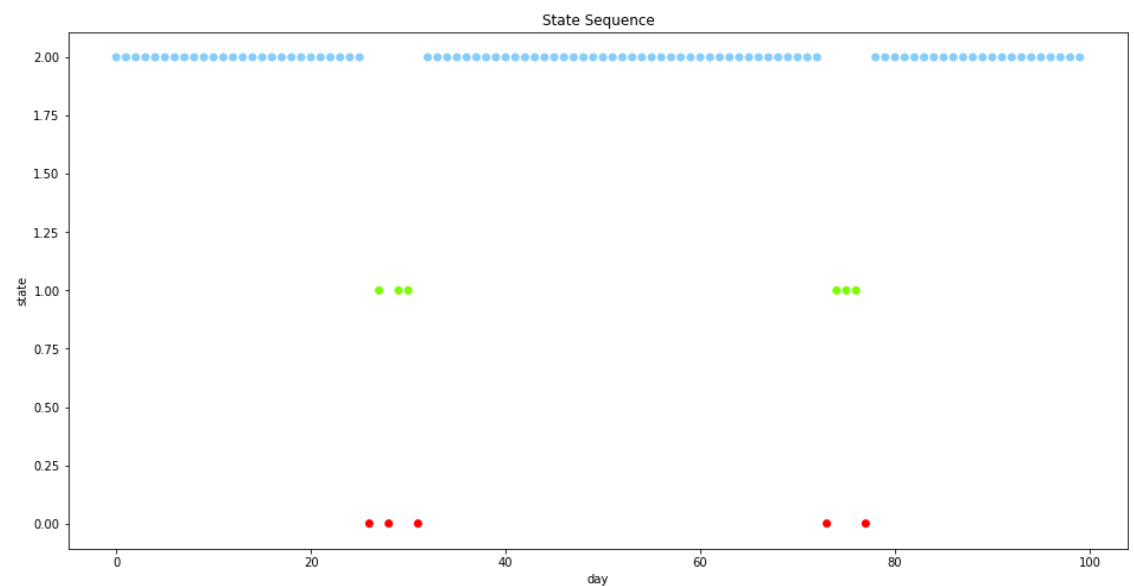
详细概率值见附录文件



• 任务二

推断时间序列 $t = 1, \dots, 100$ 上最可能的状态序列

蓝色代表稳定、绿色代表熊市、红色代表牛市



• 任务三

重复采样100次计算预测4周观测数据的标准差、均值及准确率

计算平均准确率约为0.66、均值方差详见附录文件

```
precise: 0.6610714285714284
mean: [2.97 3.01 2.96 2.97 2.96 2.99 3.01 2.89 3.07 3.    3.02 2.92 2.94 3.01
2.98 2.98 3.02 3.01 3.08 2.98 2.78 2.9   3.04 2.92 2.97 3.04 2.92 3.05]
std: [0.47864392 0.47947888 0.49507575 0.48989795 0.52877216 0.52306787
0.64529063 0.60819405 0.61611687 0.64776539 0.67793805 0.51951901
0.68818602 0.69253159 0.68673139 0.72801099 0.61514226 0.67349833
0.65505725 0.67705244 0.64311741 0.55452683]
```

Q: 如果你的老师计划投资股票, 你是否建议她使用上述你构建的模型

A: **不建议**用上述模型来真实投资炒股

1. 我们的HMM模型仅依赖于前一刻, 不能纵观全局, 有些股票交易行为涉及较远时间的数据, HMM没有长期记忆, 故不可取
2. 隐状态设计过于单一, 投资股票是否获利并不完全在于处于牛市、熊市, 还有很多其它复杂因素影响观测量
3. 投资不是0或1的选择, HMM仅提供一个粗略的股票涨势趋势图, 并未对仓位分配投资金额提供指导性意见
4. 上述模型的预测结果多为市场平稳并没有预测到足够切实可用信息, 如何投资交易还要凭借多年经验以及波澜不惊的心态

友情提醒: 股市有风险、投资需谨慎、抄底不规范、亲人两行泪!