

- content
{:toc}

HMM 隐马尔科夫模型

1 背景

1.1 数理统计学两大派

1. 贝叶斯学派和频率学派是当今数理统计学的两大学派，
频率学派，被称为古典学派。
贝叶斯学派，后期发展。

频率学派和贝叶斯学派的估计思想：

对于样本分布 $F(X, \theta)$ ，此时估计未知。

1) 频率学派（古典学派）

频率学派，对于一批样本，其分布 $F(X, Q)$ 是确定的，也即 Q 是确定的，只不过 Q 未知。

频率学派的基本宗旨，概率即是频率，某次得到的样本 X ，只是无数次可能的试验结果的一个具体实现，样本中未出现的结果不是不可能出现。只是这次抽样没有出现而已，因此综合考虑的已抽取到的样本以及未抽取的样本 X ，可以认为总体分布是确定的，不过未知，而样本来自于总体，故其样本分布 $F(X, Q)$ 也同样的特点。基于样本推断总体概率分布参数 Q 。

2) 贝叶斯学派

贝叶斯学派否定了，概率即是频率的观点，并且反对把样本 X 放到“无限多可能值之一”背景下去考虑，既然只得到样本 X ，那么就只能依靠它去做推断，而不能考虑那些可能出现而未出现的结果。与此同时，贝叶斯学派引入了主观概率的概念，认为一个事件在发生之前，人们应该对它是有所认知的， $F(X, \lambda)$ 中 Q 不是一个固定值，是一个随机变量，并且服从 $H(Q)$ 先验分布，当得到样本 X 后，我们对 Q 的分布有了新的认识， $H(Q)$ 发生改变即后验分布。此时对 Q 进行的估计，不再依赖样本，而完全依赖 Q 的后验分布。

马尔可夫性质

当前状态仅和上一状态有关，上一时刻之前的任何状态都无关，我们称其符合马尔可夫性质。

如果随机过程，E 为状态空间，任意 $t_1 < t_2 < t_3 < \dots < t_n < t_n + 1$, 任意的 $x_1, x_2, x_3, x_4 \in E$ 随机变量 $X(t)$ 在已知变量 $X(t_1)=x_1, \dots, X(t_n)=x_{n-1}$ 无关，即条件分布函数满足下列等式，此性质称为马尔可夫性，如果随机过程满足马尔可夫性，则该过程称为马尔可夫过程。

二、马尔可夫链

三元素： 状态空间S、转移概率矩阵P、初始概率分布 π

例子：马尔可夫链是指马尔可夫性质的随机过程。在过程中，在给定当前信息的情况下，过去的信息状态对于预测将来的状态是无关的。

例子：当前时间状态下的股价，可以转变下一时刻的股价，股价的转变即状态的改变。这个状态现在可以上升（股价提高），状态可以下降。我可以根据当前股票价格去决定下一刻股价上升、下降、不变的概率。这种股价变动的概率为状态转移概率。

例子：假设为认为股价有三种状态（高、中、低）；

- 1、状态空间 S 例：S 是一个集合，包含所有的状态 $S_{\text{股价}} = \text{高, 中, 低}$;
- 2、初始概率分布 π -例：
股价刚发行的时候有一个初始价格，我们认为初始价格为高的概率50%, 初始价格为中的概率为30% , 初始价格为低的概率是20%。 初始概率分布为： $\pi=(0.5, 0.3, 0.2)$
- 3、转移概率矩阵P-例：
现在有个股价为中，下一个时刻状态转变的可能性有三种，中→高、中→低、中→中；（n, n）矩阵。

初始概率 $\pi[0.1,0.6,0.3]$

$$P = \begin{bmatrix} 0.75 & 0.125 & 0.125 \\ 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

第n天	晴	阴	雨
0	0.1	0.6	0.3
1	0.45	0.3125	0.2375
2	0.5531	0.2531	0.1937
3	0.5898	0.2293	0.1809
4	0.6022	0.2215	0.1763
5	0.6065	0.2188	0.1747
6	0.6079	0.2179	0.1742
7	0.6084	0.2176	0.174
8	0.6086	0.2174	0.1739
9	0.6087	0.2174	0.1739
10	0.6087	0.2174	0.1739
11	0.6087	0.2174	0.1739

初始概率 $\pi[0.1, 0.6, 0.3]$

HMM

条件：隐状态必须是离散的

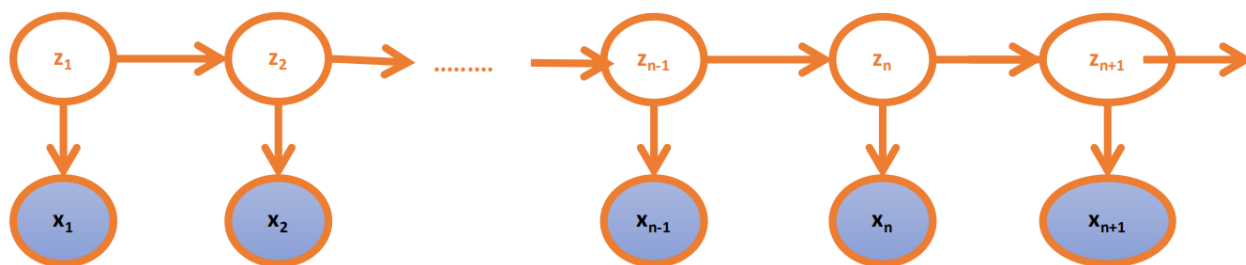
隐马尔可夫模型，是一种统计模型，在语音识别、行为识别、NLP故障诊断等领域具有高效性能。

HMM是关于时序的概率模型，描述一个含有未知参数的马尔可夫链所生成的不可观测的状态随机序列，再由各个状态生成观测随机序列的过程。

含有未知参数的马尔可夫链

HMM是一个双重随机过程——具有一定状态的隐马尔可夫链和随机的观测序列。

HMM随机生成的状态随机序列被称为状态序列；



思考： z_1, z_2, \dots, z_n 是不可观测的状态， x_1, x_2, \dots, x_n 是可观测量到的序列；不可观测的状态决定可观测量序列的值(z 的取值决定 x 的取值)；

1、在 z_1, z_2 不可观测的情况下， x_1 和 z_2 独立吗？ x_1 和 x_2 独立吗？

回答：这个问题可以回顾之前的贝叶斯网络来理解。

首先 z_1, z_2 都是离散的值，但 x_1 的值可能是离散的也可能是连续的。比如 z 是天气情况，每天天气的改变是离散的。 x 是因为天气而改变的一些其他状态，比如 x =(地面是否潮湿、路上行人数量、雨伞销售数量...)；

在 z_1 和 z_2 不可观测的情况下， x_1 和 z_2 不独立， x_1 和 x_2 也是不独立的。

2、在 z_1, z_2 可观测的情况下， x_1 和 z_2 独立吗？ x_1 和 x_2 独立吗？

回答：在 z_1 和 z_2 可观测的情况下，因为 x_1 和 z_2 的取值只和 z_1 有关，所以就独立了。同样在给定了 z_1 和 z_2 的情况下， x_1 和 x_2 也独立。

贝叶斯学派 和 频率学派的区别请看上面～

概率分布是不断变化的，后验分布决定

[贝叶斯算法](#) (这是一个博主：白尔摩斯，写了一系列文章)

2.1 一个模型，两个假设，三个问题区别:

- 有限历史假设 $p(s_i|s_{i-1}) = p(s_i|s_{i-1})$
- 齐次性假设, $P(s_{i+1}|s_i) = p(s_{j+1}, s_j)$
- 观测独立性假设, $P(o_t|q_t)$

HMM 解决的三个问题:

- 评估问题, 已知 $\lambda = (A, B, \pi)$, 计算某个观测序列发生的概率, 即求 $P(O|\lambda)$, 概率计算问题

例子: 前向-后向算法: 给定模型和观测序列 λ , 计算模型下观测到序列出现的概率 P ?

- 学习问题, 如何调整模型参数 $\lambda = (A, B, \pi)$, 使得 $P(O|\lambda)$ 最大?

Baum-Welch算法: 已知观测序列 $Q = \{q_1, q_2, q_3, \dots, q_T\}$, 估计模型的参数, 使得在该模型下观测序列下概率最大,

EM算法的一个特例, 专门用来求解HMM中的隐状态参数。即: 根据已知的观测序列, Q 序列, 去寻找整个模型的一组隐状态参数, 使得观测序列发生的可能性最大。

- 解码问题, 给出观测序列 O 和 U , 怎样选择一个隐藏状态序列 $S(s_1, s_2, s_3, s_4, \dots, s_{t+1})$

Viterbi 算法: 给定模型和观测序列, 求使观测序列条件概率, 最大的隐状态序列。

例子:

HMM案例回顾:

<https://www.jianshu.com/p/c80ca0aa4213>

概率计算问题

- 1、直接计算法
- 2、前向算法
- 3、后向算法

1、直接计算法(暴力算法)不推荐, 上面链接为原文

2、前向概率-后向概率 (快速递归求解)

1. 直接计算法

类似KNN计算最近邻时候的算法。

也就是说，暴力算法需要一个个遍历所有状态去计算当前状态发生的概率。

按照概率公式，列举所有可能长度T的状态序列 $I = \{i_1, i_2, \dots, i_T\}$, 求各个状态序列I 与观测序列 $Q = \{q_1, q_2, q_3, q_T\}$ 的联合概率 $P(Q, I; \lambda)$, 然后对所有可能的状态序列求和，从而得到最终的概率 $P(Q; \lambda)$;

分析

“白-黑-白-白-黑”结果的盒子序列的联合概率求出来 $P(Q, I; \lambda)$, 即 $\sum P(Q, I) = P(Q)$, $P(Q)$ 是我们观测到“白-黑-白-白-黑”,