Statistical Computing
Chap. 2.2: Hidden Markov Model

LIU, Ran

Department of Statistics,
Beijing Normal University (Zhuhai Campus)

March 25, 2024

Summary

介绍

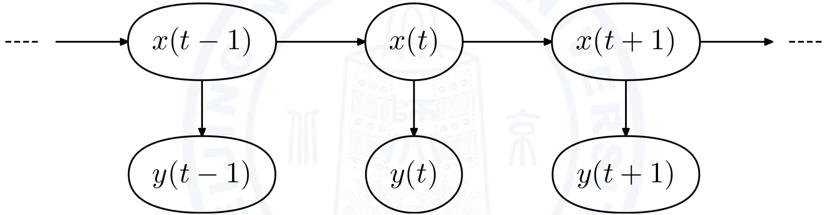
介绍

隐藏马尔可夫模型（Hidden Markov Model，HMM）是一种用于建模序列数据的概率模型。它是马尔可夫链模型的扩展，用于描述具有潜在未观察状态的动态过程。

HMM 由两个基本组成部分构成：隐藏状态（hidden state）和观察序列（observation sequence）。隐藏状态是不可直接观察到的状态，而观察序列是可观察到的数据序列。

HMM 的基本假设是，隐藏状态的转移和观察结果只依赖于有限的历史状态，即具有马尔可夫性质。这意味着在任何给定时间步骤，隐藏状态只与前一时间步骤的隐藏状态有关。

我们假设观测到的序列是 $Y = \{y_1, y_2, ..., y_T\}$，隐含的状态序列为 $X = \{x_1, x_2, ..., x_T\}$.

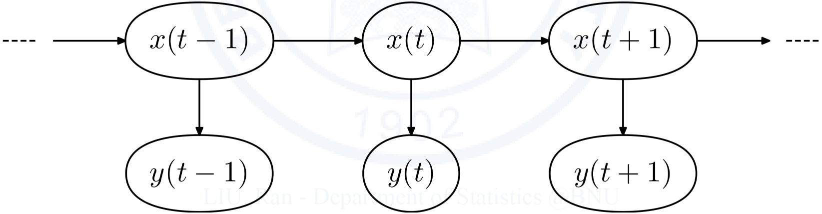


状态的转移满足马尔可夫性质，即

$$p(x_t|x_{t-1}, x_{t-2}, ...) = p(x_t|x_{t-1})$$

数据的产生

1. 由初始状态概率分布产生初始状态 x_1.
2. 初始状态 x_1 下，用观察概率矩阵生成 y_1.
3. 用状态转移概率矩阵，产生第二个状态 x_2.
4. 第二个状态 x_2 下，用观察概率矩阵生成 y_2.
5. 重复这一过程.

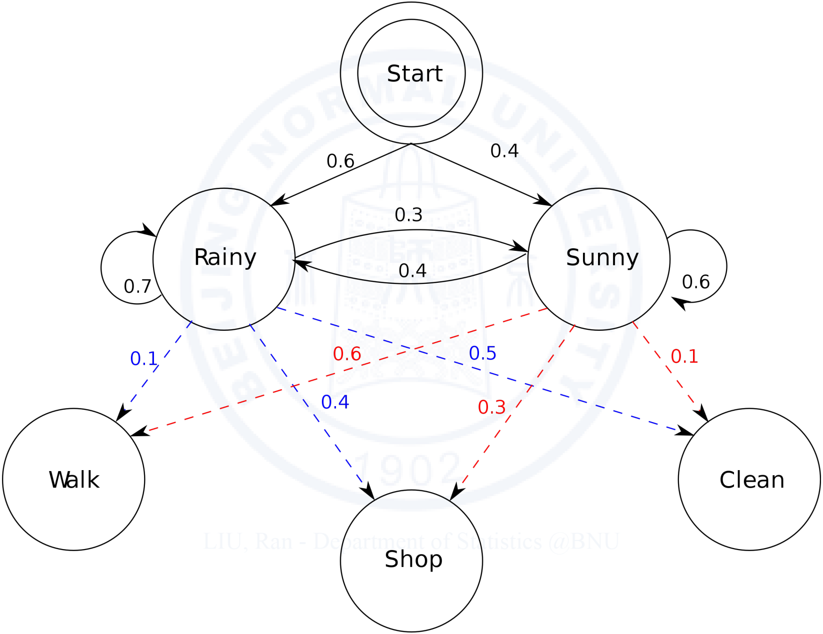


HMM 包含以下几个要素:

- 隐藏状态集合（Hidden State Set）(固定的): 表示系统可能处于的一组隐藏状态 $S = \{s_1, s_2, ..., s_N\}$。
- 离散观测数据的集合（固定的): 表示系统可能出现的一组观测值 $O = \{o_1, o_2, ..., o_K\}$

- 初始状态概率分布（Initial State Distribution）：表示系统在初始状态时处于每个隐藏状态的概率分布，通常用 Π，其中 $\pi_i = p(x_1 = s_i)$ 表示系统初始状态为 s_i 的概率。
- 状态转移概率矩阵（State Transition Probability Matrix）：表示系统从一个隐藏状态转移到另一个隐藏状态的概率分布，通常用 $A = \{a_{ij}\}$ 表示，其中 a_{ij} 表示系统从状态 s_i 转移到状态 s_j 的概率，$a_{ij} = p(x_t = s_j | x_{t-1} = s_i)$。
- 观察概率矩阵（Observation Probability Matrix）：表示在给定隐藏状态下观察到特定观察结果的概率分布，通常用 $B = \{b_{ij}\}$ 表示，其中 b_{ij} 表示在状态 s_i 下观察到观察结果 o_j 的概率，$p(y_t = o_j | x_t = s_i)$.

X 为 {Rainy, Sunny}, Y 为 {Walk, Shop, Clean}.



HMM 的三个基本问题

在 HMM 中，有三个基本问题需要解决：

1. **学习问题（Learning Problem）**：给定观测序列，估计模型参数。
2. **评估问题（Evaluation Problem）**：给定模型参数和观测序列，计算观测序列的概率。
3. **解码问题（Decoding Problem）**：给定模型参数和观测序列，找到最可能的隐藏状态序列。

解决学习问题

给定观测序列 $Y = (y_1, y_2, \ldots, y_T)$，估计 HMM 模型参数 $\theta = (A, B, \pi)$ 的方法：

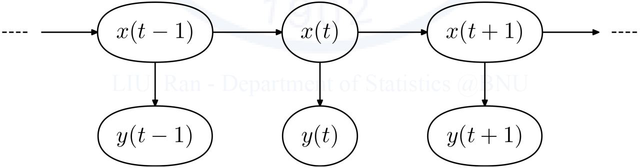
Baum-Welch 算法（Expectation-Maximization Algorithm）：通过迭代计算前向后向概率和状态转移概率，估计模型参数。

令 $\theta = (A, B, \Pi)$，我们根据 HMM 产生数据的方式写出完全数据的似然函数

$$
\begin{aligned}
p(X, Y|\theta) =& p_\theta(x_1)p_\theta(y_1|x_1) \cdot p_\theta(x_2|x_1)p_\theta(y_2|x_2) \\
& p_\theta(x_3|x_2)p_\theta(y_3|x_3) \cdot p_\theta(x_4|x_3)p_\theta(y_4|x_4)... \\
=& p_\theta(x_1)p_\theta(y_1|x_1) \prod_{t=1}^{T-1} p_\theta(x_{t+1}|x_t)p_\theta(y_{t+1}|x_{t+1})
\end{aligned}
$$

求对数之后，我们有

$$
\ell(\theta|X, Y) = \log p_\theta(x_1) + \sum_{t=1}^{T-1} \log p_\theta(x_{t+1}|x_t) + \sum_{t=1}^{T} \log p_\theta(y_t|x_t)
$$



将概率用参数表达:

$$\ell(\theta|X,Y) = \sum_{i=1}^{N} I(x_1 = s_i) \log \pi_i + \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{j=1}^{N} I(x_t = s_i, x_{t+1} = s_j) \log a_{ij} + \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{K} I(x_t = s_i, y_t = o_j) \log b_{ij}$$

求条件期望，得到 Q 函数：

$$
\begin{aligned}
Q(\theta \mid \theta^{(l)}) =& E[\ell(\theta|X,Y) \mid Y, \theta^{(l)}] \\
=& \sum_{i=1}^{N} p(x_1 = s_i \mid Y, \theta^{(l)}) \log \pi_i \qquad (1) \\
&+ \sum_{t=1}^{T-1} \sum_{i=1}^{N} \sum_{j=1}^{N} p(x_t = s_i, x_{t+1} = s_j \mid Y, \theta^{(l)}) \log a_{ij} \qquad (2) \\
&+ \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{K} I(y_t = o_j) p(x_t = s_i \mid Y, \theta^{(l)}) \log b_{ij} \qquad (3)
\end{aligned}
$$

我们分别求这三个条件概率。

我们先计算 $p(x_t \mid Y, \theta^{(l)})$:

我们考虑到马尔可夫的性质，我们在时间 t 处分开观测值

$$\begin{array}{ccc|ccc} & & x_t & & & \\ y_1, & \cdots, & y_t & y_{t+1}, & \cdots, & y_T \end{array}$$

则我们有

$$\begin{aligned} p(x_t \mid Y, \theta^{(l)}) &\propto p(x_t, Y \mid \theta^{(l)}) \\ &\propto p(y_1, y_2, \cdots, y_t, x_t \mid \theta^{(l)}) \\ &\cdot p(y_{t+1}, y_{t+2}, \cdots, y_T \mid y_1, y_2, \cdots, y_t, x_t, \theta^{(l)}) \\ &\propto p(y_1, y_2, \cdots, y_t, x_t \mid \theta^{(l)}) & \text{(a)} \\ &\cdot p(y_{t+1}, y_{t+2}, \cdots, y_T \mid x_t, \theta^{(l)}) & \text{(b)} \end{aligned}$$

Forward

我们先看(a)式，定义：

$$\alpha_i^{(l)}(t) := p(y_1, y_2, \cdots, y_t, x_t = s_i \mid \theta^{(l)})$$

想用归纳法，找出计算 $\alpha_i^{(l)}(t)$ 的递归公式：

$$p(y_1, x_1 = s_i \mid \theta^{(l)}) = p(y_1 \mid x_1 = s_i, \theta^{(l)}) p(x_1 = s_i | \theta^{(l)})$$

$$p(y_1, y_2, x_2 = s_i \mid \theta^{(l)}) = \sum_j p(y_1, y_2, x_1 = s_j, x_2 = s_i \mid \theta^{(l)})$$

$$= \sum_j p(x_1 = s_j, y_1 \mid \theta^{(l)}) p(x_2 = s_i \mid x_1 = s_j, \theta^{(l)}) p(y_2 \mid x_2 = s_i, \theta^{(l)})$$

$$p(y_1, y_2, y_3, x_3 = s_i \mid \theta^{(l)}) = \sum_j p(y_1, y_2, y_3, x_2 = s_j, x_3 = s_i \mid \theta^{(l)})$$

$$= \sum_j p(y_1, y_2, x_2 = s_j \mid \theta^{(l)}) p(x_3 = s_i \mid x_2 = s_j, \theta^{(l)}) p(y_3 \mid x_3 = s_i, \theta^{(l)})$$

所以我们能总结出递归公式:

$$\alpha_i^{(l)}(t) = \sum_{j=1}^{N} \alpha_j^{(l)}(t-1) a_{ji}^{(l)} b_i^{(l)}(y_t)$$

其中 $\alpha_i^{(l)}(1) = \pi_i b_i(y_1)$, $b_i(y_t) = \sum_{k=1}^{K} I(y_t = o_k) b_{ik}$, 我们将下列计算流程称为向前算法 (Forward):

$$\alpha_i^{(l)}(1) \to \alpha_i^{(l)}(2) \to \cdots \to \alpha_i^{(l)}(T), \quad i = 1, 2, ..., N$$

它实际求的是

$$p(y_1, x_1 = s_i \mid \theta^{(l)}) \to \cdots \to p(y_1, y_2, \cdots, y_T, x_T = s_i \mid \theta^{(l)})$$

Backward

我们再来看(b)，定义：

$$\beta_i^{(l)}(t) = p(y_{t+1}, y_{t+2}, ..., y_T \mid x_t = s_i, \theta^{(l)})$$

想用归纳法，找出计算 $\beta_i^{(l)}(t)$ 的递归公式：

$$p(y_T \mid x_{T-1} = s_i, \theta^{(l)}) = \sum_j p(y_T \mid x_T = s_j, \theta^{(l)}) p(x_T = s_j \mid x_{T-1} = s_i, \theta^{(l)})$$

$$p(y_{T-1}, y_T \mid x_{T-2} = s_i, \theta^{(l)}) = \sum_j p(y_{T-1}, y_T, x_{T-1} = s_j \mid x_{T-2} = s_i, \theta^{(l)})$$

$$= \sum_j p(y_{T-1}, y_T \mid x_{T-1} = s_j, \theta^{(l)}) p(x_{T-1} = s_j \mid x_{T-2} = s_i, \theta^{(l)})$$

$$= \sum_j p(y_{T-1} \mid x_{T-1} = s_j, \theta^{(l)}) p(y_T \mid x_{T-1} = s_j, \theta^{(l)}) p(x_{T-1} = s_j \mid x_{T-2} = s_i)$$

所以我们有:

$$\beta_i^{(l)}(t) = \sum_{j=1}^{N} b_j^{(l)}(y_{t+1})\beta_j^{(l)}(t+1)a_{ij}^{(l)}$$

其中 $\beta_i^{(l)}(T) = 1$，$b_i^{(l)}(y_{t+1}) = \sum_{k=1}^{K} I(y_{t+1} = o_k)b_{ik}^{(l)}$，我们将下列计算流程称为向后算法 (Backward):

$$\beta_i^{(l)}(T) \to \beta_i^{(l)}(T-1) \to \cdots \to \beta_i^{(l)}(1), \quad i = 1, 2, ..., N$$

它实际求的是

$$p(y_T \mid x_{T-1} = s_i, \theta^{(l)}) \to \cdots \to p(y_2, y_3, ..., y_T \mid x_1 = s_i, \theta^{(l)})$$

我们有

$$p(x_t = s_i \mid Y, \theta^{(l)}) \propto \alpha_i^{(l)}(t)\beta_i^{(l)}(t)$$

$$\gamma_i^{(l)}(t) := p(x_t = s_i \mid Y, \theta^{(l)}) = \frac{\alpha_i^{(l)}(t)\beta_i^{(l)}(t)}{\sum_{i=1}^{N} \alpha_i^{(l)}(t)\beta_i^{(l)}(t)}$$

我们再看看

$$\begin{aligned}\xi_{ij}^{(l)}(t) &:= p(x_t = s_i, x_{t+i} = s_j \mid Y, \theta^{(l)}) \\ &\propto p(x_t = s_i, x_{t+i} = s_j, Y \mid \theta^{(l)})\end{aligned}$$

还是在 t 时刻划分，上式等于

$$
\begin{aligned}
& p(x_t = s_i, x_{t+1} = s_j, Y \mid \theta^{(l)}) \\
= & \, p(y_1, y_2, ..., x_t = s_i \mid \theta) \cdot p(y_{t+1}, y_{t+2}, ..., y_T, x_{t+1} = s_j \mid x_t = s_i, \theta) \\
= & \, p(y_1, y_2, ..., x_t = s_i \mid \theta) \cdot p(y_{t+2}, ..., y_T, x_{t+1} = s_j \mid x_t = s_i, \theta) \\
& \cdot p(y_{t+1} \mid x_{t+1} = s_j) p(x_{t+1} = s_j \mid x_t = s_i) \\
= & \, \alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}
\end{aligned}
$$

所以

$$
\xi_{ij}^{(l)}(t) = \frac{\alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}}{\sum_i \sum_j \alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}}
$$

回到最初的三个条件分布，我们有

$$
\begin{aligned}
(1) &= \sum_{i=1}^{N} p(x_1 = s_i \mid Y, \theta^{(l)}) \log \pi_i \\
&= \sum_{i=1}^{N} \gamma_i^{(l)}(1) \log \pi_i \\
(2) &= \sum_{t=1}^{T-1} \sum_{i=1}^{N} \sum_{j=1}^{N} p(x_t = s_i, x_{t+1} = s_j \mid Y, \theta^{(l)}) \log a_{ij} \\
&= \sum_{t=1}^{T-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \xi_{ij}^{(l)}(t) \log a_{ij} \\
(3) &= \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{K} I(y_t = o_j) p(x_t = s_i \mid Y, \theta^{(l)}) \log b_{ij} \\
&= \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} I(y_t = o_j) \gamma_i^{(l)}(t) \log b_{ij}
\end{aligned}
$$

则我们的 $Q(\theta \mid \theta^{(l)})$ 为三者相加，我们求导算参数的估计值，则有

$$\frac{\partial Q}{\partial \pi_i} = \frac{\gamma_i^{(l)}(1)}{\pi_i} \Longrightarrow \pi_i^{(l+1)} = \frac{\gamma_i^{(l)}(1)}{\sum_{i=1}^{N} \gamma_i^{(l)}(1)}$$

$$\frac{\partial Q}{\partial a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_{ij(t)}}{a_{ij}} \Longrightarrow a_{ij}^{(l+1)} = \frac{\sum_{t=1}^{T-1} \xi_{ij(t)}^{(l)}}{\sum_{j=1}^{N} \sum_{t=1}^{T-1} \xi_{ij(t)}^{(l)}}$$

以及

$$\frac{\partial Q}{\partial b_{ik}} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_i^{(l)}(t) I(y_t = o_k)}{b_{ik}}$$

$$\Longrightarrow b_{ik}^{(l+1)} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_i^{(l)}(t) I(y_t = o_k)}{\sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_i^{(l)}(t) I(y_t = o_k)}$$

总结流程

1. 初始化 $\theta_0 = (\Pi, A, B)$.
2. 向前和向后算法计算 α 和 β.
3. 计算 $\theta^{(l+1)}$.