HMM——定义和假设

一些基本概念

在正式讲解隐马尔可夫模型（Hidden Markov Model，HMM）之前，有几个概念需要搞清楚。

概率模型（Probabilistic Model）

所谓**概率模型**，顾名思义，就是将学习任务归结于计算变量的概率分布的模型。

概率模型非常重要。在生活中，我们经常会根据一些已经观察到的现象来推测和估计未知的东西——这种需求，恰恰是概率模型的推断（Inference）行为所做的事情。

推断（Inference）的本质是：利用可观测变量，来推测未知变量的**条件分布。**

我们下面要讲的隐马尔可夫模型（HMM）和条件随机场（CRF）都是概率模型，之前讲过的朴素贝叶斯和逻辑回归也是概率模型。

生成模型 VS 判别模型

概率模型又可以分为两类：**生成模型**（Generative Model）和**判别模型**（Discriminative Model）。这两种模型有什么不同呢？我们来看一下。

既然概率模型是通过可观测变量推断部分未知变量，那么我们将可观测变量的集合命名为 O，我们感兴趣的未知变量的集合命名为 Y。

生成模型学习出来的是 O 与 Y 的联合概率分布 $P(O,Y)$，而判别模型学习的是条件概率分布：$P(Y|O)$。

之前我们学过的朴素贝叶斯模型是生成模型，而逻辑回归则是判别模型。

对于某一个给定的观察值 O，运用条件概率 $P(Y|O)$ 很容易求出它对于不同 Y 的取值。

那么当遇到分类问题时，直接就可以运用判别模型——给定 O 对于哪一个 Y 值的条件概率最大——来判断该观测样本应该属于的类别。

生成模型直接用来给观测样本分类有点困难。当然也不是不可行，通过运用贝叶斯法则，可以将生成模型转化为判别模型，但是这样显然比较麻烦。

在分类问题上，判别模型一般更具优势。不过生成模型自有其专门的用途，下面我们要讲的 HMM，就是一种生成模型。

概率图模型（Probabilistic Graphical Model）

我们还需要明确一个很重要的概念：概率图模型。

概率图模型：是一种以图（Graph）为表示工具，来表达变量间相关关系的概率模型。

这里说的图就是“数据结构”课中讲过的图概念：**一种由节点和连接节点的边组成的数据结构。**

在概率图模型中，一般用节点来表示一个或者一组随机变量，而节点之间的边则表示两个（组）变量之间的概率相关关系。

边可以是有向（有方向）的，也可以是无向的。概率图模型大致可以分为：

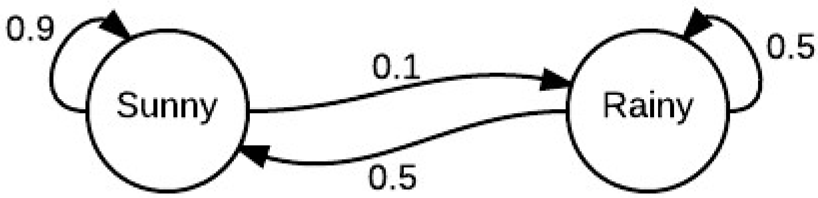
- **有向图模型**（贝叶斯网络）：用有向无环图表示变量间的依赖关系；
- **无向图模型**（马尔可夫网）：用无向图表示变量间的相关关系。

HMM 就是贝叶斯网络的一种——虽然它的名字里有和"马尔可夫网"一样的"马尔可夫"。

对变量序列建模的贝叶斯网络又叫做动态贝叶斯网络。HMM 就是最简单的动态贝叶斯网络。

马尔可夫链，马尔可夫随机场和条件随机场

马尔可夫链（Markov Chain）：一个随机过程模型，它表述了一系列可能的事件，在这个系列当中每一个事件的概率仅依赖于前一个事件。



上图就是一个非常简单的马尔可夫链。两个节点分别表示晴天和雨天，几条边表示节点之间的转移概率。

一个晴天之后，0.9的可能是又一个晴天，只有0.1的可能是一个雨天。而一个雨天之后，0.5的可能是晴天，也有0.5的可能是另外一个雨天。

假设这是某个地区的天气预报模型（这个地区只有晴天和雨天两种天气），则明天天气的概率，只和今天的天气状况有关，和前天以及更早没有关系。那么我们只要知道今天的天气，就可以推测明天是晴是雨的可能性了。

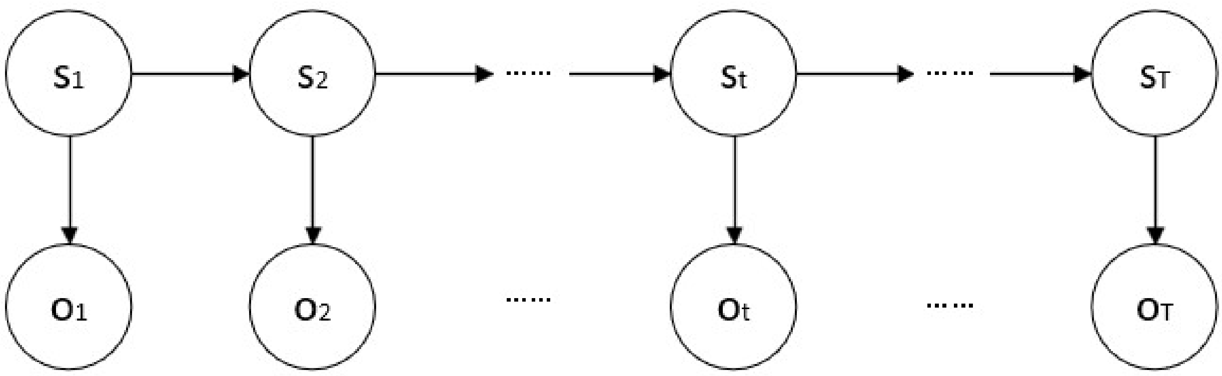
隐马尔可夫模型（Hidden Markov Model，HMM）

HMM 定义

HMM 是一个**关于时序的概率模型**，它的**变量分为两组**：

- 状态变量 $\{s_1, s_2, \ldots, s_T\}$, 其中 $s_t \in \mathcal{S}$ 表示 t 时刻的系统状态；
- 观测变量 $\{o_1, o_2, \ldots, o_T\}$, 其中$o_t \in \mathcal{O}$ 表示 t 时刻的观测值。

状态变量和观测变量各自都是一个时间序列，每个状态/观测值都和一个时刻相对应（见下图，图中箭头表示依赖关系）：



一般假定状态序列是隐藏的、不能被观测到的，因此**状态变量是隐变量**（Hidden Variable）——这就是 HMM 中 H（Hidden）的来源。

这个隐藏的、不可观测的**状态序列是由一个马尔可夫链随机生成的**——这是 HMM 中的第一个 M（Markov）的含义。

一条隐藏的马尔可夫链随机生成了一个不可观测的**状态序列**（State Sequence），然后**每个状态又对应生成了一个观测结果**，这些观测值按照时序排列后就成了**观测序列**（Observation Sequence）。这两个序列是一一对应的，每个对应的位置又对应着一个时刻。

一般而言，HMM 的状态变量取值是离散的，而观测变量的取值，则可以是离散的，也可以是连续的。

不过为了方便讨论，也因为在大多数应用中观测变量也是离散的，因此，我们下面仅讨论状态变量和观测变量都是离散的情况。

HMM 基本假设

HMM 的定义建立在两个假设之上：

假设1： 假设隐藏的马尔可夫链在任意时刻 t 的状态只依赖于前一个时刻（t-1 时）的状态，与其他时刻的状态及观测无关，也与时刻 t 无关。

用公式表达就是：

$$P(s_t|s_{t-1},o_{t-1},\dots,s_1,o_1)=P(s_t|s_{t-1}),\quad t=1,2,\dots,T$$

这一假设又叫做**齐次马尔可夫假设。**

假设2：假设任意时刻的观测只依赖于该时刻的马尔可夫链状态，与其他观测及状态无关。

用公式表达为：

$$P(o_t|s_T,o_T,s_{T-1},o_{T-1},\dots,s_{t+1},o_{t+1},s_t,o_t,\dots,s_1,o_1)=P(o_t|s_t)$$

这叫**观测独立性假设。**

确定 HMM 的两个空间和三组参数

基于上述两个假设，我们可知：所有变量（包括状态变量和观测变量）的联合分布为：

$$P(s_1,o_1,\dots,s_T,o_T)=P(s_1)P(o_1|s_1)\prod_{t=2}^{T}P(s_t|s_{t-1})P(o_t|s_t).$$

设 HMM 的状态变量（离散型），总共有 N 种取值，分别为：$\{S_1,S_2,\dots,S_N\}$。

观测变量（也是离散型），总共有 M 种取值，分别为：$\{O_1,O_2,\dots,O_M\}$。

那么，要确定一个 HMM，除了要指定其对应的状态空间 $\mathcal{S}$ 和观测空间$\mathcal{O}$ 之外，还需要**三组参数**，分别是：

- **状态转移概率**：模型在各个状态间转换的概率，通常记作矩阵 $A=[a_{ij}]_{N\times N}$。

  其中，$a_{ij}=P(s_{t+1}=S_j|s_t=S_i),1\leqslant i,j\leqslant N$ 表示在任意时刻 t，若状态为 S_i，则下一时刻状态为 S_j 的概率。

- **输出观测概率**：模型根据当前状态获得各个观测值的概率，通常记作矩阵 $B=[b_{ij}]_{N\times M}$。

  其中，$b_{ij}=P(o_t=O_j|s_t=S_i),1\leqslant i\leqslant N,1\leqslant j\leqslant M$ 表示在任意时刻 t，若状态为 S_i,则观测值 O_j 被获取的概率。

  有些时候，S_i 已知，但可能 O_j 是未知的，这个时候，b 就成了当时观测值的一个函数，因此也可以写作 $b_i(o)=P(o|s=S_i)$。

- **初始状态概率**：模型在初始时刻各状态出现的概率，通常记作 $\pi=(\pi_1,\pi_2,\dots,\pi_N)$，其中 $\pi_i=P(s_1=S_i),1\leqslant i\leqslant N$ 表示模型的初始状态为 S_i的概率。

通常我们用 $\lambda = [A, B, \pi]$ 来指代这三组参数。

有了状态空间 $\mathcal{S}$ 和观测空间 $\mathcal{O}$ ，以及参数 λ ，一个 HMM 就被确定了。