Multivariate Models

桂昊坤
HITSZ

2022 年 11 月 11 日

大纲

混合模型

定义

混合模型

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{y})$$

其中 p_k 是第 k 个混合分量，且 π_k 是满足 $0 \leq \pi_k \leq 1$ 和 $\sum_{k=1}^{K} \pi_k = 1$ 的混合权重。

隐变量

引入隐变量 $z \in \{1, \ldots, K\}$, 其用以指定产生 y 的分布。

定义

Joint Model

$$\begin{aligned} p(z|\boldsymbol{\theta}) &= Cat(z|\boldsymbol{\pi}) \\ p(\mathbf{y}|z=k,\boldsymbol{\theta}) &= p(\mathbf{y}|\boldsymbol{\theta_k}) \end{aligned}$$

其中, $\boldsymbol{\theta} = (\pi_1, \ldots, \pi_K, \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_K})$ 为模型参数。

通过边缘化 z 我们可以导出混合模型的公式:

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(z=k|\boldsymbol{\theta})p(\mathbf{y}|z=k,\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k p_k(\mathbf{y}|\boldsymbol{\theta})$$

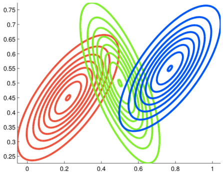
定义

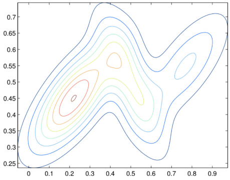
定义

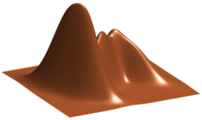
高斯混合模型 (GMM)

$$p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \Sigma_{\boldsymbol{k}})$$

图像







- 上述例子为二维中的 3 个高斯分布的混合，其中每个混合物成分都由左图中一组不同的椭圆轮廓表示。
- 右图是总体密度分布图像

应用

GMM 常用于无监督聚类。

步骤

- 通过计算 MLE 训练模型
- 联系每个数据点 y_n 同离散的隐变量 z_n

可以通过计算后验概率来计算出最有可能的隐变量

$$r_{nk} \triangleq p(z_n = k|\boldsymbol{y_n}, \boldsymbol{\theta}) = \frac{p(z_n = k|\boldsymbol{\theta})p(\boldsymbol{y_n}|z_n = k, \boldsymbol{\theta})}{\sum_{k'=1}^{K} p(z_n = k'|\boldsymbol{\theta})p(\boldsymbol{y_n}|z_n = k', \boldsymbol{\theta})}$$

应用

定义

硬聚类

$$\hat{z}_n = \arg\max_k r_{nk} = \arg\max_k [\log p(\boldsymbol{y}_n | z_n = k, \boldsymbol{\theta}) + \log p(z_n = k | \boldsymbol{\theta})]$$

软聚类

被指派了一个聚类分组，此外还指定了该样本属于该聚类分组的概率，即该样本不一定属于该聚类分组，有一定几率属于其他聚类分组

同 K-Means 关系

如果有一个关于 z_n 统一的先验，且使用 $\sum_k = \mathbf{I}$ 的球型高斯分布，则硬聚类方法简化成：

$$z_n = \arg\min_k \|\boldsymbol{y}_n - \hat{\boldsymbol{\mu}}_k\|_2^2$$

伯努利混合模型

如果数据是二值的，我们可以使用伯努利混合模型，且每一个混合成份有以下形式：

$$p(\boldsymbol{y}|z=k,\boldsymbol{\theta}) = \prod_{d=1}^{D} Ber(y_d|\mu_{dk}) = \prod_{d=1}^{D} \mu_{dk}^{yk}(1-\mu_{dk})^{1-y_d}$$

大纲

陈述

- 概率图形模型或 PGM 是一种联合概率分布，它使用图形结构来编码条件独立性假设
- 当图是有向无环图 (DAG) 的时候，模型又被叫做贝叶斯网络
- 每个节点代表了一个随机变量，每条边代表一个有向的依赖
- 我们可以将节点以拓扑序编码，并且每个节点在给定父亲节点的情况下同期所有的祖先节点条件独立

$$\mathbf{Y}_i \perp \mathbf{Y}_{pred(i)\setminus pa(i)} | \mathbf{Y}_{pa(i)}$$

公式

联合分布公式

$$p(\mathbf{Y}_{1:N_G}) = \prod_{i=1}^{N_G} p(Y_i|\mathbf{Y}_{pa(i)})$$

洒水网络

我们想要给四个随机变量之间的依赖关系进行建模处理：

(1) C(是否是多云季节)

(2) R(是否下雨)

(3) S(是否喷水系统开启)

(4) W(是否草是湿的)

根据先验知识，我们可以如下定义联合分布：

$$p(C,S,R,W)=p(C)p(S|C)p(R|C,\cancel{S})p(W|\cancel{C},S,R)$$

洒水网络

定义

- 节点 i 的条件概率分布 (CPD)：$p(Y_i|\mathbf{Y}_{pa(i)})$
- 如果每个 CPD 都为条件范畴分布，我们可以将其写成条件概率表 (CPT)
  第 i 个 CPT 可以表示为：

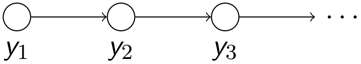
$$\theta_{ijk} \triangleq p(Y_i = k|\mathbf{Y}_{pa(i)=j})$$

$$\text{s.t.} \quad 0 \leq \theta_{ijk} \leq 1 \quad \text{and} \quad \sum_{k=1}^{K_i} \theta_{ijk} = 1$$

马尔可夫链

定义

- 语言模型：假设我们想在可变长度序列上创建一个联合概率分布 $p(y_{1:T})$。如果每个变量 y_t 代表一个具有 K 可能值的词汇表中的单词，则 $y_t \in \{1, \ldots, K\}$ ，则生成的模型表示长度为 T 的可能句子的分布；这通常被称为语言模型。
- 一阶马尔可夫条件：未来 $y_{t+1:T}$ 在给定现在 y_t 的情况下与过去 $y_{1:t-1}$ 无关



马尔可夫链

根据概率的链式计算法则，可以将其联合分布写成下述形式：

$$p(\mathbf{y}_{1:T}) = p(y_1)p(y_2|p_1)p(y_3|y_2,y_1)p(y_4|y_3,y_2,y_1)\cdots = \prod_{t=1}^{T} p(y_t|\mathbf{y}_{1:T})$$

根据一阶马尔可夫条件, 我们可以重写联合分布为：

$$p(\mathbf{y}_{1:T}) = p(y_1)p(y_2|p_1)p(y_3|y_2)p(y_4|y_3)\cdots = p(y_1)\prod_{t=2}^{T} p(y_t|y_{t-1})$$

这也被称为**马尔可夫链、马尔可夫模型或 1 阶自回归模型**

马尔可夫链

- 其中 $p(y_t|y_{t-1})$ 被叫做过渡函数、过渡核或者马尔可夫核。其满足 $p(y_t|y_{t-1}) \geq 0$ 和 $\sum_{k=1}^{K} p(y_t = k|y_{t-1} = j) = 1$
- 可以将 CPT 写成随机矩阵，$A_{jk} = p(y_t = k|y_{t-1} = j)$，其中每列的和为 1，该矩阵也被称为状态转移矩阵。
- 我们假设该矩阵不随时间改变，模型可以称为均匀、静止或时不变的。这个假设允许我们使用固定数量的参数对任意数量的变量进行建模。

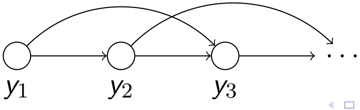
马尔可夫链

定义

M 阶马尔可夫链：每一时刻的状态取决于过去 M 时刻的状态

$$p(\boldsymbol{y}_{1:T}) = p(\boldsymbol{y}_{1:M}) \prod_{t=M+1}^{T} p(y_t|\boldsymbol{y}_{t-M:t-1})$$

当 M=2 的时候，模型被叫做三元模型；当 M=1 时，模型被称为二元模型



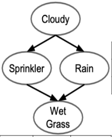
马尔可夫链

对于大词汇量，估计 M 很大的 M-gram 模型的条件分布所需的参数数量可能会变得令人望而却步。在这种情况下，我们需要做出有条件独立性之外的额外假设。
例如，我们可以假设 $p(y_t|y_{t-M:t-1})$ 可以表示为低阶矩阵，也可以用某种神经网络表示。这被称为神经语言模型。

推断

由于一个 PGM 定义了一个联合概率分布。因此，我们可以使用边缘化和条件规则来计算任何变量集 i 和 j 的 $p(\mathbf{Y}_i|\mathbf{y}_j)$。

P (C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

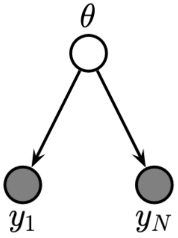
C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S R	P(W=F)	P(W=T)
F F	1.0	0.0
T F	0.1	0.9
F T	0.1	0.9
T T	0.01	0.99

- 下雨的先验信念是 $p(R=1)=0.5$。
- 如果看到草是湿的，那么关于下雨为的后验信念 $p(R=1|W=1)=0.7079$。
- 假设注意到洒水器打开了：我们对下雨的信念下降到 $p(R=1|W=1,S=1)=0.3204$。
- 一些观察的多种原因之间的这种负面相互相互作用被称为伯克森悖论。

学习

如果 CPD 的参数未知，我们可以将它们视为额外的随机变量，将它们作为节点添加到图表中，然后将它们视为要推断的隐藏变量。



该模型属性有以下性质：

- $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$
- $\mathbf{y}_n \sim p(\mathbf{y}|\boldsymbol{\theta})$

相应的联合分布函数有以下形式：

$$p(\mathcal{D}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})(\mathcal{D} = \{\mathbf{y_1}, \ldots, \mathbf{y_n}\})$$

根据独立同分布假设，似然可以被写成 $p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{y}_n|\boldsymbol{\theta})$

板符号

定义

板符号：为了避免视觉混乱，通常使用一种称为板的句法糖。这是一个符号约定，我们在重复变量周围画一个小框，当模型展开时，框中的节点将重复。