1 概率图模型

1.1 有向图: 贝叶斯网络

定义 1.1 贝叶斯网络 贝叶斯网络 (Bayesian network), 又称**信念网络 (Belief Network)**, 是一种概率图模型, 其网络拓朴结构是一个有向无环图, 边的终点所指的事件依赖于边的起点所指的事件.

1.1.1 贝叶斯网络的三种情形

讨论贝叶斯网络中事件的条件独立性时, 总会遇到以下三种情形:

1. **尾对尾 (tail-to-tail)**: 两根箭头的尾部相连:

$$a \leftarrow c \rightarrow b$$

可以简单地理解为 a, b 有共同的起因 c(尽管并不一定是因果关系). 根据贝叶斯网络的性质, 此时的联合概率为

$$p(a,b,c) = p(a|c)p(b|c)p(c)$$

于是

$$p(a,b) = \sum_c p(a|c)p(b|c)p(c)$$

一般而言, 上式并不等于

$$p(a)p(b) = \left(\sum_c p(a|c)p(c)\right)(p(b|c)p(c))$$

因此 a 与 b 并不独立, 记作 $a \not\perp\!\!\!\perp b \mid \varnothing$. 这里 $\varnothing$ 表示空集, 即没有前提条件.

当固定 (已知)c 时则有

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = p(a|c)p(b|c)$$

即 a 与 b 在 c 固定的条件下是独立的:

$$a \perp\!\!\!\perp b \mid c$$

可以举一个简单的例子: 假定 a, b, c 分别是一所小学中学生的鞋子尺码, 阅读能力和年龄. 当 c 不固定时, 将会得出阅读能力和鞋子尺码正相关的结论, 但是当固定年龄 c 时就会发现这两者是独立的.

2. **头对尾 (head-to-tail)**: 两根箭头头尾相连:

$$a \rightarrow c \rightarrow b$$

此时 a, b, c 的联合概率为

$$p(a,b,c) = p(a)p(c|a)p(b|c)$$

如果 c 不固定, 则

$$p(a,b) = p(a)\sum_c p(c|a)p(b|c)$$

一般而言, 上式并不等于

$$p(a)p(b) = p(a)\sum_a \left(p(a) \sum_c p(c|a)p(b|c) \right)$$

因此 $a \not\perp\!\!\!\perp b \mid \varnothing$. 当固定 c 时

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = \frac{p(a,c)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

即 $a \perp\!\!\!\perp b \mid c$. 也即, 固定 c 之后切断了 a 对 b 的影响.

3. 头对头 (head-to-head): 两根箭头的头部相连:

$$a \to c \leftarrow b$$

此时 a, b, c 的联合概率为

$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

如果 c 不固定, 则有

$$p(a,b) = p(a)p(b)\sum_c p(c|a,b) = p(a)p(b)$$

即 $a \perp\!\!\!\perp b \mid \varnothing$. 然而, 如果 c 固定, 则有

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(b)p(c|a,b)}{\displaystyle\sum_{a,b} p(a)p(b)p(c|a,b)}$$

而

$$p(a|c)p(b|c) = \frac{p(a)p(b)\left(\displaystyle\sum_a p(a)p(c|a,b)\right)\left(\displaystyle\sum_b p(b)p(c|a,b)\right)}{\left(\displaystyle\sum_{a,b} p(a)p(b)p(c|a,b)\right)^2}$$

一般情况下两者并不相等, 即 $a \not\perp\!\!\!\perp b \mid c$. 这表明在两个事件共同的结果确定时, 这两者具有一定的相关关系.

可以举一个简单的例子: 假定 a, b, c 分别是文化课成绩好, 体育成绩好和进入某大学. 如果文化课成绩好和体育成绩好都能进入大学, 那么观察大学里的学生就很可能发现文化课成绩好的体育成绩差, 反之亦然. 实际上两者并没有负相关关系, 但是固定 c 之后就导致了表面上的负相关.

1.2 朴素贝叶斯

1.2.1 朴素贝叶斯分类器

贝叶斯分类器是一种用于分类任务的特殊的贝叶斯网络。它假设在类别 $\mathcal{C}$ 固定时, 输入 $\boldsymbol{x}$ 的各个分量 (特征) 的分布之间是独立的. 朴素指的就是这种算法假定各个分量的分布独立. 尽管真实数据可能不满足这个简化假设, 但这一方法对很多复杂的问题还是很有效果的.

在朴素贝叶斯模型中, 类别 $\mathcal{C}$ 是父节点, 输入 $\boldsymbol{x}$ 的各个分量 $x_1, \cdots, x_D$ 是其子节点, 子节点之间没有连线. 于是联合概率分布为

$$p(\boldsymbol{x}|\mathcal{C}) = p(\mathcal{C}) \prod_{i=1}^{D} p_i(x_i|\mathcal{C})$$

于是就把分布处理成多个一元分布, 然后根据数据对每个维度进行训练. 这很好地避免了维度灾难的问题.

朴素贝叶斯的推断可以利用贝叶斯公式, 即

$$p(\mathcal{C}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{C})p(\mathcal{C})}{p(\boldsymbol{x})} = \frac{p(\mathcal{C})}{p(\boldsymbol{x})} \prod_{i=1}^{D} p_i(x_i|\mathcal{C})$$

1.3 马尔科夫随机场与玻尔兹曼机

1.3.1 马尔科夫随机场

定义 1.2 马尔科夫随机场 **马尔科夫随机场 (Markov random field)**, 又称**马尔科夫网络 (Markov network)**, 也是一类概率图模型, 属于无向图模型 (undirected graphical model), 连接结点的边是没有方向的.

1.3.2 玻尔兹曼机与受限玻尔兹曼机