

计算机科学188：人工智能导论

2024年春季

笔记10

作者（其他所有笔记）：尼基尔·夏尔马

作者（贝叶斯网络笔记）：乔希·胡格和杰基·梁，由王瑞吉编辑

作者（逻辑笔记）：亨利·朱，由考佩林编辑

致谢（机器学习与逻辑笔记）：部分内容改编自教材《人工智能：一种现代方法》。

最后更新时间：2023年8月26日

概率概述

我们假设你已经在CS70课程中学习了概率基础，因此这些笔记将假定你对概率中的标准概念有基本理解，如概率密度函数、条件概率、独立性和条件独立性。在此，我们简要总结一下我们将使用的概率规则。

随机变量表示结果未知的事件。概率分布是对结果赋予权重。概率分布必须满足以下条件：

$$0 \leq P(\omega) \leq 1$$
$$\sum_{\omega} P(\omega) = 1$$

例如，如果 A 是一个二元变量（只能取两个值），那么对于某个 $p \in [0, 1]$ 有 $P(A=0) = 1-p$ 和 $P(A=1) = p$ 。

我们将采用这样的约定：大写字母表示随机变量，小写字母表示该随机变量的某个特定结果。

我们使用符号 $P(A, B, C)$ 来表示变量 A, B, C 的联合分布。在联合分布中，顺序无关紧要，即 $P(A, B, C) = P(C, B, A)$ 。

我们可以使用链式法则（有时也称为乘积法则）来展开联合分布。

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$
$$P(A_1, A_2, \dots, A_k) = P(A_1)P(A_2|A_1) \dots P(A_k|A_1 \dots A_{k-1})$$

通过对变量 C 作为 $P(A, B) = \sum_c P(A, B, C=c)$ 时所有可能取值进行求和，可以得到 A, B 的边际分布。 A 的边际分布也可以通过 $P(A) = \sum_b \sum_c P(A, B=b, C=c)$ 得到。我们有时也会将边缘化过程称为“求和消元”。

当我们对概率分布进行操作时，有时会得到并非必然总和为1的分布。为了解决这个问题，我们进行归一化：求出分布中所有项的总和，然后将每个项除以该总和。

条件概率是基于某些已知事实为事件赋予概率。例如

$P(A \mid B = b)$ 给出了在已知 B 的值等于 b 的情况下 A 的概率分布。条件概率的定义如下：

$$P(A|B) = \frac{P(A,B)}{P(B)}.$$

结合上述条件概率的定义和链式法则，我们得到贝叶斯法则：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

为了表示随机变量 A 和 B 相互独立，我们写成 $A \perp B$ 。这等同于 $B \perp A$ 。

当 A 和 B 相互独立时， $P(A, B) = P(A)P(B)$ 。你可以想到的一个例子是两次独立的抛硬币。在其他课程中，你可能已经熟悉将相互独立简称为“独立”。我们可以从上述等式和链式法则推导出 $P(A \mid B) = P(A)$ 和 $P(B \mid A) = P(B)$ 。

要写出随机变量 A 和 B 在给定另一个随机变量 C 的条件下是条件独立的，我们写成 $A \perp B \mid C$ 。这也等同于 $B \perp A \mid C$ 。

如果给定 C 时， A 和 B 条件独立，那么 $P(A, B \mid C) = P(A \mid C)P(B \mid C)$ 。这意味着如果我们知道 C 的值，那么 B 和 A 不会相互影响。与上述条件独立定义等价的关系是 $P(A \mid B, C) = P(A \mid C)$ 和 $P(B \mid A, C) = P(B \mid C)$ 。注意这三个等式如何与相互独立的三个等式等价，只是多了一个关于 C 的条件！

概率推理

在人工智能领域，我们常常希望对各种不确定事件之间的关系进行建模。如果天气预报显示有 40% 的降雨概率，我应该带伞吗？如果我买的冰淇淋球越多，掉落的可能性就越大，那我应该买几个球呢？如果在我前往甲骨文球馆观看勇士队比赛的途中，高速公路 15 分钟前发生了一起事故，我应该现在出发还是 30 分钟后出发？所有这些问题（以及更多问题）都可以通过概率推理来回答。

在本课程前面的章节中，我们将世界建模为处于一个始终可知的特定状态。在接下来的几周里，我们将使用一种新模型，其中世界的每个可能状态都有其自身的概率。例如，我们可能构建一个天气模型，其状态由季节、温度和天气组成。我们的模型可能会说 $P(\text{winter}, 35^\circ, \text{cloudy}) = 0.023$ 。这个数字代表冬天、 35° 且多云这种特定结果的概率。

更确切地说，我们的模型是一个联合分布，即一个概率表，它捕捉了每个可能结果的可能性，也称为变量赋值。例如，考虑下面的表格：

Season	Temperature	Weather	Probability
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

这个模型使我们能够回答一些我们可能感兴趣的问题，例如：

- 晴天的概率是多少？ $P(W = \text{sun})$
- 已知是冬天，天气的概率分布是怎样的？ $P(W \mid S = \text{winter})$
- 已知是雨天且寒冷，是冬天的概率是多少？ $P(S = \text{冬天} \mid T = \text{寒冷}, W = \text{下雨})$
- 已知寒冷，天气和季节的概率分布是怎样的？ $P(S, W \mid T = \text{cold})$

枚举推理

给定一个联合概率密度函数 (joint PDF)，我们可以使用一种称为枚举推理的简单直观过程轻松计算任何所需的概率分布 $P(Q_1 \dots Q_m \mid e_1 \dots e_n)$ ，为此我们定义三种我们将处理的变量：

1. 查询变量 Q_i ，它们是未知的，并且出现在所需概率分布中条件 (|) 的左侧。
2. 证据变量 e_i ，它们是已观察到的变量，其值已知，并且出现在所需概率分布中条件 (|) 的右侧。
3. 隐藏变量，它们是存在于整体联合分布中但不存在于所需分布中的值。

在枚举推理中，我们遵循以下算法：

1. 收集与观察到的证据变量一致的所有行。
2. 对所有隐藏变量进行求和消元（边缘化）。
3. 对表格进行归一化处理，使其成为一个概率分布（即所有值的总和为1）

例如，如果我们想使用上述联合分布来计算 $P(W \mid S = \text{winter})$ ，我们会选择 S 为冬季的那四行，然后对 T 进行求和消元并归一化。这将得到以下概率表：

W	S	Unnormalized Sum	Probability
sun	winter	$0.10 + 0.15 = 0.25$	$0.25 / (0.25 + 0.25) = 0.5$
rain	winter	$0.05 + 0.20 = 0.25$	$0.25 / (0.25 + 0.25) = 0.5$

因此 $P(W = \text{sun} \mid S = \text{winter}) = 0.5$ 和 $P(W = \text{rain} \mid S = \text{winter}) = 0.5$ ，我们了解到在冬季有50%的概率出太阳，有 50% 的概率下雨。

只要有联合概率密度函数表，通过枚举推理（IBE）就可以用于计算任何所需的概率分布，即使是针对多个查询变量 $Q_1 \dots Q_m$ 。