概率建模与推断
引言

Michael Gutmann（迈克尔·古特曼）
译者：郭尚敏（Shangmin Guo）

School of Informatics, University of Edinburgh

版权声明

概率推断

- 概率推断 $\equiv$ 概率推理：
  基于能观测的事件计算不能或未曾观测的事件的概率
  - 未观察/不确定事件，例：认知功能障碍 $x = 1$
  - 观察事件 $\equiv$ 证据 $\equiv$ 数据，例：检测结果 $y = 1$
- “先验”：观测数据前，对于不确定事件的概率分布假设，例：$Pr(x = 1)$
- “后验”：观测数据前，对于不确定事件的概率分布估计，例：$Pr(x = 1|y = 1)$
- 后验可利用贝叶斯法则基于先验和证据计算得出

概率的关键运算法则

1. 乘积法则:

$$\begin{aligned} Pr(x=1, y=1) &= Pr(y=1|x=1)Pr(x=1) \\ &= Pr(x=1|y=1)Pr(y=1) \end{aligned}$$

2. 求和法则:

$$Pr(y=1) = Pr(x=1, y=1) + Pr(x=0, y=1)$$

贝叶斯法则可由乘积法则导出:

$$Pr(x=1|y=1) = \frac{Pr(x=1, y=1)}{Pr(y=1)} = \frac{Pr(y=1|x=1)Pr(x=1)}{Pr(y=1)}$$

其中，分母可根据求和与乘积法则计算:

$$Pr(y=1) = Pr(y=1|x=1)Pr(x=1) + Pr(y=1|x=0)Pr(x=0)$$

概率的关键运算法则

- 以上法则可以推广到多元随机变量（离散或连续）
- 假设存在变量 $\mathbf{x}$ 和 $\mathbf{y}$，其条件联合概率密度函数（pdf）或概率质量函数（pmf）为 $p(\mathbf{x}, \mathbf{y})$，以上法则可推广为：

1. 乘积规则：

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \\ &= p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \end{aligned}$$

2. 求和规则：

$$p(\mathbf{y}) = \begin{cases} \sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}) & \text{for discrete r.v. (随机变量)} \\ \int p(\mathbf{x}, \mathbf{y}) d\mathbf{x} & \text{for continuous r.v.} \end{cases}$$

概率建模与推断

- 概率建模：
  - 首先确定需要建模的事实，以及相关的数量因素
  - 假设它们都是随机变量，比如 $\mathbf{x}$、$\mathbf{y}$ 和 $\mathbf{z}$，并服从联合 pdf (pmf) $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$
- 概率推断：
  - 假设已知 $\mathbf{y} \in \varepsilon$(测量结果、证据)
  - 关于 $\mathbf{x}$ 的概率推断是指计算

$$p(\mathbf{x}|\mathbf{y} \in \varepsilon)$$

    或相关量，比如
    - $\operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y} \in \varepsilon)$，最大值
    - $\mathbb{E}\left[g(\mathbf{x})|\mathbf{y} \in \varepsilon\right] = \int g(\mathbf{u})p(\mathbf{u}|\mathbf{y} \in \varepsilon)d\mathbf{u}$，期望值

通过乘积、求和法则求解

假设所有的变量都是离散量，即 $\varepsilon = \{\mathbf{y}_o\}$，并且已知 $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$，可通过以下方式求解 $p(\mathbf{x}|\mathbf{y}_o)$：

1. 乘积法则：$p(\mathbf{x}|\mathbf{y}_o) = \frac{p(\mathbf{x},\mathbf{y}_o)}{p(\mathbf{y}_o)}$
2. 求和法则：$p(\mathbf{x}, \mathbf{y}_o) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})$
3. 求和法则：$p(\mathbf{y}_o) = \sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}_o) = \sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})$
4. 结果：

$$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}$$

概率建模与推断的主要问题

$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}$

假设 $\mathbf{x}, \mathbf{y}, \mathbf{z}$ 均为 $d = 500$ 维向量且向量中每个元素可取 $K = 10$ 个离散值，那么

- 问题 1：为了指定 $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$，我们需要指定 $K^{3d} - 1 = 10^{1500} - 1$ 个非负数，而这样的计算量明显是不可能的。
  主题 1——表示：为了高效地表示 $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$，我们可以做哪些合理的弱假设?

概率建模与推断的主要问题

$$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}$$

- 问题 2：上式中求和的操作存在组合爆炸问题，即分子求和的计算量为 $K^d = 10^{500}$、分母求和计算量为 $K^{2d} = 10^{1000}$，这样的计算量太大、无法完成。
  主题 2——精确推断：我们应该进一步利用 $p(\mathbf{x},\mathbf{y},\mathbf{z})$ 所做的那些假设来高效地计算后验概率或者其他想计算的值?
- 问题 3：$p(\mathbf{x},\mathbf{y},\mathbf{z})$ 中的非负数应该如何设置才合理?
  主题 3——学习：我们应该如何从数据中学习这些数字?
- 问题 4：即使在重复利用了所做的假设，对于有些模型来说，精确推断还是开销太高。
  主题 4——近似推断与学习。