BRML第一章：概率推断

emacsun

目录

最近我在阅读 David Barber 的 Bayesian Reasoning and Machine Learning（以后都用BRML来简称）。阅读过程中，随手记录了一些笔记。现在稍作整理，发表为一系列博文。本文是第一篇，也就是第一章的笔记整理。后续的博文都是各个章节的笔记整理，基本是一章一篇博文，个别重要的章节或许会作为多篇博文来写。

BRML的第一章主要复习了概率的基本知识，通过贝叶斯公式（或者条件概率公式）引入概率推断的概念，然后对先验概率，后验概率和似然值做了详细介绍。从第一章就可以窥测本文将会用大量的例子来阐述概念，这是我比较喜欢的书籍的风格。

1 贝叶斯法则

为了保证完整性，记录贝叶斯公式的定义。

定义 *1.1* 已知事件 y时，事件x发生的概率，定义为：

$$p(x|y) \triangleq \frac{p(x,y)}{p(y)} = \frac{p(x)p(y|x)}{p(y)}$$

这个公式是如此重要，基本上撑起了机器学习的半壁江山。另外，这个公式在通信系统的信道译码算法中也频频出现，尤其在消息传递算法或者置信度传递算法中。关于其重要性，就不再多言。随着学习的深入，对这个公式及其扩展的理解会愈加深刻。

从贝叶斯规则引入统计独立的概念，即当

$$p(x,y) = p(x)p(y)$$

x, y是相互独立的。此时x的发生不影响y发生的概率，即

$$p(y|x) = p(y)$$

2 概率推断

概率推断的核心是识别环境中所有相关的随机变量$x_1, \ldots, x_N$，并且根据他们的关系创建一个概率模型$p(x_1, \ldots, x_N)$。推断的过程是根据已知信息更新某个随机变量概率的过程。

例 *2.1* 医生发现一个人得*KJ*病的概率是相当低的大约是$1/100000$，但是得了*Kreuzfeld-Jacob(KJ)*病的人几乎都吃汉堡包，$p(HamburgerEater|KJ) = 0.9$。

1. 假设一个人吃汉堡包的概率是*0.5*，$p(HamburtgerEater) = 0.5$，那么一个吃汉堡包的人得*KJ*的概率是多少？

   这个概率可以表示为：

$$\begin{aligned} p(KJ|HamburtgerEater) &= \frac{p(KJ, HamburtgerEater)}{p(HamburtgerEater)} && (2.1) \\ &= \frac{p(HamburtgerEater|KJ)p(KJ)}{p(HamburtgerEater)} && (2.2) \\ &= \frac{0.9 \times 1/100000}{1/2} && (2.3) \\ &= 1.8 \times 10^{-5} && (2.4) \end{aligned}$$

2. 如果吃汉堡包的人的概率比较低，不是*0.5*而是*0.001*，那么一个吃汉堡包的人得*KJ*的概率是多少？

   重复上面的计算

$$\begin{aligned} p(KJ|HamburtgerEater) &= \frac{p(KJ, HamburtgerEater)}{p(HamburtgerEater)} && (2.5) \\ &= \frac{p(HamburtgerEater|KJ)p(KJ)}{p(HamburtgerEater)} && (2.6) \\ &= \frac{0.9 \times 1/100000}{0.001} && (2.7) \\ &\approx 1/100 && (2.8) \end{aligned}$$

这个例子告诉我们不要为一些不大可能的事情担心:得了KJ的前提下吃汉堡包的概率和吃汉堡包的前提下得KJ的概率完全是两码事。

例 *2.2* 再给一个异或门的例子，我们知道一个标准的异或门电路的逻辑关系满足表*1:*

表 1: 异或门逻辑

A	B	$C = A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	1

当我们观测到异或门的输出是0时，对于A或者B的概率我们知道多少？在这样的情况下，可能A, B都是0，也可能A, B都是1。 A, B处于0或者1的概率是等该的。

但是考虑一个软判决输出的异或门逻辑,其逻辑关系如表2，我们假定A, B是独立的且$p(A = 1) = 0.65, p(B = 1) = 0.77$，那么求$p(A = 1|C = 0)$？

表 2: 异或门逻辑

A	B	$p(C = 1\|A, B)$
0	0	0.1
0	1	0.99
1	0	0.8
1	1	0.25

由条件概率公式，得：

$$p(A = 1|C = 0) = \frac{p(A = 1, C = 0)}{p(C = 0)} = \frac{p(A = 1, C = 0)}{p(A = 1, C = 0) + p(A = 0, C = 0)} \tag{2.9}$$

接下来，我们对2.9右边分母上的两个求和项进行展开：

$$\begin{aligned} p(A = 1, C = 0) &= \sum_B p(A = 1, B, C = 0) & (2.10) \\ &= \sum_B p(C = 0|B, A = 1)p(B)p(A = 1) & (2.11) \end{aligned}$$

$$\begin{aligned} p(A = 0, C = 0) &= \sum_B p(A = 0, B, C = 0) & (2.12) \\ &= \sum_B p(C = 0|B, A = 0)p(B)p(A = 0) & (2.13) \end{aligned}$$

带入表格中相应的概率数字得出$p(A = 1|C = 0) = 0.8436$

3 先验概率，似然值和后验概率

现实生活中非常多的问题可以归类为：当我知道数据D时，告诉我随机变量θ的概率。这个问题可以归类为：

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int_\theta p(D|\theta)p(\theta)} \tag{3.1}$$

这个模型在机器学习和信道编码理论中都有普遍的实用，甚至是其基础的基础。那么从式3.1我们可以读出什么信息呢？式3.1告诉我们，我们可以从数据生成模型$p(D|\theta)$和先验概率$p(\theta)$推断后验概率$p(\theta|D)$。最大后验概率准则(maximize a posteriori,MAP)准则，可以表示为：

$$\hat{\theta} = \arg\max_\theta p(\theta|D) \tag{3.2}$$

对于等概率分布的先验概率$p(\theta)$，MAP准则和最大似然准则(maximum likelihood,ML) 是等效的，即最大化$p(D|\theta)$的θ同样最大化$p(\theta|D)$。

例 *3.1*　现在针对$p(D|\theta)$我们给出一个例子。假设一个钟摆在摆动，我们用x_t来表示钟摆在t时刻的角度。假设每次测量都是独立的。假设每次测量都是精确的，则有

$$x_t = \sin(\theta t) \tag{3.3}$$

这里假设系统没有阻尼则$\theta = \sqrt{g/L}$，其中g是地球引力场数，L是吊起钟摆的绳子的长度，但是我们是在测试θ，是根据$x_1, \ldots, x_T$来测量θ。另外，实际测量过程中(比如测量位置仪器质量很差或者定时器不准)，测量总是存在误差，假设误差是ϵ_t，测量结果可以表示为:

$$x_t = \sin(\theta t) + \theta_t \tag{3.4}$$

一般情况我们定义ϵ_t服从均值为零方差为δ^2的高斯随机变量。所以关于θ的后验概率可以表示为:

$$p(\theta|x_1, \ldots, x_T) \propto p(\theta) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\delta^2}} e^{\frac{1}{2\delta^2}(x_t - \sin(\theta)t)^2} \tag{3.5}$$

例 *3.2*　我们来考虑投掷两个均匀骰子的场景。假设现在有人告诉你两个骰子的数字之和为9。求此时关于两个骰子上数字的后验概率分布。

首先我们用s_a, s_b代表两个骰子的数字，其取值范围是$\{1,2,3,4,5,6\}$。两者之和为$t = s_a + s_b$这三个随机变量的模型遵循:

$$p(t, s_a, s_b) = \underbrace{p(t|s_a, s_b)}_{likelihood} \underbrace{p(s_a, s_b)}_{prior} \tag{3.6}$$

假设两个骰子是均匀的$p(s_a, s_b)$则$p(s_a, s_b) = p(s_a)p(s_b)$其概率分布表格为:

表 3:　$p(s_a)p(s_b)$

	$s_a=1$	$s_a=2$	$s_a=3$	$s_a=4$	$s_a=5$	$s_a=6$
$s_a=1$	1/36	1/36	1/36	1/36	1/36	1/36
$s_a=2$	1/36	1/36	1/36	1/36	1/36	1/36
$s_a=3$	1/36	1/36	1/36	1/36	1/36	1/36
$s_a=4$	1/36	1/36	1/36	1/36	1/36	1/36
$s_a=5$	1/36	1/36	1/36	1/36	1/36	1/36
$s_a=6$	1/36	1/36	1/36	1/36	1/36	1/36

由于骰子是均匀的则$p(s_a) = p(s_b) = 1/6$。另外，我们有$p(t|s_a, s_b)$，表格如下:

表 4:　$p(t|s_a, s_b)$

	$s_a=1$	$s_a=2$	$s_a=3$	$s_a=4$	$s_a=5$	$s_a=6$
$s_a=1$	0	0	0	0	0	0
$s_a=2$	0	0	0	0	0	0
$s_a=3$	0	0	0	0	0	1
$s_a=4$	0	0	0	0	1	0
$s_a=5$	0	0	0	1	0	0
$s_a=6$	0	0	1	0	0	0

后验概率 $p(s_a, s_b|t=9) = \frac{p(t=9|s_a,s_b)p(s_a)p(s_b)}{p(t=9)}$，其中

$$p(t=9) = \sum_{s_a s_b} p(t=9|s_a, s_b)p(s_a)p(s_b)$$

综上我们可以得到后验概率表:

表 5: $p(s_a, s_b|t=9)$

	$s_a=1$	$s_a=2$	$s_a=3$	$s_a=4$	$s_a=5$	$s_a=6$
$s_a=1$	0	0	0	0	0	0
$s_a=2$	0	0	0	0	0	0
$s_a=3$	0	0	0	0	0	1/4
$s_a=4$	0	0	0	0	1/4	0
$s_a=5$	0	0	0	1/4	0	0
$s_a=6$	0	0	1/4	0	0	0