INT104W4_朴素贝叶斯v0.1

假设有两个事件：A事件 与 B事件

我们记发生事件A的概率为P(A)，发生事件B的概率为P(B)

乘法公式：P(AB)=P(A)×P(B|A)=P(B)×P(A|B)；
$P(A_1A_2A_3...A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1A_2...A_n)$

如果事件间相互独立：事件与事件间没有因果或顺序关系，或者说一个观测值的取值不会对其他观测值产生影响

独立事件有性质：AB两事件同时发生的概率：

$$P(AB) = P(A)\cdot P(B)$$

非独立事件有性质：在发生A事件的情况下B事件发生的概率：

$$\text{条件概率公式：}P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(B)\cdot P(A|B)}{P(A)}$$

*其中P(A∩B)是A事件和B事件同时发生的概率

在频率主义学派中，他们认为参数虽然未知，但存在客观的固定值，通过优化似然函数可以确定参数。

在贝叶斯学派中，一个条件发生的概率的分布θ（你可以认为是一个参数，代表着**随机变量的可能取值及取得对应值的概率**）是不确定的，而是服从一个函数。故我们可以假定一个参数服从一个先验分布，之后再用观测到的数据来计算参数的后验分布。

有一个人前来买瓜，如果在瓜摊的所有瓜中，

- 出现好瓜的概率是P(A)=0.3
- 出现声音清脆的瓜的概率是$P(B_1)$=0.4
- 在清脆瓜中出现好瓜的概率是P(A|B)=0.2

那么我们怎么**求出好瓜中出现清脆瓜的概率**$P(B_1|A)$呢？（读者不妨画一个韦恩图尝试）

贝叶斯定理

利用贝叶斯定理，我们可以**求出导致一个结果A的原因B_k的概率。其中{B1...Bn}是完备事件组，且P(A)，$P(B_n)$均大于0。**

贝叶斯定理：

$$P(B_k|A)=\frac{P(A|B_k)\cdot P(B_k)}{\sum_1^n P(A_i)P(B|A_i)}=\frac{P(A|B_k)\cdot P(B_k)}{P(A)}\text{（利用全概率公式）}=\frac{P(B_kA)}{P(A)}\text{(利用乘法公式)}$$

*可以将A视为样本属性x，将B_k视为标签c

其中，我们称

- 出现清脆瓜的概率P(B_1)是**先验概率**（Prior Probability）：是在没有训练样本数据前，根据以往经验和主观判断得到的概率，往往是好求的一个概率

  p(θ)->先验：见到数据集D之前，对参数θ的认识

- 在清脆瓜中出现好瓜的概率P(A|B_1)是**类条件概率**（class-conditional probability/CCP）：当下事件由果及因发生的概率。样本属性A相对于类标签B的概率。已知一个条件下，结果发生的概率。条件概率实际上把一个完整的问题集合S通过特征进行了划分，划分成S1/S2/S3…。类条件概率中的类指的是把造成结果的所有原因一 一进行列举，分别讨论。

  p(D|θ)->似然：在给定参数θ下，数据集D被观测到的概率

- 出现好瓜的概率P(A)是**支持因子**（evidence factor /observation）：在贝叶斯推断中用来表示观测数据对不同假设或模型的支持程度的因子。

- 好瓜是清脆瓜的概率P(B_1|A)是**后验概率**（Posterior Probability）：是当下事件由因及果发生的概率，求导致事件B_1发生的原因是由某个因素A引起的可能性的大小。由样本属性x导致标签为c的概率P(c|x)就称为后验概率。得到他，我们就能知道引起这一结果的原因是什么，好比发烧去医院，我们就可以用贝叶斯求出导致你发烧的原因是什么，从而对症下药。

  p(θ|D)->后验：在见到数据集D后，对θ的重新认识

即

$$\text{后验概率}=\frac{CCP\times\text{先验概率}}{\text{支持因子}}$$

由于支持因子在同一个数据集中的概率相同，可发现正比关系：

$$\Rightarrow\text{后验概率}\propto CCP(\text{似然})\times\text{先验概率}$$

那么，要求出后验概率，我们只需要找到CCP和先验概率即可。

根据大数定律，**先验概率P(B_1)**可认为是 观测得到的 发生该现象的概率。

那么只要求出CCP（似然）就可以了，但似然并不好求。

似然函数：P(D|θ) 表示在给定参数 θ 的情况下观测到数据 D 的概率（P(x|C_k) 表示在类别 k 的条件下样本x的特征分布概率，一样的）

为简化操作，我们假设参数集中的参数（特征）彼此独立，就可以将似然函数 P(D|θ) 展开为各参数（特征）的**条件概率的乘积**。

$$有参数集：\Theta=(\theta_1,\theta_2,...,\theta_j),其中\theta_j是一个参数（特征）$$
$$特征独立时，似然：P(D|\Theta)=P(x_1,x_2,...,x_n|\Theta)=\prod_{j=1}^{n}p(x_j|\theta)$$

可计算**好瓜中是清脆瓜的概率**$P(B_1|A)$：

$$P(B_1|A)==\frac{P(A|B_1)\cdot P(B_1)}{P(A)}=\frac{0.2*0.4}{0.3}=0.27$$

例题1-单个参数

气温	外出	不外出	p
热	2	2	0.28(4)
适中	4	2	0.43(6)
冷	3	1	0.28(4)
p	0.64(9)	0.36(5)	1(14)

求气温适中的天气是更有可能外出还是不外出?

p(适中|外出)=4/9=0.44；p(适中|不外出)=2/5=0.4

后验概率：

$$在适中天气外出：p(外出|适中)=\frac{p(适中|外出)\cdot p(外出))}{p(适中)}=0.65 \quad (1)$$
$$在适中天气不外出：p(不外出|适中)=\frac{p(适中|不外出)\cdot p(不外出))}{p(适中)}=0.33 \quad (2)$$

*如果仅需比较可能性大小，无需计算p(适中)

我们发现在适中天气更有可能外出。

例题2-多个参数

Outlook	Temprature	Humidity	Windy	Play
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No
Sunny	Hot	High	True	No
Sunny	Hot	High	False	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

求晴天且有风的情况更有可能外出还是不外出?

后验概率:

$$p(\text{外出}|\text{晴天},\text{有风}) = \frac{p(\text{晴天},\text{有风}|\text{外出})p(\text{外出})}{p(\text{晴天},\text{有风})} \propto p(\text{晴天},\text{有风}|\text{外出})p(\text{外出}) \tag{3}$$

$$\Rightarrow p(\text{晴天}|\text{外出})p(\text{有风}|\text{外出})p(\text{外出}) = 2/9 * 2/9 * 9/14 = 0.05 \tag{4}$$

$$p(\text{不外出}|\text{晴天},\text{有风}) = \frac{p(\text{晴天},\text{有风}|\text{不外出})p(\text{不外出})}{p(\text{晴天},\text{有风})} \propto p(\text{晴天},\text{有风}|\text{不外出})p(\text{不外出}) \tag{5}$$

$$\Rightarrow p(\text{晴天}|\text{不外出})p(\text{有风}|\text{不外出})p(\text{不外出}) = 3/5 * 3/5 * 5/14 = 0.13 \tag{6}$$

我们发现，在晴天且有风的情况下，不外出的概率比外出大，故更有可能不外出。

朴素贝叶斯分类器

选择具有最大后验概率的类别作为样本的预测类别即可。

$$\hat{y} = argmax_k P(C_k|x) \tag{7}$$

$$h_{nb}(x) = argmax_{c \in y} P(c) \prod_{i=1}^{d} P(x_i|c) \tag{8}$$

但如果某个属性值未在训练集中与某个类同时出现过，在估计时该项概率会变为0，导致其他属性的估计被抹去。我们需要进行**平滑**以防止错误的发生，一般使用拉普拉斯修正。

从这个角度看，机器学习要实现的目的就是根据有限的样本集估计出尽可能准确的后验概率。

为此，有两种策略方法：判别式模型和生成式模型。

最大后验估计(MAP)

这是频率主义的想法，最优的参数应该能让后验概率最大。