

浅谈频率学派和贝叶斯学派

原创 石头 机器学习算法那些事 2018-10-17

【写在前面的话】

终于可以写贝叶斯相关的文章啦，心情有点小激动，最近一段时间反复看Bishop老师编写的<<Pattern Recognition and Machine Learning>>前三章章节，发现贝叶斯思想真是太强大了，瞬间成为该书作者的忠实粉丝。在后续的文章中，笔者会尽自己最大的努力去阐述与机器学习算法所蕴含的贝叶斯思想。若有错误之处，欢迎更正指出。笔者认为互相学习、资源互享是最好最快速的学习方法。

【频率学派】

高中数学对概率的定义：在大量重复进行同一实验事件A发生的频率总是接近某一个常数，并在它附近进行摆动，这时将这个常数叫事件A的概率，记作 $P(A)$ 。

这是古典频率学派对概率的定义，定义包含了二个要点：

- (1)、事件A发生的概率是常数。
- (2)、事件A发生的概率是重复多次进行同一实验得到的。

频率学派的局限性：

频率学派评估可重复实验事件发生的概率具有一定的现实意义。

但是假如评估本世纪末北极圈的冰川消失的概率，按照频率学派的思想，首先需要创造无数个平行世界，然后计算北极圈冰川消失的平行世界的频率，记该频率为冰川消失的概率。

目前，创造无数个平行世界的技术还不成熟，因此频率学派在评估不可重复实验事件发生的概率具有很大的限制性。

【贝叶斯学派】

贝叶斯学派对概率的定义：贝叶斯学派评估事件A发生的概率带有主观性，且事件A发生的概率是当前观测数据集D下的概率，即条件概率 $P(A|D)$ ，当观测数据集更新为D1时，则事件A发生的概率为

$P(A|D1)$ ，不同的数据集预测A事件发生的概率不同。贝叶斯学派评估事件A发生的概率会引用先验概率和后验概率两个概念，贝叶斯定理是搭建先验概率和后验概率的桥梁。

定义包含了三个要点：

- (1)、事件A发生的概率是变化的，并非常数。
- (2)、事件A发生的概率是特定数据集下的条件概率。
- (3)、事件A发生的概率是后验概率，且事件A发生的先验概率已给定。

贝叶斯学派的难点在于如何设置合理反映事件A发生的先验概率，不同的先验概率得到的结果不一样。

概率论基本知识回顾

条件概率：设A，B是两个事件，且 $P(A)>0$ ，称

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

为在事件A发生的条件下事件B发生的概率。

乘法定理：设 $P(A)>0$ ，称

$$P(AB) = P(B|A)P(A)$$

事件A，B同时发生的概率等于事件A发生的概率与事件A发生的条件下事件B发生的概率的乘积。

求和定理：设 $P(A)>0$ ，称

$$P(A) = \sum_B P(A, B)$$

为事件A发生概率的边缘化。

全概率公式：事件B发生的所有可能结果 B_1, B_2, \dots, B_n ，事件A发生的概率 $P(A)$ ，则

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n) \\ \Rightarrow P(A) &= \sum_B P(A)P(A|B) \end{aligned}$$

贝叶斯定理：

$$\begin{aligned} P(A, B) &= P(B, A) \\ \Rightarrow P(A|B)P(B) &= P(B|A)P(A) \\ \Rightarrow P(A|B) &= \frac{P(A)P(B|A)}{P(B)} \end{aligned}$$

其中， $P(A|B)$ 为已知事件B下A发生的概率，称为**后验概率**；等式右边分子部分 $P(A)$ 为事件A发生的概率，称为**先验概率**。**贝叶斯定理是先验概率和后验概率转换的桥梁。**

频率学派和贝叶斯学派在评估模型参数的异同

相同点：最大似然函数在频率学派和贝叶斯学派都具有重要的作用，最大似然函数的思想是存在即合理，认为已观测数据的概率分布是最大概率，最大概率对应的模型就是我们需要找的模型。

不同点：频率学派认为模型是一成不变的，即模型参数是个常数；贝叶斯学派认为模型是一直在变的，当获取新的信息后，模型也相应的在改变，即模型参数是个变量，用概率去描述模型参数的不确定性。

【例】小明在做抛硬币试验，已观测数据集D为五次正面向上，求正面向上的概率 w

频率学派解法：

$$\begin{aligned}\text{似然函数: } P(D|w) &= (w)^5; \\ \text{最大似然函数: } P(D|w)_{\max} &= 1; \\ \Rightarrow (w)^5 &= 1 \\ \Rightarrow w &= 1\end{aligned}$$

硬币正面向上的概率为1，模型明显存在问题，称为**过拟合**

贝叶斯学派解法：

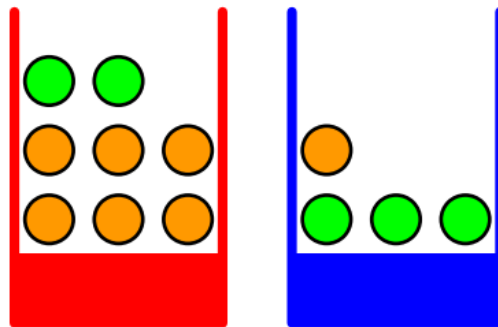
假设硬币正面向上的先验概率 $p(w)$ ，根据贝叶斯定理得：

$$\begin{aligned}P(w|D) &= \frac{P(D|w) * P(w)}{P(D)} \\ \Rightarrow \text{最大化后验概率 } P(w|D) \\ \Rightarrow w &= P(w|D)_{\max} \\ w &\text{即是正面向上的概率}\end{aligned}$$

由于篇幅的关系，本文只在这里描述了贝叶斯学派求解硬币正面向上概率的流程图，笔者会在后面的文章详细描述这一现象。

贝叶斯定理在先验概率和后验概率的应用举例

【例】 一个红盒子有六个橘子一个苹果，一个蓝色盒子有一个橘子三个苹果，选择红盒子的概率为0.4，选择蓝盒子的概率为0.6，随机从盒子抽取一次水果，（1）求水果为橘子的概率；（2）当抽取的水果为橘子时，求随机选择盒子为红色的概率；



解：

假设选择盒子的事件记为 B ， B 有两种可能的结果，选择红盒子记为 r ，选择蓝盒子记为 b ；

假设抽取水果的事件记为 F ， F 有两种可能的结果，抽取橘子记为 o ，抽取苹果记为 a ；

(1)、由全概率公式得：

$$\begin{aligned}
 P(F=o) &= P(B=r, F=o) + P(B=b, F=o) \\
 &\Rightarrow P(F=o) = P(B=r)P(F=o|B=r) + P(B=b)P(F=o|B=b) \\
 &\Rightarrow P(F=o) = 0.4 * (3/4) + 0.6 * (1/4) \\
 &\Rightarrow P(F=o) = 0.45
 \end{aligned}$$

因此，随机抽取水果为橘子的该为0.45。

(2)、问题转化为求解 $P(B=r|F=o)$

由贝叶斯定理得：

$$\begin{aligned}
 P(B=r|F=o) &= \frac{P(B=r, F=o)}{P(F=o)} \\
 P(B=r|F=o) &= \frac{P(B=r)P(F=o|B=r)}{P(F=o)} \\
 \text{由 (1) 可知 } P(F=o) &= 0.45, \text{ 得:} \\
 P(B=r|F=o) &= \frac{0.4 * (3/4)}{0.45} \\
 P(B=r|F=o) &= 0.67
 \end{aligned}$$

由 (2) 可知，选择红色盒子概率为0.4，该概率为先验概率；当观测数据为橘子时，选择红色盒子的概率变成0.67，该概率为后验概率。再次证明了贝叶斯估计模型的概率是随着观测数据的变化而变化的。

总结

本文介绍了频率学派和贝叶斯学派的概率定义，频率学派认为模型是一成不变的，贝叶斯学派认为模型是随着数据的更新而不断更新，频率学派和贝叶斯学派都可以使用最大似然函数来估计模型。



-END-



长按二维码关注

机器学习算法那些事
微信: beautifulife244

砥砺前行 不忘初心