机器学习数据分析和建模中级教程系列讲座-第5讲

**朴素贝叶斯(Naïve Bayes)(上)**

主讲人：元俊

**预备知识**

基本算法基础（开平方）

概率论中级知识（伯努利分布）

逻辑回归的回顾（极大似然，随机梯度）

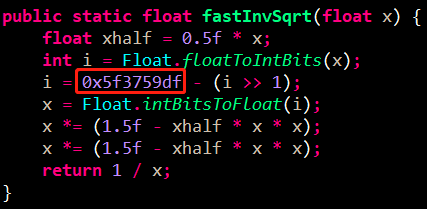
**华丽的打个岔**

计算机科学是严谨的明确的科学，理论重要还是经验重要？

--------------------------------------------------------------------------------------

C数学函数库中的sqrt具有理想的精度，但对于3D游戏的实时光影计算来说速度太慢。我们希望能够在保证足够的精度的同时，进一步提高速度。

坊间传言，***John Carmack***在QUAKE3中使用了快速开平方法，采用了0x5f3759df这个魔数，另外又有人说是nVidia的某位高人发明



该函数的相对误差约为0.177585%，速度比C标准库的sqrt提高超过20%

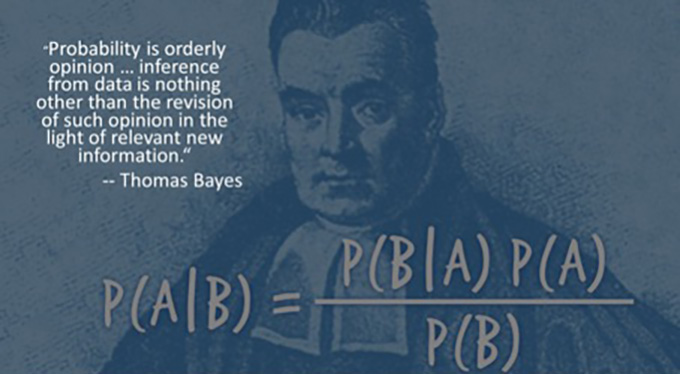
这个神奇的魔数目前还没有太好的数学证明（可能是太繁琐了）

**贝叶斯规则**

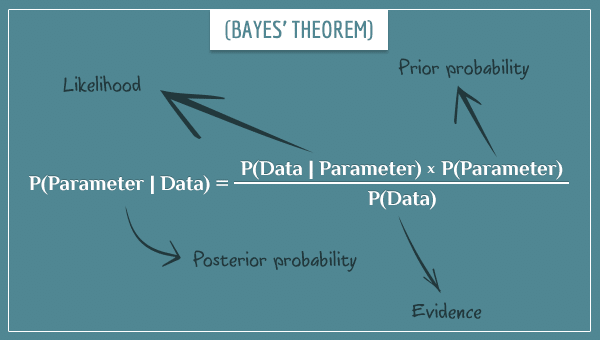
托马斯·贝叶斯 (Thomas Bayes),英国神学家、数学家、数理统计学家和哲学家，1702年出生于英国伦敦，做过神甫；1742年成为英国皇家学会会员；1763年4月7日逝世。

贝叶斯在数学方面主要研究概率论。他首先将归纳推理法用于概率论基础理论，并创立了贝叶斯统计理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。1763年发表了这方面的论著，对于现代概率论和数理统计都有很重要的作用。贝叶斯的另一著作《机会的学说概论》发表于1758年.贝叶斯所采用的许多术语被沿用至今。

他对统计推理的主要贡献是使用了"逆概率"这个概念，并把它作为一种普遍的推理方法提出来。贝叶斯定理原本是概率论中的一个定理，这一定理可用一个数学公式来表达，这个公式就是著名的贝叶斯公式。



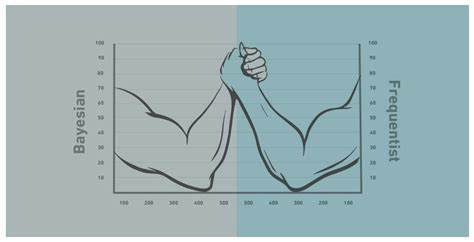
用客观的新信息更新我们最初关于某个事物的信念后，我们就会得到一个新的、改进了的信念。



**简单的贝叶斯思维例子**



**频率主义学派vs贝叶斯学派**



二十世纪二十年代开始，统计学界分裂为两个阵营（Frequentists/Bayesians），开始了至今长达百年的论战。

**频率主义学派认为概率为长期的频率，是唯一确定的，因此需要采集大量数据，尝试不断的逼近唯一真实的那个值（回顾一下LR的推导过程? w初始化，然后随机梯度上升不断修正w，目的是为了似然度最大，直到不能再优化为止）**

**贝叶斯学派认为，概率不确定(本身也会有分布)，但数据是概率的体现，因此可假定一个先验的概率经验值，然后根据数据去修正**

**MLE** = Maximum Likelihood Estimation

**MAP** = Maximum A Posteriori

我们经常在不经意间使用MLE和MAP

逻辑回归就是经典的MLE

入学择校，选医生，择偶，买股票 MAP也有比较多的应用

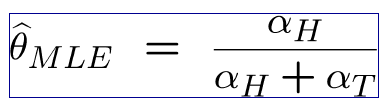
**进阶**

**抛5枚硬币，Frequentist和Bayesian各自的思路**

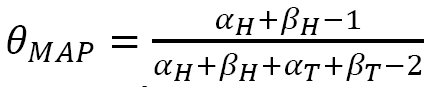
****

**MLE**

****



**MAP**



小知识：

**贝叶斯的实战威力**

第二次世界大战期间，阿兰·图灵发展贝叶斯法则，破解了德国海军的密码。当理论家们把贝叶斯法则视作禁忌时，它把贝尔电话系统从1907年的金融恐慌中拯救了出来，保险精算师用它确定赔率；它指引联军的炮火，找到德国人的潜艇；它确定地震的震中，英国地球物理学家哈罗德·杰弗里斯用它推测出地核是液态的——也许是熔化的铁，或者混合了少量镍。“冷战”期间，贝叶斯法则帮助预测了“挑战者号”的悲剧，证明抽烟会致癌、高胆固醇会引发心脏病等等。今天，贝叶斯法则把色情图片过滤到我们电脑的回收站中。当船沉的时候，海岸警卫队用它寻找也许会在海上漂浮数周的幸存者。科学家发现了基因是如何被控制的。在网上，贝叶斯法则在网上爬梳，售出歌曲和电影。它渗透到了计算机科学、人工智能、机器的学习、华尔街、天文学和物理学、国土安全部、微软和谷歌。它帮助电脑把一种语言翻译成另一种语言。它成了我们的大脑如何学习和运转的一个比喻。杰出的贝叶斯论者甚至给政府部门就教育、能源和科研提出建议。

**频率的本质**

频率学派和贝叶斯学派对世界的认知有本质不同：

频率学派认为世界是确定的，有一个本体，这个本体的真值是不变的，我们的目标就是要找到这个真值或真值所在的范围；

而贝叶斯学派认为世界是不确定的，人们对世界先有一个预判，而后通过观测数据对这个预判做调整，我们的目标是要找到最优的描述这个世界的概率分布。



**Ronald Fisher （1890.2.17－1962.7.29） – Probability as Long-Term Frequency**

**概率 = 长期的频率**



**Frank Plumpton Ramsey （1903.2.22 - 1930.1.19） – Probability as Degree of Belief**

**概率 = 信念的强度**



**Rudolf Carnap（1891.5.18 – 1970.9.14） – Logical Probability**

**概率 = 逻辑上的可能性**

****



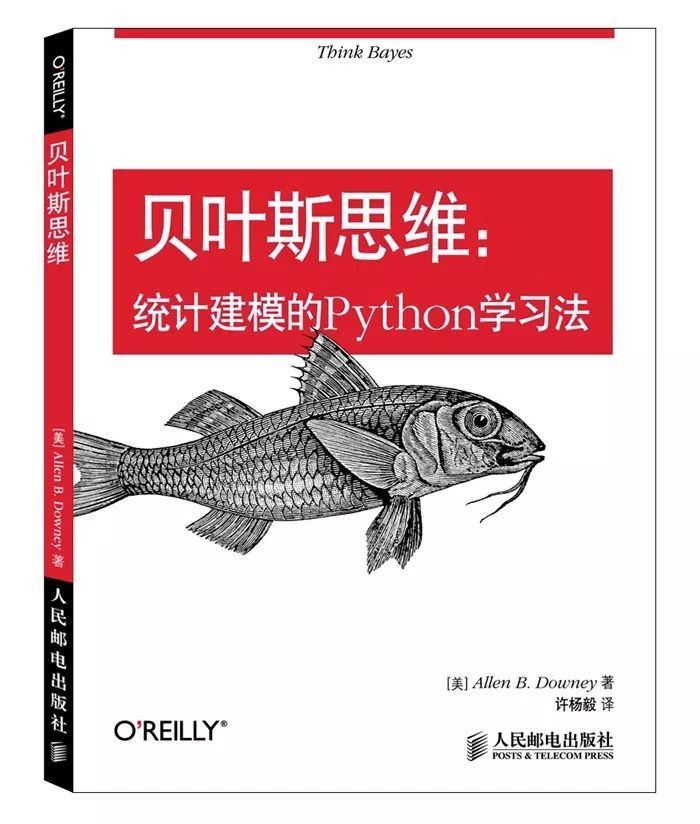
**扩展阅读**

德克萨斯大学达拉斯分校计算机系的贝叶斯讲义，简明的介绍了贝叶斯思维和频率主义的对比<http://www.utdallas.edu/~nrr150130/cs7301/2016fa/lects/Lecture_14_Bayes.pdf>

勘误：（P21的第二个MLE应该为MAP！！！）

卡内基梅隆大学计算机系的机器学习介绍,详细介绍贝叶斯思维最大后验MAP和极大似然估计MLE之间的异同<http://www.cs.cmu.edu/~aarti/Class/10701_Spring14/slides/MLE_MAP_Part1.pdf>

频率主义和贝叶斯主义之间的代表人物和纷争<https://www.digitalvidya.com/blog/frequentist-vs-bayesian/>



Allen Downey写了一本Think Bayes,中文书名《贝叶斯思维：统计建模的Python学习方法》，里面有详尽的理论和Python 2.7代码，非常适合有一定编程基础的同学。

三联生活周刊2011年7月刊贝叶斯法则及其应用<http://www.lifeweek.com.cn/2011/0705/33852.shtml>

**代码附录：**

Git地址：<https://github.com/Yao-lab001/machine_learning_lecture.git>

运行环境：Python 3.7.6 or anaconda 2020.02

预装环境：matplotlib 3.1.3

**下一讲预告**

朴素贝叶斯(下)（代码课）

MLE vs MAP

逻辑回归和朴素贝叶斯在文本领域的具体思路

