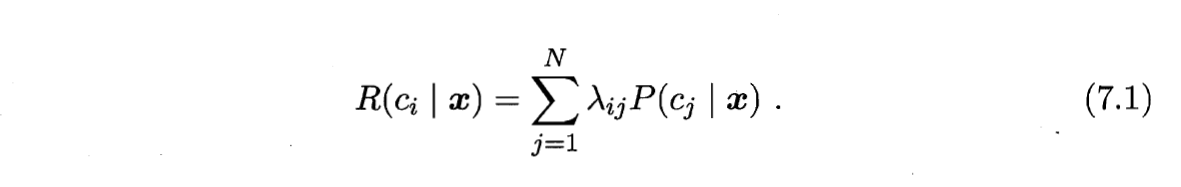
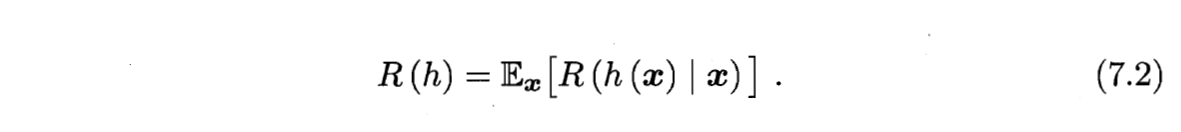
# 7.1贝叶斯决策论

贝叶斯决策论是**概率框架**下实施决策的基本方法。

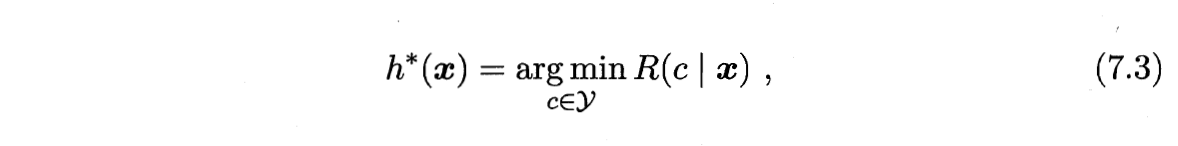
假设有N个类别，λij是将本属于类别j的样本误认为类别i的损失。基于后验概率可获得将样本x分类为第i类（ci）的期望损失：



我们需要找到一个判定准则h以最小化总体风险，表达式如下：



当每一个样本x的期望损失R(h(x) | x)最小，则总体期望损失R(h)最小。引出贝叶斯判定准则：为了最小化整体风险，在每个样本上选择能使条件风险R(c | x)最小的类别标记，即：



此时h\*称为**贝叶斯最优分类器**，与之对应的总体风险R(h\*)称为**贝叶斯风险**，**值越小性能越好**。

**\*\***argmin()当函数值取最小时自变量取值**\*\***

欲使用贝叶斯判定准则来最小化决策风险，首先需要获取后验概率，然而后验概率往往是难以获取的。采用的策略是根据有限的训练样本集估计后验概率（**已知某样本被认为是某个分类，则他其实属于某个分类的概率**）。

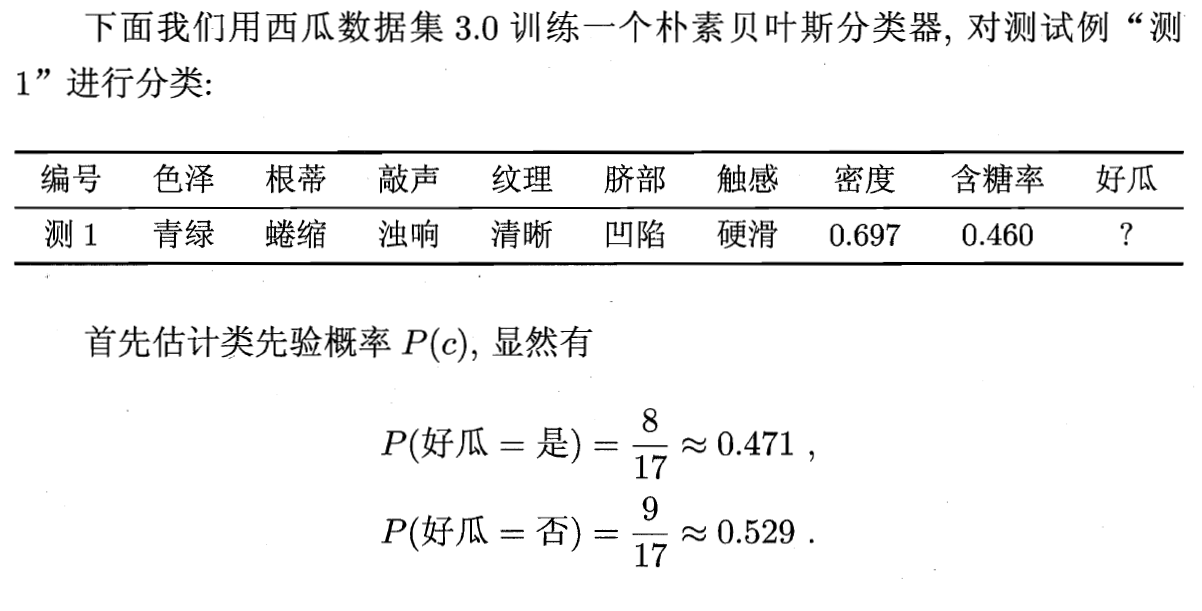
常见策略：1）判别式模型；2）生成式模型

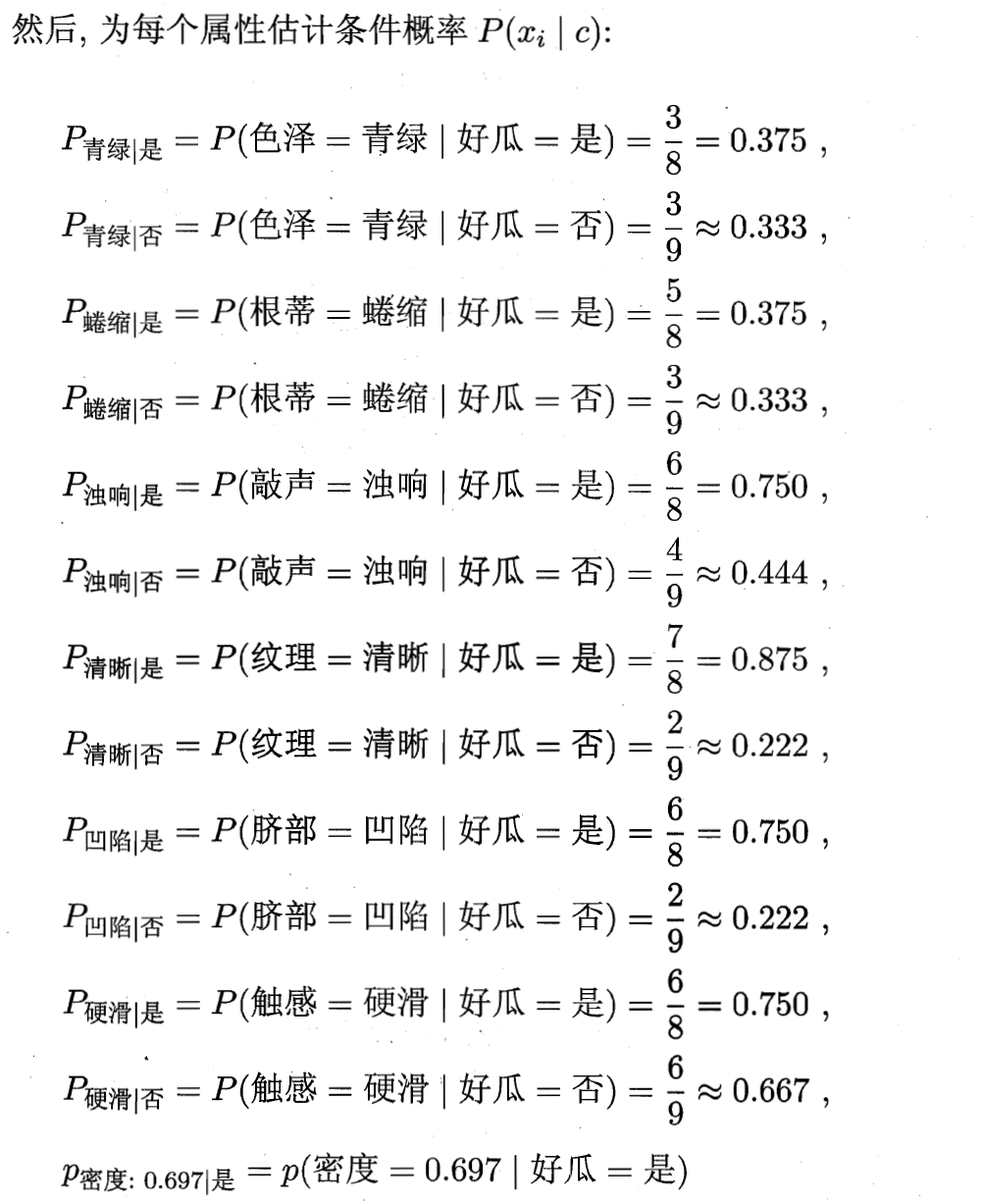
事实上，概率模型的训练过程就是**参数估计**过程。对于参数估计，统计学界的两个学派分别提供了不同的解决方案：**频率主义学派**认为参数虽然未知，但却是**客观存在的固定值**，因此，可通过优化似然函数等准则来确定参数值；**贝叶斯学派**则认为参数是**未观察到的随机变量**，其本身也可有分布，因此，可假定参数服从一个**先验分布**，然后基于观测到的数据来计算参数的**后验分布。**

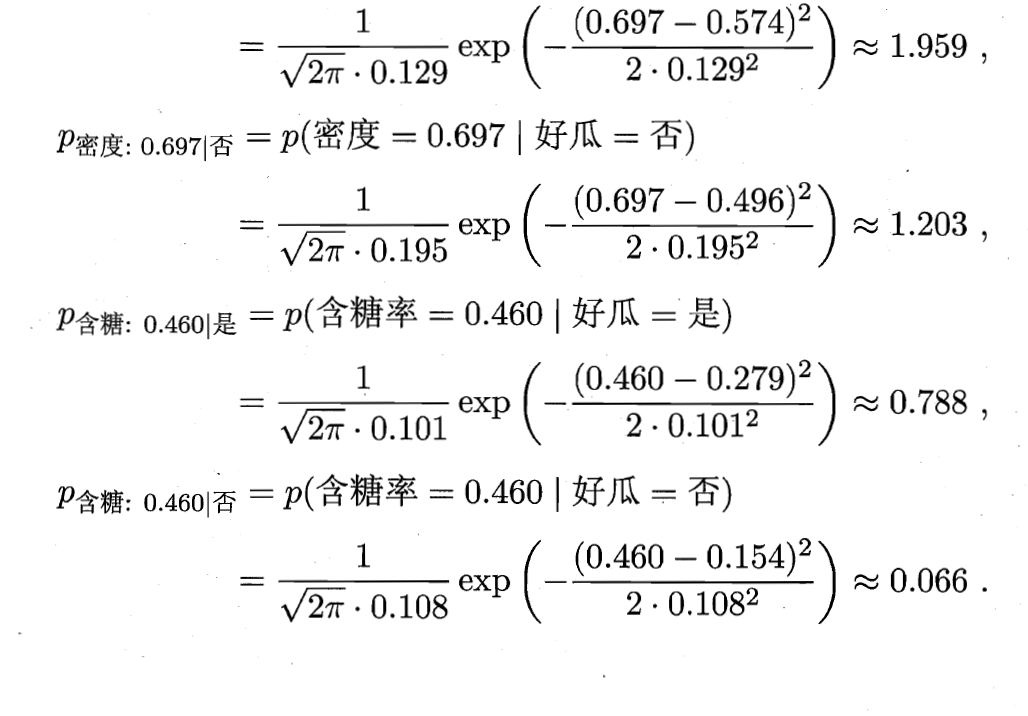
# 7.3朴素贝叶斯分类器

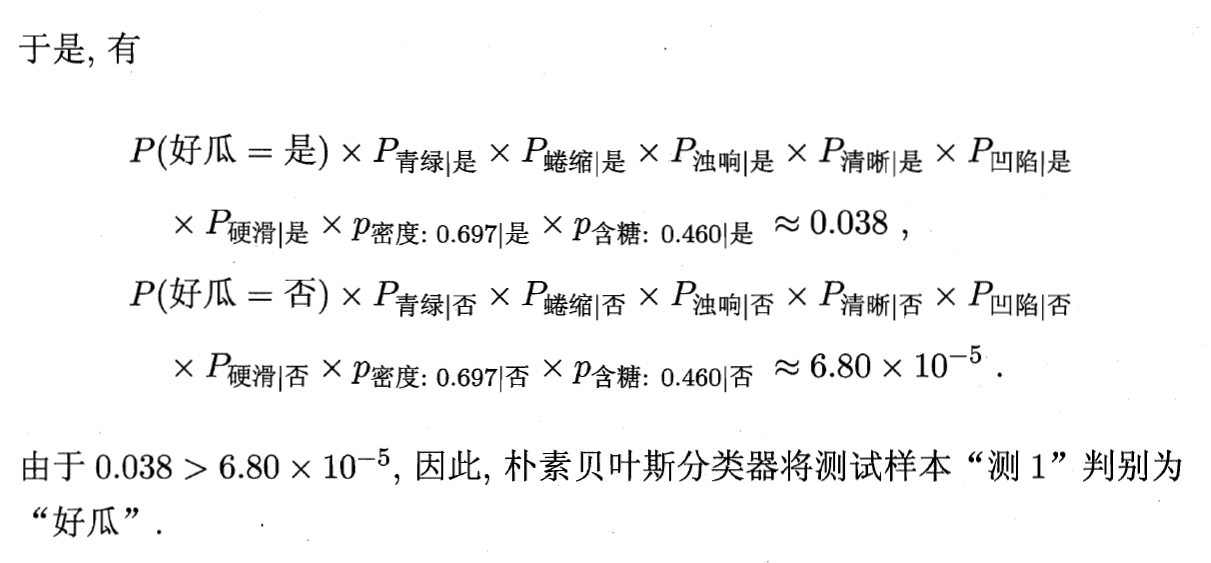
重要思想：假设每个属性**独立**地对分类结果产生影响.

实例：









**拉普拉斯修正：**避免其他属性携带的信息被训练集中**未出现的属性值**“抹去”，避免了因**训练样本不充分**而导致概率估值为零的问题。

# 7.4半朴素贝叶斯分类器

现实任务中，朴素贝叶斯分类器的**属性条件独立性假设**往往不成立，因此尝试对该假设进行一定程度的放松，提出了半朴素贝叶斯分类器。

**“独依赖估计”：**半朴素贝叶斯分类器最常用的一种策略，其假设是“**假设每个属性在类别之外最多仅依赖于一个其他属性**”，如何找到每个属性的父属性是关键问题。

确定父属性的方法：

1.超父独依赖估计分类器（SPODE）：假设所有属性都依赖于同一个属性，称为“超父”。

2.树增广朴素贝叶斯网络分类器（TAN）：基于最大带权生成树算法。

3.平均独依赖估测器（AODE）：基于集成学习机制，将每个属性作为超父构建SPODE，再集成。

# 7.5贝叶斯网

贝叶斯网也称作“信念网“，借助**有向无环图（DAG）**刻画属性依赖关系，并用**条件概率表（CPT）**描述属性的联合概率分布。

贝叶斯网学习的**首要步骤**是根据训练数据集确定结构最恰当的贝叶斯网（结构恰当说明准确描述了属性依赖关系）。

上述问题常用的**评分函数**通常基于信息论准则，将学习问题看作一个数据压缩任务，目标是找到一个能以最短编码长度描述训练数据的模型。此为“最小描述长度“准则。

找到最优结构的问题是一个NP难问题，无法快速求解，常用的求近似解的方法是从某个网络结构出发，每次调整一条边，直到评分函数值不再降低（**值越低编码越短，结构越好**）。

推断时若根据联合概率分布精确计算则是一个NP难问题，因此通常采用吉布斯采样降低精度近似计算。

吉布斯采样核心思想：在贝叶斯网所有变量的联合状态空间内与证据一致的子空间进行随机漫步。每一步仅依赖于前一步的状态，是一个“马尔可夫链“。由于马尔可夫链通常需要很长时间才能趋于平稳分布，因此吉布斯采样算法收敛速度较慢。

# 7.6EM算法

考虑数据部分缺失的情况（隐变量）。

算法步骤：

1.期望（E）步：利用当前估计的参数值来计算对数似然的期望值

2.最大化（M）步：寻找能使E步产生的似然期望最大化的参数值，再执行E步，循环直至收敛。