贝叶斯决策论通过**相关概率已知**的情况下利用**误判损失**来选择最优的类别分类

贝叶斯定理

$$P(c_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|c_i)P(c_i)}{P(\boldsymbol{x})}$$

利用贝叶斯定理，可将计算转换成对$P(x|c_i)$、(x)、$P(c_i)$的计算。

一般来说，$P(c_i)$为先验概率，$P(\boldsymbol{x}|c_i)$为条件概率，$P(\boldsymbol{x})$是用于归一化的证据因子。对于$P(c_i)$可以通过训练样本中类别为c_i的样本所占的比例进行估计；此外，由于只需要找出最大的$P(\boldsymbol{x}|c_i)$，因此我们并不需要计算$P(\boldsymbol{x})$。

朴素贝叶斯分类器

$ P(x|c)$难以直接计算，朴素贝叶斯分类器采用"属性条件独立假设"：（假设每个属性独立地对结果发生影响），则

$$P(c|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|c)P(c)}{P(\boldsymbol{x})} = \frac{P(c)}{P(x)}\prod_{i=1}^{d}P(x_i|c)$$

贝叶斯决策理论的核心思想，**就是选择最高概率的决策**：对于每一个输入x，分别计算这几个类概率的值，哪个大x属于哪一个类。可以证明，若要最小化损失函数的期望，就要使其概率最高，即预测输出

$$\hat{y} = \arg\max_{c_i \in Y} P(c_i)\prod_{j=1}^{d}P(x_j|c_i)$$

如果x_j是标签属性，那么可以通过计数的方法估计$P(x_j|c_i)$

$$P(x_j|c_i) = \frac{P(x_j, c_i)}{P(c_i)} \approx \frac{\#(x_j, c_i)}{\#(c_i)}$$

其中，$\#(x_j, c_i)$表示在训练样本中x_j与c_i共同出现的次数。

例子

使用经典的西瓜训练集如下：

编号	色泽	根蒂	敲声	纹理	脐部	触感	密度	含糖率	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	0.697	0.460	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	0.774	0.376	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	0.634	0.264	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	0.608	0.318	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	0.556	0.215	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	0.403	0.237	是

编号	色泽	根蒂	敲声	纹理	脐部	触感	密度	含糖率	好瓜
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	0.481	0.149	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	0.437	0.211	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	0.666	0.091	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	0.243	0.267	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	0.245	0.057	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	0.343	0.099	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	0.639	0.161	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	0.657	0.198	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	0.360	0.370	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	0.593	0.042	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	0.719	0.103	否

对下面的测试例“测1”进行 分类：

编号	色泽	根蒂	敲声	纹理	脐部	触感	密度	含糖率	好瓜
测1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	0.697	0.460	?

首先，估计类先验概率$P(c_j)$，有

$$P(\text{好瓜} = \text{是}) = \frac{8}{17} = 0.471$$
$$P(\text{好瓜} = \text{否}) = \frac{9}{17} = 0.529$$

然后，为每个属性估计条件概率（这里，对于连续属性，假定它们服从正态分布）

$$P_{\text{青绿}|\text{是}} = P(\text{色泽} = \text{青绿} | \text{好瓜} = \text{是}) = \frac{3}{8} = 0.375$$
$$P_{\text{青绿}|\text{否}} = P(\text{色泽} = \text{青绿} | \text{好瓜} = \text{否}) = \frac{3}{9} \approx 0.333$$
$$P_{\text{蜷缩}|\text{是}} = P(\text{根蒂} = \text{蜷缩} | \text{好瓜} = \text{是}) = \frac{5}{8} = 0.625$$
$$P_{\text{蜷缩}|\text{否}} = P(\text{根蒂} = \text{蜷缩} | \text{好瓜} = \text{否}) = \frac{3}{9} = 0.333$$
$$P_{\text{浊响}|\text{是}} = P(\text{敲声} = \text{浊响} | \text{好瓜} = \text{是}) = \frac{6}{8} = 0.750$$
$$P_{\text{浊响}|\text{否}} = P(\text{敲声} = \text{浊响} | \text{好瓜} = \text{否}) = \frac{4}{9} \approx 0.444$$
$$P_{\text{清晰}|\text{是}} = P(\text{纹理} = \text{清晰} | \text{好瓜} = \text{是}) = \frac{7}{8} = 0.875$$
$$P_{\text{清晰}|\text{否}} = P(\text{纹理} = \text{清晰} | \text{好瓜} = \text{否}) = \frac{2}{9} \approx 0.222$$

$$P_{凹陷|是}=P(脐部=凹陷|好瓜=是)=\frac{6}{8}=0.750$$

$$P_{凹陷|否}=P(脐部=凹陷|好瓜=否)=\frac{2}{9}\approx 0.222$$

$$P_{硬滑|是}=P(触感=硬滑|好瓜=是)=\frac{6}{8}=0.750$$

$$P_{硬滑|否}=P(触感=硬滑|好瓜=否)=\frac{6}{9}\approx 0.667$$

$$\begin{aligned}\rho_{密度：0.697|是}&=\rho(密度=0.697|好瓜=是)\\&=\frac{1}{\sqrt{2\pi}\times 0.129}exp\left(-\frac{(0.697-0.574)^2}{2\times 0.129^2}\right)\approx 1.959\end{aligned}$$

$$\begin{aligned}\rho_{密度：0.697|否}&=\rho(密度=0.697|好瓜=否)\\&=\frac{1}{\sqrt{2\pi}\times 0.195}exp\left(-\frac{(0.697-0.496)^2}{2\times 0.195^2}\right)\approx 1.203\end{aligned}$$

$$\begin{aligned}\rho_{含糖：0.460|是}&=\rho(密度=0.460|好瓜=是)\\&=\frac{1}{\sqrt{2\pi}\times 0.101}exp\left(-\frac{(0.460-0.279)^2}{2\times 0.101^2}\right)\approx 0.788\end{aligned}$$

$$\begin{aligned}\rho_{含糖：0.460|否}&=\rho(密度=0.460|好瓜=是)\\&=\frac{1}{\sqrt{2\pi}\times 0.108}exp\left(-\frac{(0.460-0.154)^2}{2\times 0.108^2}\right)\approx 0.066\end{aligned}$$

于是有

$$\begin{aligned}&P(好瓜=是)\times P_{青绿|是}\times P_{蜷缩|是}\times P_{浊响|是}\times P_{清晰|是}\times P_{凹陷|是}\\&\times P_{硬滑|是}\times p_{密度：0.697|是}\times p_{含糖：0.460|是}\approx 0.063\end{aligned}$$

$$\begin{aligned}&P(好瓜=否)\times P_{青绿|否}\times P_{蜷缩|否}\times P_{浊响|否}\times P_{清晰|否}\times P_{凹陷|否}\\&\times P_{硬滑|否}\times p_{密度：0.697|否}\times p_{含糖：0.460|否}\approx 6.80\times 10^{-5}\end{aligned}$$

由于$0.063>6.80\times 10^{-5}$，因此，朴素贝叶斯分类器将测试样本“测1”判别为“好瓜”。