

贝叶斯分类器

贝叶斯决策论是概率框架下实施决策的基本方法，对于分类任务来说，在所有相关概率都已知的理想情况下，贝叶斯决策论考虑如何基于这些概率和误判损失来选择最优的类别标记。

假设有N种可能的类别标记，即 $Y = c_1, c_2, c_3 \dots c_N$ ， λ_{ij} 是将一个真实标记为 c_j 的样误分类为 c_i 所产生的损失。基于后验概率 $P(c_i | x)$ 可以获得将样本 x 误分类成 c_i 所产生的期望损失，记载样本 x 上的“条件风险”

$$R(c_i | x) = \sum_{j=1}^N \lambda_{ij} P(c_j | x)$$

我们的任务是寻找一个判定准则 $h: X \rightarrow Y$ 以最小化总体风险

$$R(h) = E_x [R(h(x) | x)]$$

显然，对于每个样本 x ，若 h 能最小化条件风险 $R(h(x) | x)$ 则总体风险 $R(h)$ 也将被最小化，这就产生了贝叶斯判定准则：为最小化总体风险，只需在每个样本上选择那个能使条件风险 $R(c | x)$ 最小的类别标记，即

$$h^*(x) = \operatorname{argmin}_{c \in Y} R(c | x)$$

此时，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， $1 - R(h^*)$ 反映了分类器所能达到的最好性能，即通过机器学习所能产生的模型精度的理论上限。

具体来说，若目标是最小化分类错误率，则误判损失 λ_{ij} 可以表示为，当 $i=j$ 时， $\lambda_{ij} = 0$ ，否则为1。此时条件风险为*

$$R(c | x) = 1 - P(c | x)$$

*,于是，最小化分类错误率的贝叶斯最优分类器为

$$h^*(x) = \operatorname{argmax}_{c \in Y} P(c | x)$$

朴素贝叶斯分类器

朴素贝叶斯分类器采用了“属性条件独立性假设”：对已知类别，假设所有属性相互独立，换言之，假设每个属性独立地对分类结果发生影响。基于属性条件独立性假

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

由于对于所有类别来说 $P(x)$ 相同，贝叶斯判定准则有

$$h_{nb}(x) = \operatorname{argmax}_{c \in Y} P(c) \prod_{i=1}^d P(x_i | c)$$