朴素贝叶斯

Li Liang[*]

1 贝叶斯决策论

考虑一个多分类问题，假设有一个输入向量 x，我们的目标是对于 x 预测它的类别 y，其中 $y = \{C_1, C_2, ..., C_N\}$。记 x 是类别 C_j 的概率为 $p(x, C_j)$。假设 x 的真实类别是 C_i，则 x 被错误分类的概率为：

$$p(mistake) = \sum_{j=1, j\neq i}^{N} p(x, C_j) \tag{1}$$

我们当然希望错误分类的概率越小越好：

$$\arg\min \sum_{j=1, j\neq i}^{N} p(x, C_j) \tag{2}$$

这就是最小化分类错误率。但是，对于许多应用来说，犯错所得的损失是不一样的，我们要权衡不同犯错的损失。计 λ_{ij} 是将真实类别为 C_i 的样本误分类为 C_j 所产生的损失。错误分类产生的损失，也称期望损失为：

$$EL = \sum_{j=1}^{N} \lambda_{ij} p(x, C_j) \tag{3}$$

对于输入 x，分别计算划分为不同类的期望损失，将其分类为期望损失最小的类。

$$\arg\min \sum_{j=1}^{N} \lambda_{ij} p(x, C_j) \tag{4}$$

也就是最小化期望损失。将联合概率形式写成条件概率：

$$\arg\min \sum_{j=1}^{N} \lambda_{ij} p(C_j \mid x)\ p(x) \tag{5}$$

注意到 $p(x)$ 在所有类的期望损失都相同，上式等价于：

$$\arg\min \sum_{j=1}^{N} \lambda_{ij} p(C_j \mid x) \tag{6}$$

[*] https://github.com/leeliang/

2 朴素贝叶斯分类器

若我们的损失函数简化为（期望损失简化为最小化错误分类）:

$$\lambda_{ij} = \begin{cases} 1, & k \neq i \\ 0, & k = i \end{cases} \tag{7}$$

(6) 式等价于:

$$\begin{aligned} \arg\min \sum_{j=1}^{N} \lambda_{ij} p(C_j \mid x) &= \arg\min \sum_{j=1, j\neq i}^{N} p(C_j \mid x) \\ &= \arg\min [1 - p(C_i \mid x)] \\ &= \arg\max p(C_i \mid x) \end{aligned} \tag{8}$$

对于输入 x，分别计算划分为不同类后验概率（$p(C_i \mid x)$），将其分类为后验概率最大的类。在现实任务中，后验概率通常难以直接获取，需要需用贝叶斯公式。基于贝叶斯公式，有:

$$p(C_i \mid x) = \frac{p(C_i)\ p(x \mid C_i)}{p(x)} \tag{9}$$

最大化后验概率，等价于:

$$arg \max p(C_i)\ p(x \mid C_i) \tag{10}$$

也就是需要估计先验概率 $p(C_i)$ 和类条件概率 $p(x \mid C_i)$。针对具体问题，类条件概率 $p(x \mid C_i)$ 是所有特征 x 上的联合概率，难以从有限的样本直接估计而得。为避开这个问题，朴素贝叶斯分类器采用了特征独立性假设：即对已知类别，假设所有属性相互独立。基于该假设，(10) 式重写为:

$$arg \max p(c) \prod_{i=1}^{d} p(x_i \mid C_i) \tag{11}$$

d 为 x 的属性个数。

因此，朴素贝叶斯分类器的训练过程就是基于数据集 D 来估算先验概率 $p(C_i)$ 和每个属性的条件概率 $p(x_i \mid C_i)$。