# 朴素贝叶斯

## 前言

关于朴素贝叶斯，我在之前的线性回归的高斯判别分析(GDA)和朴素贝叶斯(NB)提到过，但是并未详细进行讲解。是不相瞒，NB 朴素贝叶斯其实就是基于我们高中学过的贝叶斯简单推导得到的。还记得么？先验概率可以用来计算后验概率，而朴素贝叶斯就是一种基于贝叶斯公式的分类算法，它从训练数据中估计先验概率和似然度，然后利用这些估计来预测后验概率用于分类。

## 朴素贝叶斯

在前面的推导中，我们有贝叶斯公式(7.8)来估计后验概率 $P(c|x)$ ，但是一般来说，我们都会面临一些困难，比如：条件概率 $P(x|c)$ 是所有属性上的联合概率（什么是联合概率，见GDA），很难从已有的有限训练样本得到。为了解决这个问题，我们采用了属性条件独立性假设，也就是属性之间相互独立（注意区分独立与不相关的区别）。我们可以将(7.8)重写为：

$$P(c|x)=\frac{P(c)P(x|c)}{P(x)}=\frac{P(c)}{P(x)}\prod_{i=1}^{d}P(x_i|c),\quad(7.14)$$

这个公式中，我们单独提出了条件概率，将条件概率 $P(x|c)$ 进行了分解，得到了所有 $x$ 属性( $x_i$ )之和，也就是 $\prod_{i=1}^{d}P(x_i|c)$ ，它的前提就是，所有属性条件独立（这就是它为什么叫朴素的原因，因为它的假设太大胆了）。其余的， $d$ 为属性数目， $x_i$ 是 $x$ 的第 $i$ 个属性。

如果你觉得抽象，那么就看看下面这个例子：对于“如果天气是热、高温、晴”三种情况下，我们不出门。它的联合概率为 $P(\text{热，高温，晴}|\text{不出门})$ ，但是如果属性条件独立，那么就可以写为 $P(\text{热}|\text{不出门})\times P(\text{高温}|\text{不出门})\times P(\text{晴}|\text{不出门})$ 。

接下来，由于对于所有类别来说 $P(x)$ 证据概率都相同，所以我们其实可以不考虑证据概率。那么我们基于(7.6)的贝叶斯判定准则则有：

$$h_{nb}(x)=\arg\max_{c\in\mathcal{Y}}\frac{P(c)\prod_{i=1}^{d}P(x_i|c)}{P(x)}=\arg\max_{c\in\mathcal{Y}}P(c)\prod_{i=1}^{d}P(x_i|c),\quad(7.15)$$

之前说过的，这个公式是对于每个样本 $x$ ，选择能使后验概率 $P(c|x)$ 最大的类别标记。这就是朴素贝叶斯分类器的表达式，这就是基于训练集 $D$ 来估计先验概率 $P(c)$ ，并为每个属性估计条件概率 $P(x_i|c)$ 。

而对于先验概率，我们则是令 $D_c$ 表示训练集 $D$ 中第 $c$ 类样本组合成的集合，若有充分独立同分布样本，则可以估计出先验概率（注意，独立同分布！）：

$$P(c)=\frac{\text{某一类}}{\text{整个集合}}=\frac{|D_c|}{|D|},\quad(7.16)$$

对于离散属性而言，令 $D_{c,x_i}$ 表示 $D_c$ 中在第 $i$ 个属性上取值为 $x_i$ 的样本组成的集合，则条件概率 $P(x_i|c)$ 可估计为：

$$P(x_i|c) = \frac{\text{类中某个属性的个数}}{\text{类总体}} = \frac{|D_{c,x_i}|}{|D_c|}, \quad (7.17)$$

而对于非离散数据，则可以使用相应的概率密度函数，还记得我上面说的独立同分布么？就是那个！如果符合正态分布，那就用正态，其他的同理！

## 拉普拉斯修正和其他

说白了就是为了防止概率为零，然后在分子分母上同时加 1，就这么简单。专业点说就是：避免了因训练集样本不充分而导致概率估值为零的问题，并且在训练集变大时，修正过程所引入的先验(prior)的影响也会逐渐变得可忽略，使得估值渐趋向于实际概率值。

所以拉普拉斯修正为：

$$\hat{P}(c) = \frac{|D_c + 1|}{|D| + N}, \quad (7.19)$$

$$\hat{P}(x_i|c) = \frac{|D_{c,x_i}| + 1}{|D_c| + N_i}, \quad (7.20)$$

在现实任务中朴素贝叶斯分类器有多种使用方式。例如，若任务对预测速度要求较高，则对给定训练集，可将朴素贝叶斯分类器涉及的所有概率估值事先计算好存储起来，这样在进行预测时只需“查表”即可进行判别；若任务数据更替频繁，则可采用“懒惰学习”(lazy learning)方式，先不进行任何训练，待收到预测请求时再根据当前数据集进行概率估值；若数据不断增加，则可在现有估值基础上，仅对新增样本的属性值所涉及的概率估值进行计数修正即可实现增量学习。