朴素贝叶斯法

朴素贝叶斯法
- 4.1 朴素贝叶斯的学习与分类
  - 4.1.1 基本方法
  - 4.1.2 后验概率最大化含义
- 4.2 参数估计
  - 4.2.1 极大似然估计
  - 4.2.2 学习与分类算法
  - 4.2.3 贝叶斯估计
- 后记-连续型属性

> - 朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法;
> - 给定训练数据集,目的是基于特征条件独立假设学习输入/输出的联合概率分布,然后基于此模型,对于给定的输入x,利用贝叶斯定理求出后验概率最大的输出y;
> - **朴素贝叶斯(naive Bayes)与贝叶斯估计(Bayesian estimation)是不同的概念**;

4.1 朴素贝叶斯的学习与分类

4.1.1 基本方法

- 朴素贝叶斯通过训练数据集学习**联合概率分布**$P(X, Y)$,具体地学习以下*先验概率分布*和*条件概率分布*,于是学习到联合概率分布
  - 先验概率分布:

$$P(Y=c_k), k=1,2,\ldots,K \tag{1}$$

  - 条件概率分布:

$$P(X=x|Y=c_k)=P(X^{(1)}=x^{(1)},\ldots,X^{(n)}=x^{(n)}|Y=c_k), k=1,2,\ldots,K \tag{2}$$

- 上述条件概率分布具有*指数级*数量的参数，其估计是不可行的（假设$x^{(d)}$可取值S_d个，那么参数个数为$K\prod_{d=1}^{n}S_d$）；
- **条件独立性假设**：一个强假设，因此得名

$$P(X=x|Y=c_k)=\prod_{d=1}^{n}P(X^{(d)}=x^{(d)}|Y=c_k) \tag{3}$$

- 朴素贝叶斯实际上学习到生成数据的机制，属于生成模型；
- **分类预测**：对于给定输入x，通过模型计算后验概率分布$P(Y=c_k|X=x)$，将后验概率最大的类作为输出，根据贝叶斯定理

$$\begin{aligned}P(Y=c_k|X=x)&=\frac{P(X=x|Y=c_k)P(Y=c_k)}{\sum_k P(X=x|Y=c_k)P(Y=c_k)}\\&=\frac{P(Y=c_k)\prod_d P(X^{(d)}=x^{(d)}|Y=c_k)}{\sum_k P(Y=c_k)\prod_d P(X^{(d)}=x^{(d)}|Y=c_k)}\end{aligned}$$

- **朴素贝叶斯分类器**（分母是全概率公式，结果一致，可以省略）：

$$y=f(x)=\arg\max_{c_k}P(Y=c_k)\prod_d P(X^{(d)}=x^{(d)}|Y=c_k) \tag{4}$$

4.1.2 后验概率最大化含义

- 可以证明取后验概率最大的类等价于*期望风险最小化*；

4.2 参数估计

4.2.1 极大似然估计

- 先验概率：

$$P(Y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N},k=1,2,\ldots,K \tag{5}$$

- 条件概率（设第d个特征$x^{(d)}$的可能取值集合为$\{a_{d1},a_{d2},\ldots,a_{dS_d}\}$）：

$$P(X^{(d)}=a_{dl}|Y=c_k)=\frac{\sum_{i=1}^{N}I(x_i^{(d)}=a_{dl},y_i=c_k)}{\sum_{i=1}^{N}I(y_i=c_k)},l=1,2,\ldots,S_d \tag{6}$$

4.2.2 学习与分类算法

- **朴素贝叶斯算法**：
    - 计算训练集的先验概率和条件概率；
    - 对于给定的实例$x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T$，计算：

$$P(Y=c_k)\prod_d P(X^{(d)}=x^{(d)}|Y=c_k), k=1,2,\dots,K \tag{7}$$

    - 选择后验概率最大的作为输出类；

4.2.3 贝叶斯估计

- 使用MLE可能会出现所要的估计概率为0的情况，会影响后验概率结果，使分类产生偏差；
- **条件概率的贝叶斯估计**：

$$P_\lambda(X^{(d)}=a_{dl}|Y=c_k)=\frac{\sum_{i=1}^N I(x_i^{(d)}=a_{dl}, y_i=c_k)+\lambda}{\sum_{i=1}^N I(y_i=c_k)+S_d\lambda}, l=1,2,\dots,S_d \tag{8}$$

其中$\lambda \geq 0$，相当于在各个频数上赋予了一个基础正数λ；当$\lambda = 0$时就是极大似然估计；经常取值$\lambda = 1$，称为拉普拉斯平滑（Laplace smoothing）；可以证明经过上述调整后仍然是一种概率分布（$p > 0, \sum p = 1$）；

- **先验概率的贝叶斯估计**：

$$P_\lambda(Y=c_k)=\frac{\sum_{i=1}^N I(y_i=c_k)+\lambda}{N+K\lambda}, k=1,2,\dots,K \tag{9}$$

后记-连续型属性

- 当属性值是连续型时，可以假设连续变量服从某种概率分布，然后使用训练数据估计分布的参数，*高斯分布*经常被用来表示连续属性的类条件概率分布：

$$P(X^{(d)}=a_d|Y=c_k)=\frac{1}{\sqrt{2\pi}\sigma_{dk}^2}\exp\left(-\frac{(a_d-\mu_{dk})^2}{2\sigma_{dk}^2}\right) \tag{10}$$

其中参数μ_{dk}与σ_{dk}^2可以用类c_k的所有训练数据关于特征$X^{(d)}$的样本均值和方差估计；