

Linear Classification 06 Naive Bayes

Chen Gong

04 November 2019

本节主要是介绍一下 Naive Bayes Classification，也就是朴素贝叶斯分类。朴素贝叶斯分类器的核心思想也就是，条件独立性假设。这是一种最简单的概率图模型，也就是一种有向图模型。

1 条件独立性假设

条件独立性假设用简单的图来进行表述，可以表示为如下图所示的形式：

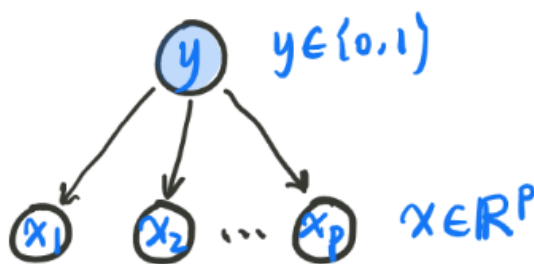


图 1: 条件独立性假设

我们可以将其定义为 $x_i \perp x_j | y$ ($i \neq j$)。根据贝叶斯公式可以得：

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x,y)}{p(x)} \propto p(x,y) \quad (1)$$

而做条件独立性假设的最终目的，是为了简化运算。因为对于一个数据序列 $x = (x_1, x_2, \dots, x_p)$ 。如果 x_i 和 x_j 之间有关系的话，这个计算难度可能会变得很难，所以就假设各个变量之间是相互独立的。而且，马尔可夫决策链也是这样类似的思想。

2 Naive Bayes Classification

朴素贝叶斯算法的优化目的即为：

$$\begin{aligned} \hat{y} &= \arg \max_{y \in \{0,1\}} p(y|x) \\ &= \arg \max_{y \in \{0,1\}} p(x|y)p(y) \end{aligned} \quad (2)$$

其中,

$$p(x|y) = \prod_{i=1}^N p(x_i|y) \quad (3)$$

对于 $p(y)$ 这个先验概率密度函数的确定, 对于二分类问题, 也就是 $y \sim \text{Bernoulli Distribution}$, 而对于多分类问题, 先验概率为 $y \sim \text{Categorical Distribution}$ 。而对于, $p(x|y) = \prod_{i=1}^N p(x_i|y)$ 。如果 x 是离散的, 那么 $x_i|y \sim \text{Categorical Distribution}$; 如果 x 是连续的, 那么 $x_i|y \sim \mathcal{N}(\mu_y, \Sigma_y^2)$ 。对于每一类都有一个高斯分布。

而有关于 $p(x|y)$ 用极大似然估计 MLE, 估计出来就行。因为分布的形式我们已经知道了, 那么只要利用数据来进行学习, 使用极大似然估计就可以得到想要的结果了。其实对于多分类的情况, Naive Bayes Classification 和 Guassian Discriminate Analysis 很像的。