朴素贝叶斯

xxr

2021 年 3 月 7 日

1 算法简介

1.1 原理

监督学习分为生成模型 (generative model) 与判别模型 (discriminative model)，贝叶斯方法是生成模型的代表。在概率论与统计学中，贝叶斯定理 (Bayes' theorem) 表达了一个事件发生的概率，而确定这一概率的方法是基于与该事件相关的条件先验知识 (prior knowledge)。而利用相应先验知识进行概率推断的过程为贝叶斯推断 (Bayesian inference)。

1.2 贝叶斯定理

条件概率 (conditional probability) 是指在事件 B 发生的情况下，事件 A 发生的概率。通常记为 P(A | B)。

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

因此

$$P(A \cap B) = P(A|B)P(B) \tag{2}$$

$$P(A \cap B) = P(B|A)P(A) \tag{3}$$

可得：

$$P(A|B)P(B) = P(B|A)P(A) \tag{4}$$

由此可以推断出贝叶斯公式：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5}$$

1.3 贝叶斯推断

贝叶斯公式中，P(A) 称为" 先验概率"（Prior probability），即在 B 事件发生之前，对 A 事件概率的一个判断。P(A|B) 称为" 后验概率"（Posterior probability），即在 B 事件发生之后，对 A 事件概率的重新评估。P(B|A)/P(B) 称为" 可能性函数"（Likelyhood），这是一个调整因子，使得预估概率更接近真实概率。所以，条件概率可以理解成下面的式子：

$$Posterior_{probability} = Prior_{probability} * Likelyhood$$

这就是贝叶斯推断的含义。我们先预估一个" 先验概率"，然后加入实验结果，看这个实验到底是增强还是削弱了" 先验概率"，由此得到更接近事实的" 后验概率"。因为在分类中，只需要找出可能性最大的那个选项，而不需要知道具体那个类别的概率是多少，所以为了减少计算量，全概率公式在实际编程中可以不使用。

而朴素贝叶斯推断，是在贝叶斯推断的基础上，对条件概率分布做了条件独立性的假设。因此可得朴素贝叶斯分类器的表达式。因为以自变量之间的独立（条件特征独立）性和连续变量的正态性假设为前提，就会导致算法精度在某种程度上受影响:

$$\widetilde{y} = arg\,\underset{c \in Y}{max} \sum_{i=1}^{d} P(x_i|c) \tag{6}$$

1.4 朴素贝叶斯的参数推断

实际在机器学习的分类问题的应用中，朴素贝叶斯分类器的训练过程就是基于训练集 D 来估计类先验概率 P(c) ，并为每个属性估计条件概率 P(xi | c) 。这里就需要使用极大似然估计 (maximum likelihood estimation, 简称 MLE) 来估计相应的概率.

令 D_c 表示训练集 D 中的第 c 类样本组成的集合，若有充足的独立同分布样本，则可容易地估计出类别的先验概率：

$$P(C) = \frac{|D_c|}{D} \tag{7}$$

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:

$$P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|} \tag{8}$$

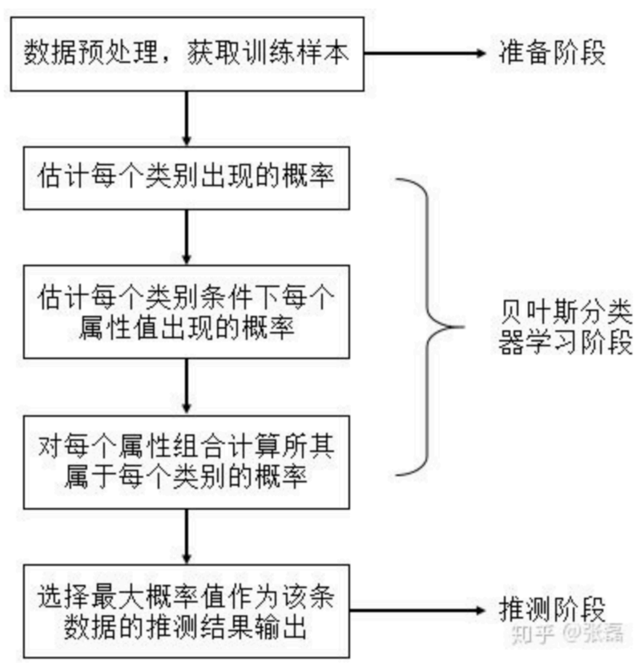
对于连续属性可考虑概率密度函数，假定:

$$P(x_i|c) \sim \mathcal{N}(\mu_c^i, \sigma_{c,i}^2) \tag{9}$$

μ 和 σ 分别是第 c 类样本在第 i 个属性上取值的均值和方差，则有：

$$P(x_i|c) = \frac{1}{\sqrt{2 * \pi}\sigma_{c,i}} exp(-\frac{(x_i - \mu_{c,i}^2)}{2\sigma_{c,i}2}) \tag{10}$$

1.5 算法流程



贝叶斯方法流程