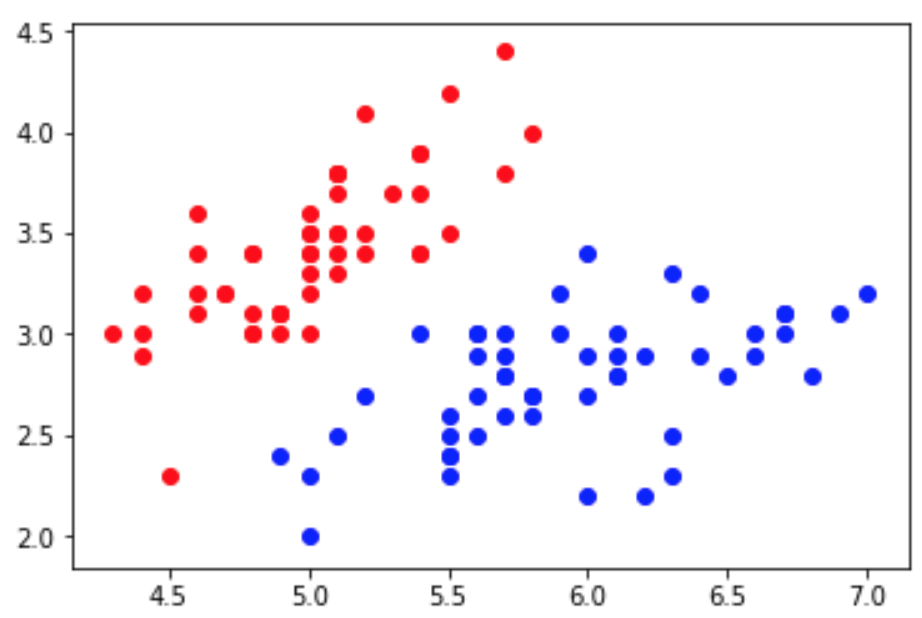
# 朴素贝叶斯

* 概述

**朴素贝叶斯(Native Bayes)算法，属于监督学习算法，基于贝叶斯定理对样本类型进行分类的一种解决分类问题的算法。**

**朴素：给定目标值时属性之间相互条件独立。**

****

**假设存在一数据集，由两类数据组成且分类结果已知晓。**

**现出现一新样本，分类未知。**

**可用p1(x,y)表示样本为红色一类的概率。**

**可用p2(x,y)表示样本为蓝色一类的概率。**

**我们提出如下规则：**

**如果p1(x,y)>p2(x,y)，则样本属于红色一类**

**如果p1(x,y)<p2(x,y)，则样本属于蓝色一类**

**也就是说，选择概率高的一类作为新样本的分类结果。**

**这就是贝叶斯决策理论的核心思想，即选择具有最高概率的决策。**

* 贝叶斯公式推导

根据条件概率的定义，在事件B发生的条件下事件 A发生的概率为：



同样地，在事件 A发生的条件下事件 B发生的概率为



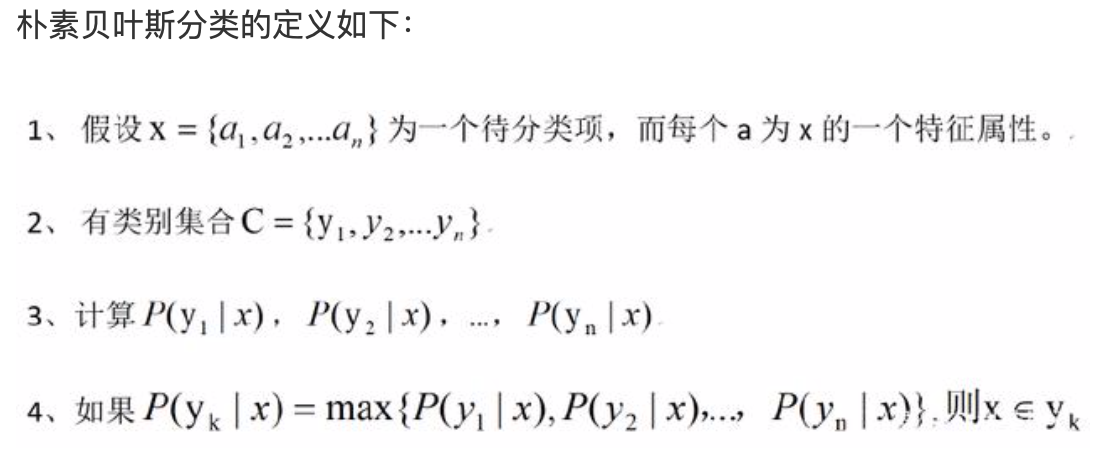
**结合这两个方程式，我们可以得到**

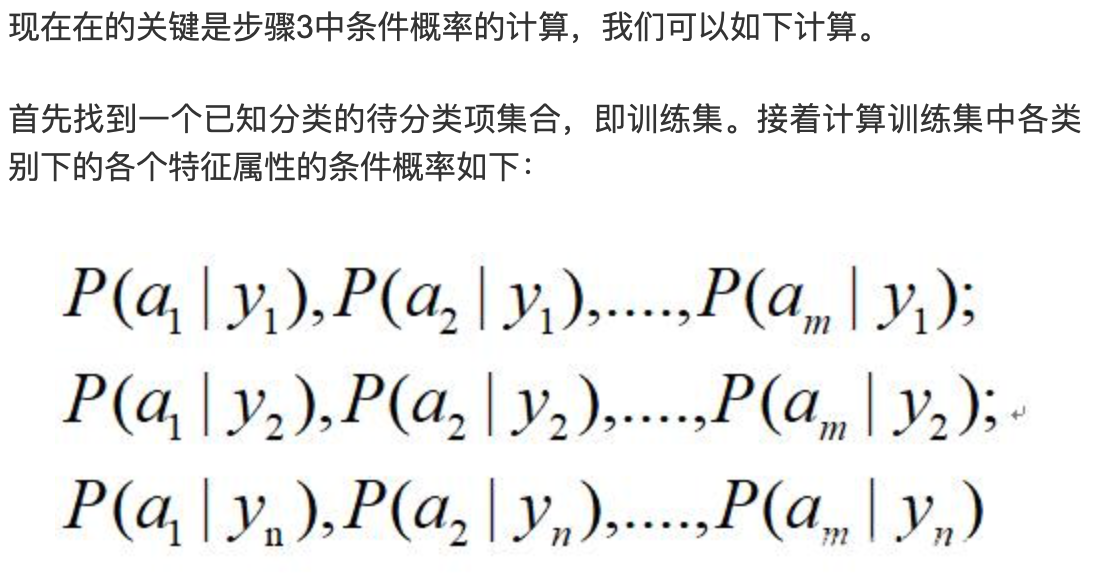


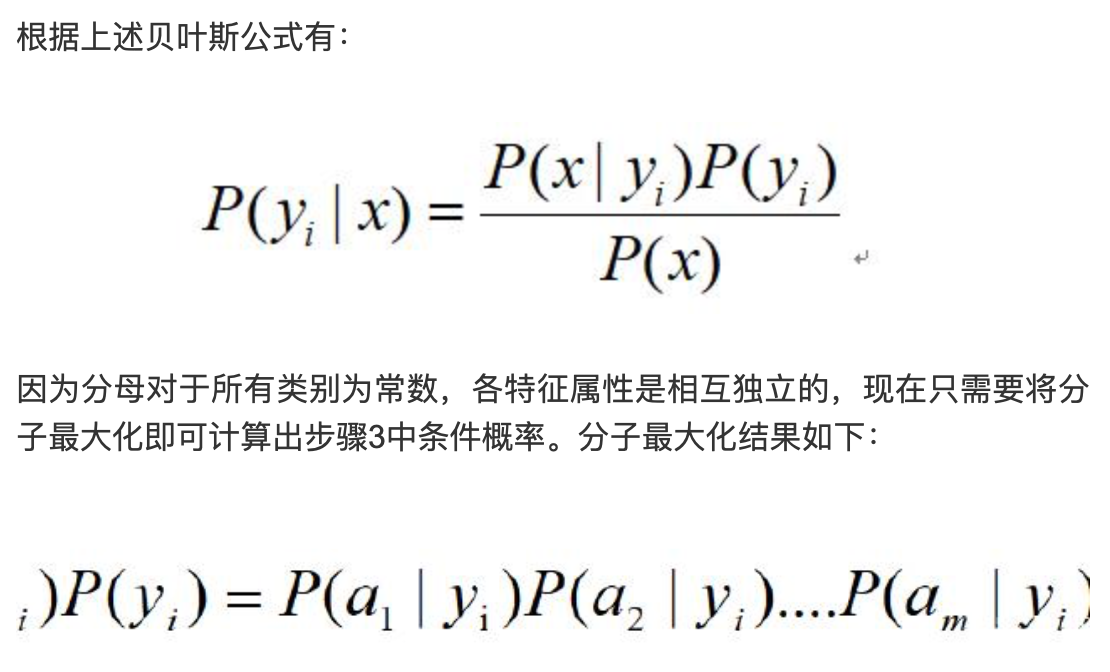
上式两边同除以 P(A)，若P(A)是非零的，我们可以得到贝叶斯定理:



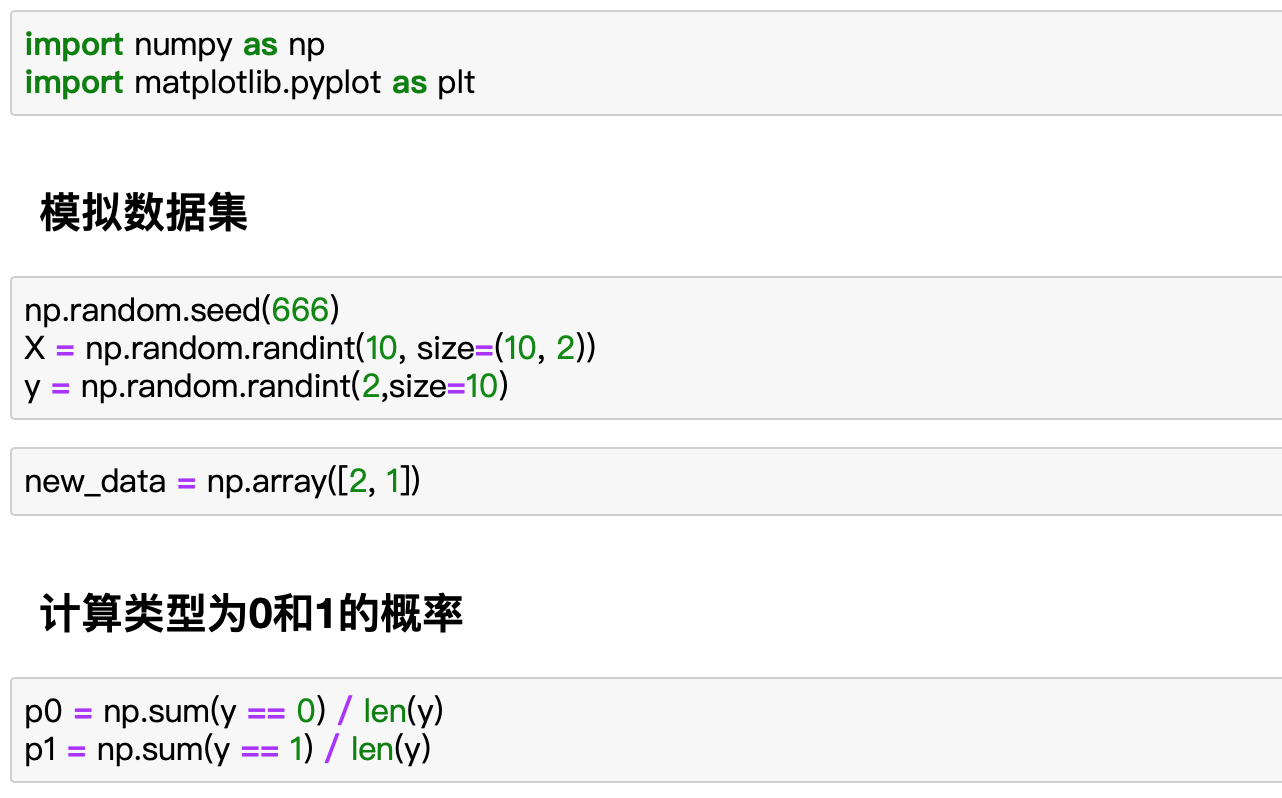
通常，事件 A 在事件 B 发生的条件下的概率，与事件 B在事件 A发生的条件下的概率是不一样的；然而，这两者是有确定关系的，贝叶斯定理就是这种关系的陈述。

****

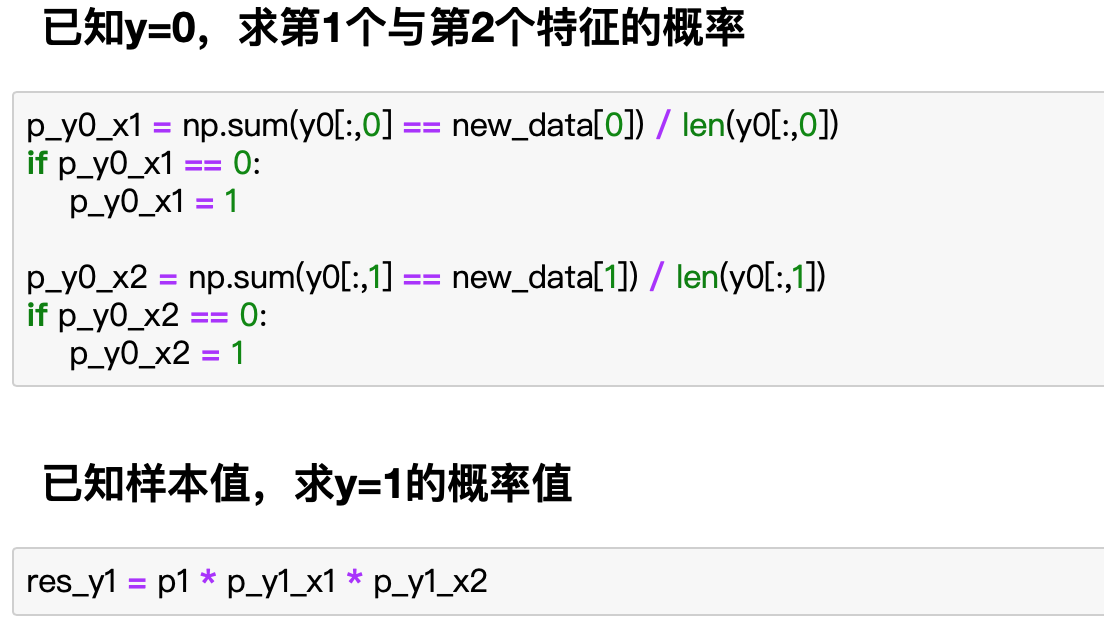
****

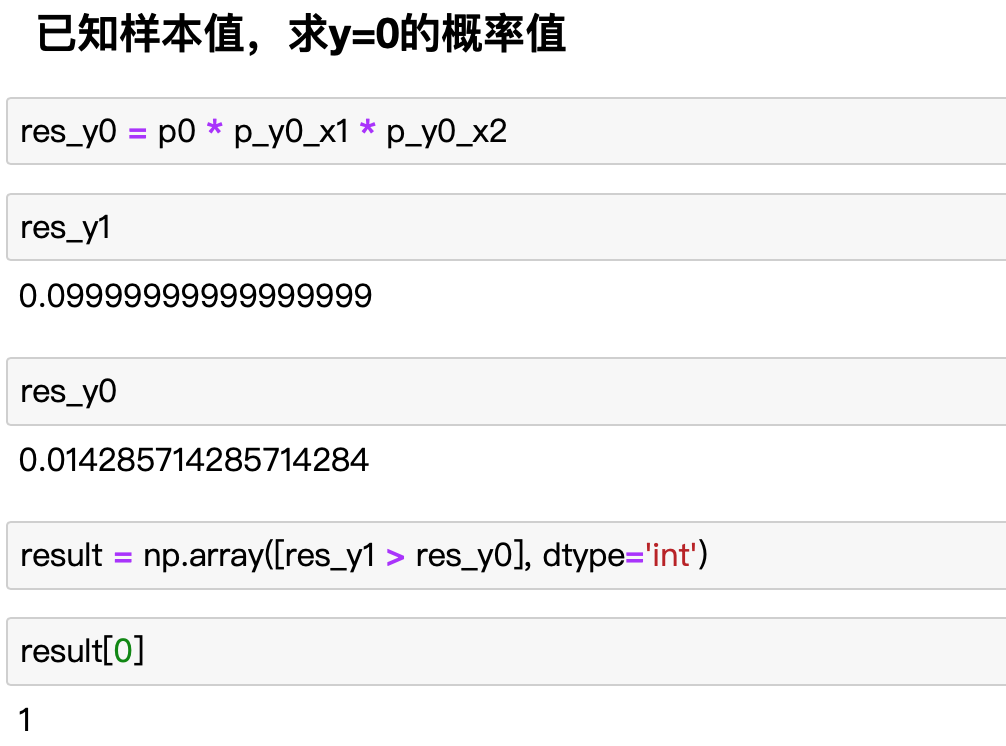
****

* 贝叶斯公式实现

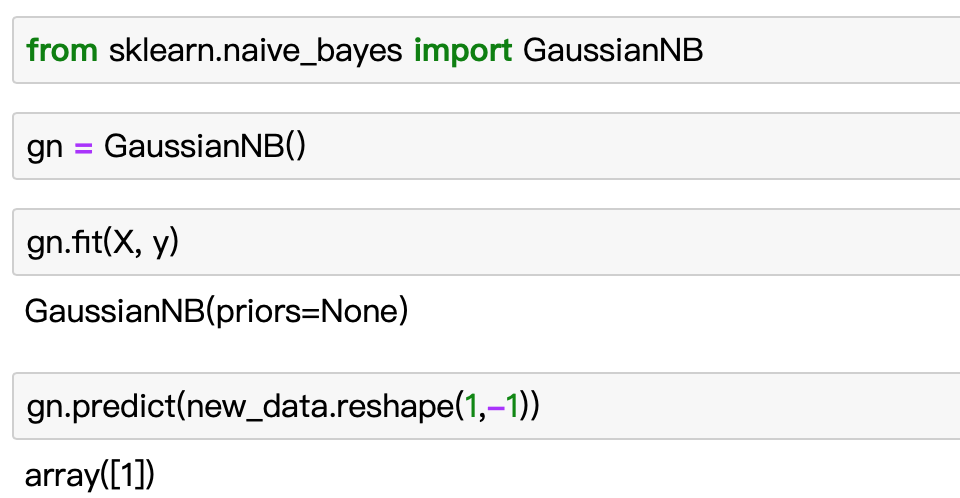
****

****

****

****

* Sklearn中的朴素贝叶斯

****