写出全概率公式&贝叶斯公式？

全概率公式

$$P(A) = \sum_{n} P(A|B_n)P(B_n)$$

贝叶斯公式

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

朴素贝叶斯为什么“朴素 naive”？

朴素贝叶斯，(Navie Bayesian)中的朴素可以理解为是“简单、理想化”的意思，因为“朴素”是假设了样本特征之间是相互独立、没有相关关系。这个假设在现实世界中是很不真实的，属性之间并不是都是互相独立的，有些属性也会存在相关性，所以说朴素贝叶斯是一种很“朴素”的算法。

朴素贝叶斯有没有超参数可以调？

基础朴素贝叶斯模型的训练过程，本质上是通过数学统计方法从训练数据中统计先验概率$P(c)$和后验概率$P(x_i|c)$；而这个过程是不需要超参数调节的。所以朴素贝叶斯模型没有可调节的超参数。虽然在实际应用中朴素贝叶斯会与拉普拉斯平滑修正（Laplacian smoothing correction）一起使用，而拉普拉斯平滑修正方法中有平滑系数这一超参数，但是这并不属于朴素贝叶斯模型本身的范畴。

朴素贝叶斯的工作流程是怎样的？

朴素贝叶斯的工作流程可以分为三个阶段进行，分别是准备阶段、分类器训练阶段和应用阶段。

准备阶段：这个阶段的任务是为朴素贝叶斯分类做必要的准备，主要工作是根据具体情况确定特征属性，并对每个特征属性进行适当划分，去除高度相关性的属性(如果两个属性具有高度相关性的话，那么该属性将会在模型中发挥了 2 次作用，会使得朴素贝叶斯所预测的结果向该属性所希望的方向偏离，导致分类出现偏差)，然后由人工对一部分待分类项进行分类，形成训练样本集合。这一阶段的输入是所有待分类数据，输出是特征属性和训练样本。(这一阶段是整个朴素贝叶斯分类中唯一需要人工完成的阶段，其质量对整个过程将有重要影响。)

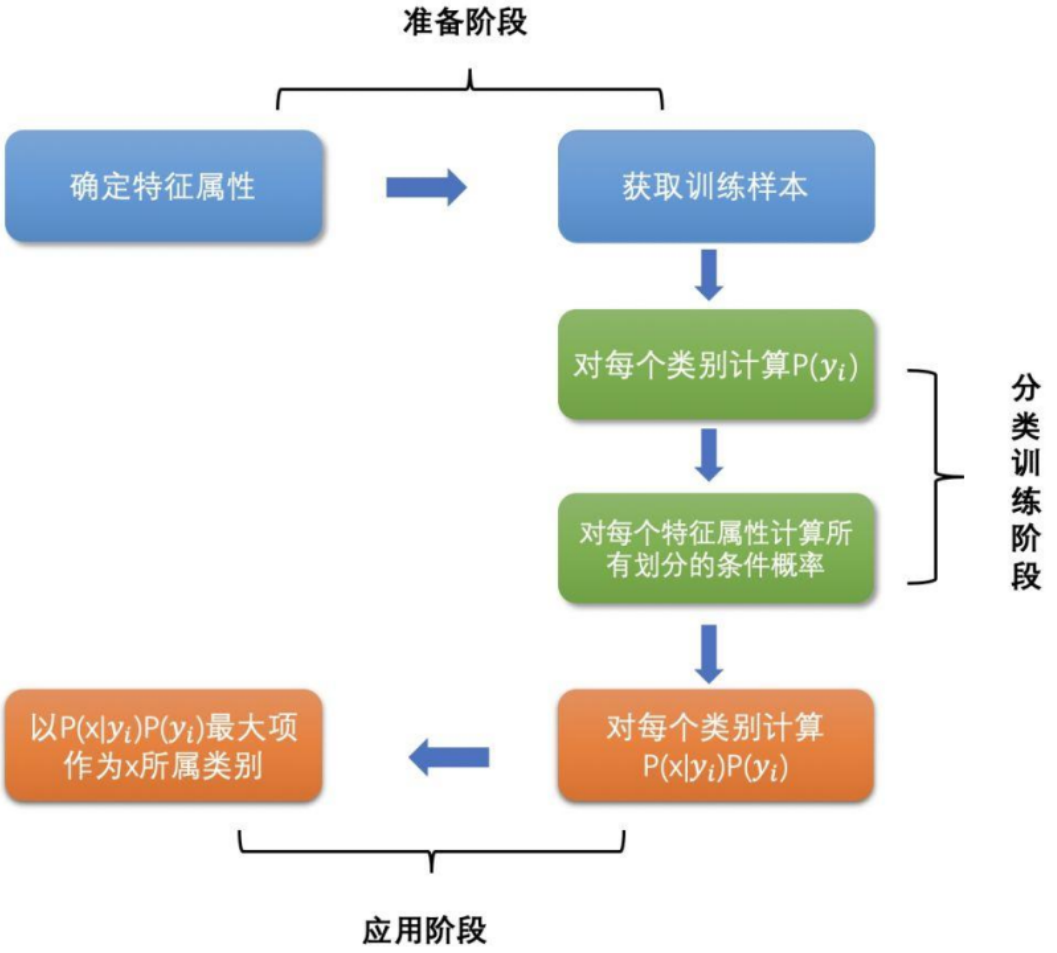
分类器训练阶段：这个阶段的任务就是生成分类器，主要工作是计算每个类别在训练样本中的出现频率及每个特征属性划分对每个类别的条件概率估计，并将结果记录。其输入是特征属性和训练样本，输出是分类器。从公式上理解，朴素贝叶斯分类器模型的训练目的就是要计算一个后验概率$P(c|x)$使得在给定特征的情况下，模型可以估计出每个类别出现的概率情况。

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

因为$P(x)$是一个先验概率，它对所有类别来说是相同的；而我们在预测的时候会比较每个类别相对的概率情况，选取最大的那个作为输出值。所以我们可以不计算$P(x)$。所以贝叶斯学习的过程就是要根据训练数据统计计算先验概率$P(c)$和后验概率$P(x_i|c)$。

应用阶段：这个阶段的任务是使用分类器对待分类项进行分类，其输入是分类器和待分类项，输出是待分类项与类别的映射关系。并选出概率值最高所对应的类别；用公式表示即为：

$$h(x) = argmax_{c\in y} P(c)\prod_{i=1}^{d} P(x_i|c)$$



朴素贝叶斯对异常值敏不敏感？

基础的朴素贝叶斯模型的训练过程，本质上是通过数学统计方法从训练数据中统计先验概率$P(c)$和后验概率$P(x_i|c)$，少数的异常值，不会对统计结果造成比较大的影响。所以朴素贝叶斯模型对异常值不敏感。