高斯模型是对于每个属性的值是连续的、且服从高斯分布

priors:先验概率大小,如果没有给定,模型则根据样本数 据自己计算(利用极大似然法)。

如果大部分是多元离散值,则采用多项式模型要好些。多 项式模型,通常就是构造参数向量,然后通过极大似然估 计来寻求参数的最有值。

alpha:先验平滑因子,默认等于1,当等于1时表示拉普 拉斯平滑。

fit\_prior:是否去学习类的先验概率,默认是True

class\_prior:各个类别的先验概率,如果没有指定,则模 型会根据数据自动学习,每个类别的先验概率相同,等 于类标记总个数N分之一。

用于多重伯努利分布的数据,即有多个特征,但每个特征 都假设是一个二元 (Bernoulli, boolean) 变量

alpha:平滑因子,与多项式中的alpha一致。

binarize:样本特征二值化的阈值,默认是0。如果不输 入,则模型会认为所有特征都已经是二值化形式了;如果 输入具体的值,则模型会把大于该值的部分归为一类,小 于的归为另一类。

fit\_prior:是否去学习类的先验概率,默认是True

class\_prior:各个类别的先验概率,如果没有指定,则模 型会根据数据自动学习,每个类别的先验概率相同,等 于类标记总个数N分之一。

sklearn.naive bayes.GaussianNB

sklearn.naive\_bayes.MultinomialNB

sklearn.naive\_bayes.BernoulliNB

sklearn.naive\_bayes

朴素贝叶斯

贝叶斯: p(x/y)=p(x,y)/p(y)=p(y/x)\*p(x)/p(y) 已知结 果求原因的概率

总概

朴素: 特征向量的各个特征独立同分布

朴素贝叶斯高效、但可能分类性能一般

求参数时一般使用最大似然估计,这样可能出现概率为0 影响后验概率的计算,所以会采用拉普拉斯平滑,即分子 分母加上一个整数。

GaussianNB、先验为高斯分布的朴素贝叶斯。

MultinomialNB,先验为多项式分布的朴素贝叶斯。

BernoulliNB,先验为伯努利分布的朴素贝叶斯。

分类

注意