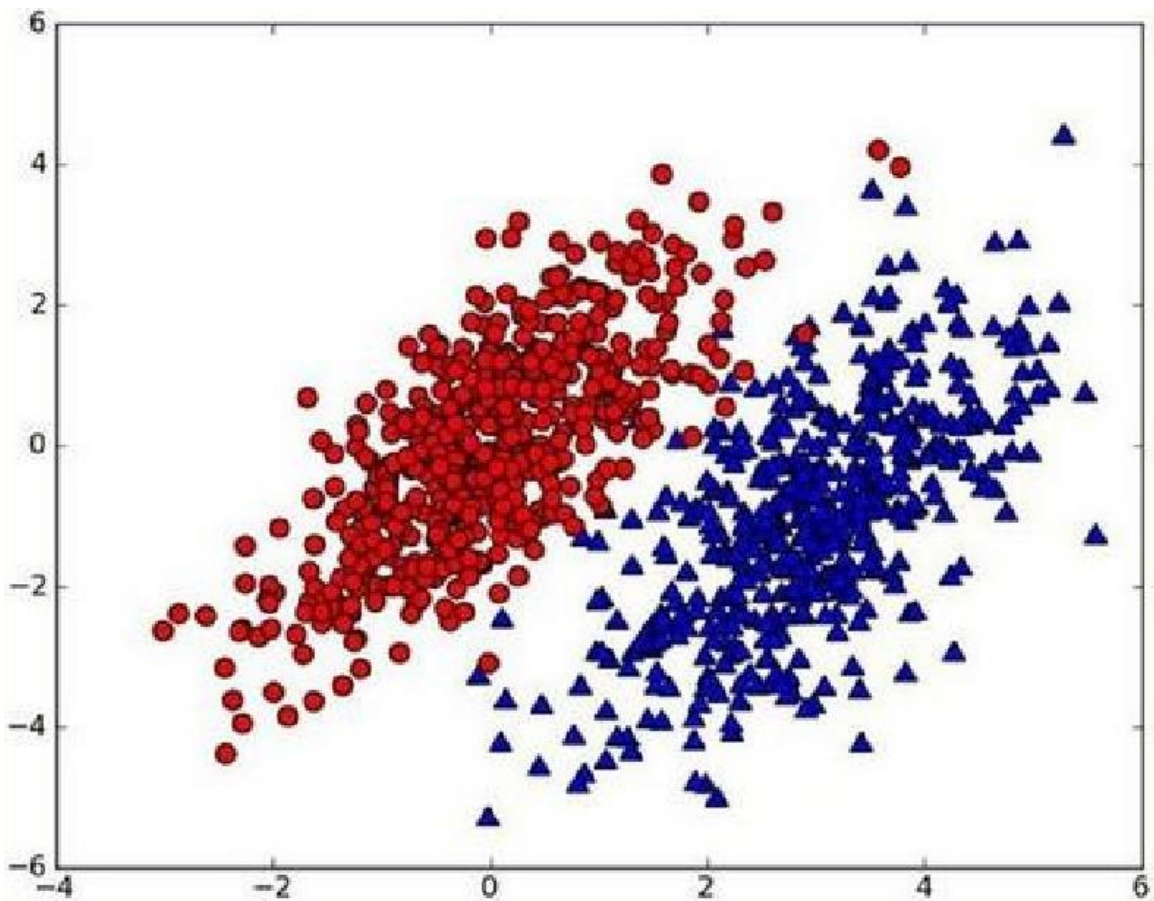


1、贝叶斯

假设我们有一个数据集，组成成分如图所示：



我们现在用 $p1(x, y)$ 表示数据点 (x, y) 属于类别 1 (图中红色圆点表示的类别) 的概率，用 $p2(x, y)$ 表示数据点 (x, y) 属于类别 2 (图中蓝色三角形表示的类别) 的概率，那么对于一个新数据点 (x, y) ，可以用下面的规则来判断它的类别：

如果 $p1(x, y) > p2(x, y)$ ，那么类别为 1

如果 $p1(x, y) < p2(x, y)$ ，那么类别为 2

贝叶斯算法核心思想是利用条件概率和全概率公式推断出 $P(A|B) = P(A) \frac{P(B|A)}{P(B)}$ 来进行计算的

$P(A)$ 称为“先验概率”，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”，即在 B 事件发生之后，我们对 A 事件概率的重新评估。

$P(B|A)/P(B)$ 称为“可能性函数”，即调整因子，使得预估概率更接近真实概率。

2、朴素贝叶斯

朴素贝叶斯对条件个概率分布做了条件独立性的假设，有 n 个特征的计算为：

$$P(a|X) = p(X|a)p(a) = p(x_1, x_2, x_3, \dots x_n|a)p(a)$$

3、总结

朴素贝叶斯推断的一些优点：

生成式模型，通过计算概率来进行分类，可以用来处理多分类问题。

对小规模的数据表现很好，适合多分类任务，适合增量式训练，算法也比较简单。

朴素贝叶斯推断的一些缺点：

对输入数据的表达形式很敏感。

由于朴素贝叶斯的“朴素”特点，所以会带来一些准确率上的损失。

需要计算先验概率，分类决策存在错误率。