

软件环境：

pycharm

实现贝叶斯分类器，并实现分类

实验步骤：

1. 先利用样本点求出正态分布的均值和方差（协方差）。
2. 然后根据贝叶斯公式，用判别函数 $g_i(x) = \ln p(x | \omega_i) + \ln P(\omega_i)$ 计算最大后验概率。

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

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^t \sum_i^{-1} (x - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\sum_i| + \ln P(w_i)$$

3. 比较同一样本点 x 的各个 $g_i(x)$ 的大小，确定样本点 x 的类别。
4. 比较原来的已知类别和通过分类器确定的类别，可以计算出分类器的误差；通过求 Bhattacharyya 界，可以知道分类器的误差率的上界。
5. 第一次只用第一个特征值 x_1 构造分类器，之后重复以上步骤，分别用两个和三个特征值构造分类器。

结果：

对于只用一个特征值 x_1 分类：

样本的经验训练误差为 $0.3000 = 30\%$

Bhattacharyya 误差界为 $p(\text{error}) \leq 0.474$

[1, 2, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 2, 2, 2, 1, 2, 2, 2, 2]

错误的点的个数：

6

30.0

对于用特征值 x_1 和 x_2 分类：

样本的经验训练误差为 $0.4500 = 45\%$

Bhattacharyya 误差界为 $p(\text{error}) \leq 0.46047$

[1, 2, 1, 1, 2, 2, 2, 1, 2, 1, 1, 1, 2, 2, 2, 1, 2, 2, 2, 1]

错误的点的个数:

9

45.0

对于用特征值 x_1 、 x_2 和 x_3 分类:

样本的经验训练误差为 $0.1500 = 15\%$

Bhattacharyya 误差界为 $p(\text{error}) \leq 0.41193$

[1, 0.000, 1.000]

[1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2]

错误的点的个数:

3

15.0