

实验三 基于单高斯模型的二分类

最大似然准则下的均值与协方差推导

高斯模型:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\therefore L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\text{令 } \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = 0 \quad \therefore \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad \therefore \mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = 0 \quad \therefore -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

$$\therefore \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

编程思路

- 先用 `textscan()` 结合提取格式提取出Train.txt中的A、B两组数据。
- 由A、B两组数据计算对应的高斯模型的均值和协方差矩阵。
- 利用得到的均值和协方差矩阵构建出高斯概率模型。
- 依次比较Test.txt中各点在两高斯概率模型中的概率，并将该点归类为概率较大的那组。
- 由分类结果与实际对比获得准确率。

实验结果

- 对测试集的分类准确率为0.7301，且在用单高斯模型对原训练集进行分类测试时结果为0.6473。
- 由此可见单高斯模型具有一定的分类能力，但总得来说分类准确率较差，可用于简单的数据分类。