

实验三 基于单高斯模型的二分类

实验内容：

用多元高斯模型解决二分类问题。假设两类有相同的先验概率。
每个观测样本的特征是 3 维向量。基于训练数据(网站上：
Train.txt)创建单高斯模型，然后用估计的模型对测试数据进行分
类(网站上：Test.txt)

每个类用一个 3-D 多元高斯分布来建模：

$$p(x) = |(2\pi)^d \Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

最大似然估计： μ 用样本的均值估计，协方差矩阵 Σ 为

$$\Sigma = \begin{bmatrix} \sigma(x_1, x_1) & \cdots & \sigma(x_1, x_d) \\ \vdots & \ddots & \vdots \\ \sigma(x_d, x_1) & \cdots & \sigma(x_d, x_d) \end{bmatrix} \in \mathbb{R}^{d \times d}$$

对于这些随机变量，我们还可以根据协方差的定义，用样本估计出两两之间的协方差，即

$$\sigma(x_m, x_k) = \frac{1}{N-1} \sum_{i=1}^N (x_{mi} - \mu_m)(x_{ki} - \mu_k)$$

其中, 为方便书写, x_{ki} 表示随机变量 x_k 中的第*i*个观测样本, N 表示样本量, 每个随机变量所对应的观测样本数量均为 N。

这里假设每个多元高斯的协方差矩阵为对角矩阵

实验要求：

1. 本次实验不限制编程语言，可以使用 MATLAB，python 等语言。
2. 请将以下文件打包发送至 algorithm_2022@126.com

文件名如：实验一_姓名_学号.zip

邮件主题:实验一_姓名_学号

a) 实验报告：推导最大似然准则下均值和协方差的估计公式；

提交在测试集上的预测准确率

b) 实验代码