

# 第一次 课后作业

---

提交截止时间：2022 年 10 月 30 日 20: 00

问题一：

- 1) (20 分) 生成两个各包含  $N = 1000$  个二维样本的数据集  $X_1$  和  $X_2$ 。数据集中的样本来自于三个正态分布，分布的均值矢量分别为  $\mathbf{m}_1 = [1, 1]^T$ 、 $\mathbf{m}_2 = [4, 4]^T$  和  $\mathbf{m}_3 = [8, 1]^T$ ，分布的协方差矩阵为  $\Sigma_1 = \Sigma_2 = \Sigma_3 = 2\mathbf{I}$ ，其中  $\mathbf{I}$  是  $2 \times 2$  维的单位矩阵。在生成数据集  $X_1$  时，样本来自三个分布模型的先验概率相同，而在生成数据集  $X_2$  时，先验概率分别为  $p(\omega_1) = 0.6$ 、 $p(\omega_2) = 0.3$  和  $p(\omega_3) = 0.1$ 。分别画出所生成两个数据集中样本的散布图。
- 2) (30 分) 在两个数据集上分别应用“似然率决策规则”、“贝叶斯风险决策规则”（其中  $C_{12} = 2, C_{13} = 3, C_{23} = 2.5, C_{11} = C_{22} = C_{33} = 0, C_{21} = C_{31} = C_{32} = 1$ ）、“最小欧几里得距离分类器”进行样本分类实验，给出实验过程设计（包括源代码）和实验结果。
- 3) (20 分) 对每个数据集给出应用上述每种分类规则的分类决策界面和分类错误率，分析实验结果并给出结论。

问题二： (30 分)

在两类分类问题中，可以约束某个类别的错误率不变，即  $\epsilon_1 = \epsilon$ 。试证明最小化另一类别错误率得到的似然率测试规则为：

$$\text{decide } x \in \omega_1 \text{ if } \frac{P(\omega_1|x)}{P(\omega_2|x)} > \theta$$

其中，选择参数  $\theta$  来满足约束条件。这个规则称为 *Neyman-Pearson* 测试规则，它与贝叶斯最小风险规则相似。

提示：可以使用拉格朗日乘子法证明这个问题等价于最小化  $q = \theta(\epsilon_1 - \epsilon) + \epsilon_2$  的问题。