

研究生课模式识别考题，张长水的。

1. 一个两类分类问题如下：第一类的先验概率是 $2/3$ ，第二类的先验概率是 $1/3$ ，两类的概率密度函数为： p_1 是一个 $[1, 10]$ 的均匀分布， p_2 是一个 $[8, 12]$ 的均匀分布。请用贝叶斯决策方法按照最小错误率准则设计分类器。

2 已知一批样本为：0, 2, 3, 3, 3, 4, 6。

a. 如果这些样本来自正态分布，请用最大似然方法估计该分布的均值和方差。并请分析样本数的多少对所估计的结果的影响。

b. 如果不知道样本的分布，给出一种估计这些样本的概率密度函数的方法。详细说明估计方法、估计步骤，并给出估计结果。

3 已知 2 类样本如下，请用 Fisher 准则设计分类器，画出样本分布和投影直线，给出分界面方程和图形。简单说明类内协方差矩阵在 Fisher 判别中的作用。

$w_1: (1, 0)^T, (0, 1)^T$

$w_2: (3, 3)^T, (3, 5)^T, (5, 5)^T, (5, 3)^T$

4 有下列两类样本，请用 1 近邻法设计分类器，给出设计好的分类器，并画出样本分布和界面。

$w_1: (0, 0)^T, (1, 1)^T, (1, -1)^T$

$w_2: (1, 0)^T, (0, 2)^T, (0, -2)^T$

5 有下列样本。请用多级聚类方法对其聚类。类间距离采用最近距离。给出计算过程，画出多级聚类树，说明应该分成几类，并说明理由。用最近距离准则有什么缺点，请举例说明。

$(1, 0)^T, (0, 2)^T, (0, 4)^T, (2, 0)^T$

6 问答题

a. 利用贝叶斯理论设计分类器对于学习模式识别有什么意义？其缺点是什么？

b. 分析压缩近邻法的优点和缺点。

c. 在特征提取和特征选择时，应找什么特征？

d. 特征提取和特征选择的困难是什么？为什么说特征提取和特征选择是设计分类器的关键？

e. 人工神经网络理论可以用于模式识别吗？就你所了解给出说明。

7 由 N 个样本构成的集合 S 被划分成 c 个互不相交的子集 S_1, S_2, \dots, S_c ，如果 S_i 是空集，则该集合样本的均值 m_i 不定义。这种情况下，误差平方和仅仅和非空子集有关；

$J = \sum_i \sum_{x \in S_i} \|x - m_i\|^2$

其中， i 是不包含空子集的子集符号。假定 $N \geq c$ ，请证明使得 J 最小的划分中没有空子集。