

中国科学院自动化研究所

2005 年招收攻读博士学位研究生入学考试试题

模式识别

1. 简述模式的概念和它的直观特性，同时绘出模式识别系统的组成框图并说明各部分的主要功能特性。
2. 试阐述线性判别函数的基本概念，并说明既然有线性判别函数，为什么还需要非线性判别函数？假设有两类模式，每类包括 5 个 3 维不同的模式，且良好分布。如果它们是线性可分的，问权向量至少需要几个系数分量？假如要建立二次的多项式判别函数，又至少需要几个系数分量？（设模式的良好分布不因模式变化而改变）
3. 假定对一类特定人群进行某种疾病检查，正常人以 ω_1 类代表，患病者以 ω_2 类代表。设被检查的人中正常者和患病者的先验概率分别为

正常人： $P(\omega_1)=0.9$

患病者： $P(\omega_2)=0.1$

现有一被检查者，其观察值为 x ，从类条件概率密度分布曲线上查得

$$P(x|\omega_1)=0.2, P(x|\omega_2)=0.4$$

同时已知风险损失函数为

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 1 & 0 \end{pmatrix}$$

试对该被检查者用以下两种方法进行分类：

- (1) 基于最小错误率的贝叶斯决策，并写出其判别函数和决策面方程；
- (2) 基于最小风险的贝叶斯决策，并写出其判别函数和决策面方程。

4. 给定先验概率相等的两类, 其均值向量分别为: $\mu_1=[1,3,-1]^T$ 和 $\mu_2=[-1,-1,1]^T$, 协方差矩阵如下

$$\Sigma_1 = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

求用 $J_s = \frac{|S_w + S_b|}{|S_w|}$ 判据的最优特征选择。

5. 给出两类模式样本集

$$\omega_1: \{(-5 \ -5)^T, (-5 \ -4)^T, (-4 \ -5)^T, (-5 \ -6)^T, (-6 \ -5)^T\}$$

$$\omega_2: \{(5 \ 5)^T, (5 \ 6)^T, (6 \ 5)^T, (5 \ 4)^T, (4 \ 5)^T\}$$

设其先验概率相等, 即

$$P(\omega_1) = P(\omega_2) = 0.5$$

试用 K-L 变换把特征空间的维数降到一维, 并用图画出样本在特征空间中的位置。

6. 试阐述 ISODATA 算法的原理, 并用 ISODATA 算法对如下 8 个二维的模式样本进行聚类分析

$$x_1(0 \ 0), x_2(1 \ 1), x_3(2 \ 2), x_4(4 \ 3), x_5(5 \ 3), x_6(4 \ 4), x_7(5 \ 4), x_8(6 \ 5)$$

初始参数选取如下:

期望得到的聚类数 $K=2$

一个聚类中的最少样本数 $\theta_N=1$

标准偏差参数 $\theta_S=1$

合并参数 $\theta_C=4$

每次迭代允许合并的最大聚类对数 $L=0$

允许迭代的次数 $I=4$