

总复习题

一、填空题

1. 一个典型的模式识别系统主要由 () 等部分组成。
2. 后验概率是 ()。
3. 近邻法的基本思想是 ()。
4. 分级聚类算法的 2 种基本途径是 () 和 ()。
5. 特征抽取与特征选择的区别是 ()。
6. 最优搜索算法是指 ()。
7. 统计学习理论的核心问题是 ()。

$$P(x | \omega_1) = 0.1 \quad 0.6$$

$$P(x | \omega_2) = 0.8 \quad 0.3$$

状态 损失 决策	W_1	W_2
a_1	0.5	2
a_2	4	1.0

(1) 试用最小错误率贝叶斯准则判决 2 个样本各属于哪一类？

(2) 试用最小风险决策规则判决 2 个样本各属于哪一类？

二、问答题

1. 什么是模式与模式识别？
2. 确定线性分类器的主要步骤。
3. 描述贝叶斯公式及其主要作用。
4. 请详细写出感知器训练算法步骤。
5. 请详细写出 Fisher 算法实现步骤。
6. 什么是 K 近邻法？
7. 监督学习与非监督学习的区别？
8. 什么是误差平方和准则？
9. 什么是两分剪辑近邻法与压缩近邻法。
10. 请详细介绍初始聚类中心的选择方法。
11. 请描述 K 均值聚类算法。
12. 什么是离散 K-L 变换以及离散有限 K-L 展开。
13. 什么是支持向量机？

2. 已知两类的训练样本：

$w_1(0,0)^T, (0,2)^T; w_2(2,0)^T, (2,2)^T$ ，试用最小均方误差准则进行感知器分类器训练，求解向量 w^* 。

3. 已知欧氏二维空间中两类 9 个训练样本

$w_1: (-1,0)^T, (-2,0)^T, (-2,1)^T, (-2,-1)^T$

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

试分别用最近邻法和 K 近邻法求测试样本 $(0,0)^T$ 的分类，取 $K=5, 7$ 。

三、计算题

1. 在图像识别中，假定有灌木和坦克 2 种类型，它们的先验概率分别是 0.7 和 0.3，损失函数如下表所示。其中，类型 w_1 和 w_2 分别表示灌木和坦克，判决 $a_1=w_1, a_2=w_2$ 。现在做了 2 次实验，获得 2 个样本的类概率密度如下：

4. 已知两类的数据：

$w_1: (1,0), (2,0), (1,1)$

$w_2: (-1,0), (0,1), (-1,1)$

试求该组数据的类内与类间散布矩阵。

5. 给出二维样本数据 $(-1,1), (2,2), (1,-1), (-2,-2)$ ，试用 K-L 变换作一维数据压缩。

6. 假设有 20 个二维样本数据： $x_1(0,0)$

$x_2(1,0)$ $x_3(0,1)$ $x_4(1,1)$ $x_5(2,1)$ $x_6(1,2)$ $x_7(2,2)$ $x_8(3,2)$ $x_9(6,6)$ $x_{10}(7,6)$ $x_{11}(8,4)$ $x_{12}(4,7)$ $x_{13}(7,7)$ $x_{14}(8,7)$ $x_{15}(9,7)$ $x_{16}(7,8)$ $x_{17}(8,8)$ $x_{18}(9,8)$ $x_{19}(8,9)$ $x_{20}(9,9)$ 。

(1) 使用基于误差平方和准则函数的 C 均值算法 (即 K 均值聚类算法) 对这些数据进行聚类。假设选择 x_1 和 x_{20} 为初始聚类中心, 计算相应的误差平方和准则函数的值, 以及经 1 次迭代后误差平方和准则函数的值。

(2) 假设这些样本中 $x_1 \sim x_{10}$ 属于第一类, $x_{11} \sim x_{20}$ 属于第二类, 请问它们线性可分吗? 请用学习过的任一种线性分类器算法计算出分界面。

四、综述题

1. 针对某个识别对象 (如手写阿拉伯数字、人脸等等) 设计自己的模式识别系统, 并叙述各步骤主要工作。
2. 举出你在日常生活中遇到的一个模式识别应用系统的例子, 说明其基本构成、工作步骤, 并讨论其优缺点。