

模式识别练习题

一、简单题

1. 列举目前常用的分类算法。
2. 试说明以下问题求解是基于监督学习或是非监督学习：
 - (1) 求数据集的主分量
 - (2) 汉字识别
 - (3) 自组织特征映射
 - (4) CT 图像的分割
3. 怎么用先验概率和似然函数计算后验概率？公式是什么？
4. 贝叶斯决策为什么称为最小错误率贝叶斯决策？
5. 最小错误率贝叶斯分类器设计过程？
6. 写出多元正态概率下的最小错误率贝叶斯决策的判别函数。
7. 朴素贝叶斯分类器的优缺点？
8. 最小错误率贝叶斯决策规则有哪几种常用的表示形式？
9. 最大似然方法与 Bayes 估计的差异。
10. 阐述参数估计与非参数估计的区别，并分别列举参数估计与非参数估计两个常用的方法。
11. 对多元高斯分布，推导用最大似然估计方法估计均值和协方差时的公式，并且明确地给出可能需要的假设条件。
12. 写出非参数估计概率密度的基本公式，阐述 Parzen 窗法和 k_N 近邻法的区别。
13. 写出 Parzen 窗法几种常用的窗函数及对应公式。
14. 若用 Parzen 窗法估计模式的类概率密度函数，窗口尺寸 h_N 过小可能会带来什么问题？ h_N 过大可能会产生什么问题？

15. 为什么 PCA 能够移除特征之间的相关性?
16. PCA 和 LDA 的区别?
17. 利用两类别的线性分类器如何解决多类别的分类问题?
18. Fisher 准则的基本原理及准则函数?
19. LDA 算法的主要优点和缺点。

二、阐述题

1. 请论述模式识别系统的主要组成部分及其设计流程，并简述各组成部分中常用方法的主要思想。
2. 请简述 KNN 算法实现的步骤
3. 分析 KNN 算法的缺点并给出改进策略。
4. 简述贝叶斯判别法的基本思想
5. 怎样利用朴素贝叶斯方法获得各个属性的类条件概率分布?
6. 请简述一下 PCA 算法的主要步骤。
7. 请描述线性判别函数与非线性判别函数的差别? 非线性判别函数相较于线性判别函数有什么优势?
8. LDA 分类的目标与步骤
9. 线性分类器设计中的最小平方准则函数方法采用的准则函数公式是什么? 当利用伪逆解方法求解时，遇到计算量过大时，可以代替采用何种方法来训练分类器参数?

三、计算题

1. 假设在某地区切片细胞中正常(ω_1)和异常(ω_2)两类的先验概率分别为 $P(\omega_1)=0.9$, $P(\omega_2)=0.1$ 。现有一待识别细胞呈现出状态 x ，由其类条件概率密度

分布曲线查得 $P(x|\omega_1)=0.2$, $P(x|\omega_2)=0.4$, 试对细胞 x 进行分类。

2. 假设在某个地区细胞识别中正常 (w_1) 和异常 (w_2) 两类先验概率分别为 $P(w_1)=0.8$, $P(w_2)=0.2$, 现有一待识别的细胞, 其观察值为 x , 从类条件概率密度分布曲线上查得 $P(x/w_1)=0.25$, $P(x/w_2)=0.6$, 并且已知 $\lambda_{11}=0$, $\lambda_{12}=6$, $\lambda_{21}=1$, $\lambda_{22}=0$

试对该细胞 x 用一下两种方法进行分类:

(1) 基于最小错误率的贝叶斯决策;

(2) 基于最小风险的贝叶斯决策;

请分析两种结果的异同及原因

3. 在图像识别中, 假定有灌木丛和坦克两种类型, 分别用 ω_1 和 ω_2 表示, 它们的先验概率分别为 0.7 和 0.3, 损失函数如表所示。现在做了四次试验, 获得四个样本的类概率密度如下:

$$p(x/\omega_1): 0.1, 0.15, 0.3, 0.6$$

$$p(x/\omega_2): 0.8, 0.7, 0.55, 0.3$$

(1) 试用贝叶斯最小误判概率准则判决四个样本各属于哪个类型;

(2) 假定只考虑前两种判决, 试用贝叶斯最小风险准则判决四个样本各属于哪个类型;

表

<div> <div>类型</div> <div>损失</div> <div>判决</div> </div>	ω_1	ω_2
	0.5	2.0

α_2 (判为 ω_2)	4.0	1.0
α_3 (拒绝判决)	1.5	1.5

4. 一人在某实验室做了一次核辐射检测，结果是阳性的。当核辐射真正存在时，检测结果返回正确的阳性概率是99%；当核辐射不存在时，结果返回正确的阴性的概率是95%。而且，所有被测人群中有3%的人确实被辐射污染了。那么这个人被辐射污染了吗？

5. 利用乘法法则和全概率公式证明贝叶斯公式

6. 为了提高某产品的质量，企业决策人考虑增加投资来改进生产设备，预计需投资 90 万元。但从投资效果看，下属部门有两种意见：一是认为改进设备后高质量产品可占 90%；二是认为改进设备后高质量产品可占 70%。根据经验决策人认为第一种意见可信度有 40%，第二种意见可信度有 60%。为慎重起见，决策人先做了个小规模试验：试制了 5 个产品，结果全是高质量产品。问现在决策人对两种意见的可信程度有没有变化？

7. 假设正常细胞和癌细胞的样本的类条件概率服从单变量正态分布

$$p(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

使用最大似然估计方法求概率密度的参数，即

均值和方差。

8. 设一维两类模式满足正态分布，它们的均值和方差分别为， $\mu_1=0, \sigma_1=2, \mu_2=2, \sigma_2=2$ ， $p(x) \sim N(\mu, \sigma)$ ，先验概率 $P(\omega_1) = P(\omega_2)$ ，取0-1损失函数，试算出判决边界点，并绘出它们的概率密度函数曲线；试确定样本-3，-2，1，3，5各属哪一类。

9. 已知两个一维模式类别的类概率密度函数为：

$$p(x/\omega_1) = \begin{cases} x & 0 \leq x < 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{其它} \end{cases} \quad p(x/\omega_2) = \begin{cases} x-1 & 1 \leq x < 2 \\ 3-x & 2 \leq x \leq 3 \\ 0 & \text{其它} \end{cases}$$

先验概率 $P(\omega_1)=0.6, P(\omega_2)=0.4$,

(1) 求 0-1 代价 Bayes 判决函数;

(2) 判断样本 $\{x_1=1.35, x_2=1.45, x_3=1.55, x_4=1.65\}$ 各属于哪一类别。

10. 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。

11. 已知有两类数据, 分别为

$$\omega_1 : (1,0)(2,0)(1,1)$$

$$\omega_2 : (-1,0)(0,1)(-1,1)$$

试求: 该组数据的类内及类间离散矩阵 S_w 及 S_b 。

12. 设两类样本的类内离散矩阵分别是 $S_1 = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}, S_2 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$, 各类

样本均值分别为 $m_1 = (1,0)^T, m_2 = (3,2)^T$, 试用 Fisher 准则求其决策面方程, 并判断 $x = (2,2)^T$ 的类别。

13. 对于基于 $Y=a$ 的最小平方差方法, 如果我们有合适的 b 和判别函数, 那么最小平方差方法就会和 Fisher 判别方法等价。这么说对吗?

14. 已知两类训练样本: $w_1(0,0)^T, (0,2)^T; w_2(2,0)^T, (2,2)^T$, 试用最小平方差准则算法进行分类器训练, 求解向量与判决边界。

15. 设有两类的二维点: $w_1(1,2)^T, (2,0)^T; w_2(3,1)^T, (2,3)^T$ 试求伪逆解及判决边界。