一、考试大纲

1. 统计决策方法：
    1. 最小错误率贝叶斯决策
    2. 最小风险贝叶斯决策
    3. 两类错误率
    4. ROC 曲线
    5. 正态分布时的统计决策
    6. 错误率的计算
2. 概率密度函数估计
    1. 最大似然估计
    2. 贝叶斯估计与贝叶斯学习
    3. 概率密度估计的非参数方法
3. 线性分类器
    1. 线性判别分析的基本概念
    2. Fisher 线性判别分析
    3. 感知器
    4. 最小平方误差估计
    5. 多类线性分类器
4. 非线性分类器与神经网络
    1. 分段线性分类器
    2. 二次判别函数
    3. 多层感知器
5. 支持向量机与核方法
    1. 最优化分类超平面与线性支持向量机
    2. 核支持向量机
    3. 多类支持向量机
    4. 核 Fisher 判别分析
6. 其他分类方法
    1. 近邻法
    2. 决策树
    3. 随机森林基本概念
    4. 罗杰斯特 (Logistic) 回归基本概念
    5. Boosting 方法基本概念
7. 特征选择
    1. 第 7 章所有内容
8. 特征提取
    1. 第 8 单所有内容
9. 聚类
    1. 动态聚类算法
    2. 模糊取类方法

3. 分级聚类方法

二、样题

1、 简答题 (每题约 10 分)

1.1、简述模式识别系统的典型构成（P10）

答： 一个模式识别系统通常包括**原始数据的获取和预处理**、**特征提取与选择**、**分类或聚类**、**后处理**四个主要部分

1.2、简述在实际问题中，对样本 x 进行最小错误率风险贝叶斯决策的计算步骤（P15）

答：

1.3、简述在实际问题中，对样本 x 进行最小风险贝叶斯决策的计算步骤

答：

1.4、写出多元正态分布的概率密度函数

答:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$$

式中：

$\pmb x = [x_1, x_2, \dots, x_d]^T$ 是 d 维列向量；

$\pmb \mu = [\mu_1, \mu_2, \dots, \mu_d]^T$ 是 d 维均值向量；

$\pmb \Sigma$ 是 $d \times d$ 维协方差矩阵， $\pmb \Sigma^{-1}$ 是 $\pmb \Sigma$ 的逆矩阵， $|\pmb \Sigma|$ 是 $\pmb \Sigma$ 的行列式；

向量 $(\pmb x - \pmb \mu)^T$ 是向量 $(\pmb x - \pmb \mu)$ 的转置；

$\pmb \mu = E{\pmb x}$, 向量 $\pmb x$ 的期望；

$\pmb \Sigma = E{ (\pmb x - \pmb \mu)(\pmb x - \pmb \mu)^T}$, 矩阵 $(\pmb x - \pmb \mu)(\pmb x - \pmb \mu)^T$ 的期望。

1.5、写出多元正态分布的性质

答:

多元正态分布的性质:

1. 参数 $\pmb \mu$ 和 $ \pmb \Sigma $ 对分布的决定性。

   多元正态分布被均值向量 $\pmb \mu$ 和协方差 $\pmb \Sigma$ 所完全确定。
2. 等密度点的轨迹为一超椭球面。
3. 不相关性等价与独立性。
4. 边缘分布和条件分布的正态性。

   多元正态分布的边缘分布和条件分布仍然是正态分布。

5. 线性变换的正态性。

   多元正态随机向量的线性变换仍为多元正态分布的随机向量。

6. 线性组合的正态性。

   若 $\pmb x$ 为多元正态随机向量，则线性组合 $y = \pmb \alpha^T \pmb x$ 是一维的正态随机向量。其中， $\pmb \alpha$ 是与 $\pmb x$ 同维的向量。

1.6、什么是 ROC 曲线？针对两类问题，请描述 ROC 曲线的绘制步骤

答:

1.7、两类问题的似然比决策规则。两类问题中，两类的先验概率相等，每一类的类条件概率密度均为正态分布，已知每一类的均值和协矩阵，请写出负对数似然比。

答:

1.8、概率密度函数估计的方法有哪些？

答:

1. 最大似然估计
2. 贝叶斯估计
3. 概率密度估计的非参方法
   1. 直方图法
   2. k_N 近邻估计方法
   3. $Parzen$ 窗法

1.9、K-近邻的判别函数以及决策规则。

答: 设有 N 个已知样本分属于 c 个类 $w_i, i=1, \dots, c$ ，考查新样本 x 在这些样本中前 k 个近邻， 设其中有 k_i 个属于 w_i 类，则 w_i 类的判别函数就是

$$g_i(\boldsymbol{x}) = k_i, i = 1, \dots, c$$

决策规则是

$$\text{若 } g_k(\boldsymbol{x}) = \max_{i=1,\dots,c} g_i(\boldsymbol{x}), \text{则 } \boldsymbol{x} \in w_k$$

1.10、简述类间离散度以及类内离散度矩阵。

答:

1.11、类别可分性准则 J 应该满足的要求。

1.12、感知器的准则函数以及解的迭代公式。

1.13、压缩近邻法

1.14、IsoMap 方法

2、叙述题（每题约 15 分）

2.1、叙述 Fisther 线性判别分析（linear discriminant analysis, LDA）的主要计算步骤和分类决策规则。

2.2、叙述主成分分析（principal component analysis, PCA）。

2.3、叙述核主成分分析的工作原理。

2.4、针对多层前馈神经网络，给出反向传播算法的工作原理和训练步骤。

2.5、分析前馈神经网络中，隐含层数对分类预测可能产生的影响。

2.6、线性支持向量机的求解过程。

2.7、求最优变换 W，使得变换后的准则 $J_1(W) = tr(W^T(S_w+S_b)W)$ 最优。公式中相关符号的含义同参考书。$ (P162) $

解:

引入一个约束条件 $tr(W^TS_wW) = c$，设 $c=1$ 。则优化问题变为：

$$\max J_1(W)$$

$$s.t. \quad tr(W^T S_w W) = 1$$

采用拉格朗日方法将有约束优化问题变成无约束问题，拉格朗日函数是：

$$g(W) = J_1(W) - tr[\Lambda(W^T S_w W) - I]$$

其中，I 是单位矩阵， Λ 是对角阵，对角线元素是拉格朗日乘子。

在拉格朗日函数的极值点上，应该满足$ \frac{\partial}{\partial W}g(W) = 0$，

$$\therefore \quad S_w^{-1}(S_w + S_b)W = W\Lambda$$

$$\therefore \quad S_w^{-1}S_bW = W(\Lambda - I)$$

$$\Lambda = ?$$

$$\begin{aligned} & S_w^{-1}(S_w + S_b)W = W\Lambda \\ \Leftrightarrow \quad & S_wS_w^{-1}(S_w + S_b)W = S_wW\Lambda \\ \Leftrightarrow \quad & I(S_w + S_b)W = S_wW\Lambda \\ \Leftrightarrow \quad & (S_w + S_b)W = S_wW\Lambda \end{aligned}$$

$$\left.\begin{aligned} W^T(S_w + S_b)W \\ (S_w + S_b)W = S_wW\Lambda \end{aligned}\right\} \Rightarrow W^TS_wW\Lambda \xlongequal{W^TS_wW=1} \Lambda$$

$$\therefore \quad J_1(W) = tr(W^T(S_w + S_b)W) = tr(W^TS_wW\Lambda) = tr(\Lambda)$$

对于 $D \times d$ 维的变换矩阵，

$$J_1(W) = \sum_{i=1}^{d}(1 + \lambda_i)$$

$\therefore$ 最优变换阵 W 就是由 $S_w^{-1}S_b$ 的前 d 个本征值所对应的本征向量组面

2.8、叙述一个你所熟悉的模式识别的典型应用，例如：联机手写汉字识别。从已知条件、需要解决的问题、信息获取与预处理、特征提取、分类器设计（或聚类）、分类决策（或结果）等方面叙述。