

# 807 信息技术基础综合

## 考试大纲

### 一、考试性质

《信息技术基础综合》是 2022 年清华大学深圳国际研究生院电子信息专业【人工智能】方向全国硕士生统一入学考试专业课考试科目。【人工智能】项目是一个基于自动化、计算机以及电子通信等一级学科基础上设立的一个多学科交叉工程硕士项目，本专业课考试科目力求能够科学、公平、准确、规范地测评考生在信息技术领域，特别是人工智能相关理论及技术上所具备的基础知识、核心技能、自主创新等方面的综合能力，选拔具有较强科研能力、知识技能创新能力和发展潜质的优秀考生入学。

### 二、考试要求

测试考生对信息技术领域特别是人工智能方向相关的基本概念、基础理论与核心技能的掌握和运用能力。

### 三、考试方式与分值

满分 150 分，题型包括：填空题、是非判断题、选择题，名词解释，计算题等。

### 四、参考书目：

《模式识别》 张学工编著 清华大学出版社，第三版（2010 年 3 月）

### 五、考试内容

#### 1 模式识别基础

- 1.1 模式与模式识别的概念
- 1.2 模式识别的主要方法
- 1.3 监督模式识别与非监督模式识别
- 1.4 模式识别系统举例
- 1.5 模式识别系统的典型构成

#### 2 统计决策方法

- 2.1 统计决策方法的基本概念
- 2.2 最小错误率贝叶斯决策
- 2.3 最小风险贝叶斯决策

- 2.4 两类错误率、Neyman-Pearson 决策与 ROC 曲线
- 2.5 正态分布时的统计决策
  - 2.5.1 正态分布及其性质回顾
  - 2.5.2 正态分布概率模型下的最小错误率贝叶斯决策
- 2.6 错误率的计算
  - 2.6.1 正态分布且各类协方差矩阵相等情况下错误率的计算
  - 2.6.2 高维独立随机变量时错误率的估计
- 2.7 离散概率模型下的统计决策

### 3 概率密度函数估计

- 3.1 概率密度函数的估计的基本概念
- 3.2 最大似然估计
  - 3.2.1 最大似然估计的基本原理
  - 3.2.2 最大似然估计的求解
  - 3.2.3 正态分布下的最大似然估计
- 3.3 贝叶斯估计与贝叶斯学习
  - 3.3.1 贝叶斯估计
  - 3.3.2 贝叶斯学习
  - 3.3.3 正态分布时的贝叶斯估计
  - 3.3.4 其他分布的情况
- 3.4 概率密度估计的非参数方法
  - 3.4.1 非参数估计的基本原理与直方图方法
  - 3.4.2 k 近邻估计方法
  - 3.4.3 Parzen 窗法

### 4 线性分类器

- 4.1 线性判别函数
- 4.2 Fisher 线性判别分析
- 4.3 感知器
- 4.4 最小平方误差判别
- 4.5 最优分类超平面与线性支持向量机

4.5.1 最优分类超平面

4.5.2 大间隔与推广能力

4.5.3 线性不可分情况

4.6 多线性分类器

4.6.1 多个两类分类器的组合

4.6.2 多类线性判别函数

## 5 非线性分类器

5.1 分段线性判别函数

5.1.1 分段线性距离分类器

5.1.2 一般分段线性判别函数

5.2 二次判别函数

5.3 多层感知器神经网络

5.3.1 神经元与感知器

5.3.2 多个感知器组合

5.3.3 基于反向传播算法的多层感知器

5.4 支持向量机

5.4.1 广义线性判别函数

5.4.2 核函数变换与支持向量机

5.4.3 多类支持向量机

5.4.4 用于函数拟合的支持向量机

5.5 核函数机器

5.5.1 大间隔机器与核函数机器

5.5.2 核 Fisher 判别

## 6 其他分类方法

6.1 近邻法

6.1.1 最近邻法

6.1.2 k-近邻法

6.1.3 近邻法的快速算法

6.1.4 剪辑近邻法

6.1.5 压缩近邻法

6.2 决策树与随机森林

6.2.1 非数值特征

6.2.2 决策树

6.2.3 过学习与决策树的剪枝

6.2.4 随机森林

6.3 Logistic 回归

6.4 Boosting 方法

## 7 特征选择

7.1 特征的评价准则

7.1.1 基于类内类间距离的可分性判据

7.1.2 基于概率分布的可分性判据

7.1.3 基于熵的可分性判据

7.1.4 利用统计检验作为可分性判据

7.2 特征选择的最优算法

7.3 特征选择的次优算法

7.4 特征选择的遗传算法

7.5 以分类性能为准则的特征选择方法

## 8 特征提取

8.1 基于类别可分性判据的特征提取

8.2 主成分分析方法

8.3 Karhunen-Loève 变换

8.3.1 K-L 变换的基本原理

8.3.2 用于监督模式识别的 K-L 变换

8.4 K-L 变换在人脸识别中的应用举例

8.5 高维数据的低维显示

8.6 多维尺度法

8.6.1 MDS 的基本概念

8.6.2 古典尺度法

- 8.6.3 度量型 MDS
- 8.6.4 非度量型 MDS
- 8.6.5 MDS 在模式识别中的引用
- 8.7 非线性变换方法简介
  - 8.7.1 核主成分分析 (KPCA)
  - 8.7.2 IsoMap 方法和 LLE 方法

## 9 非监督模式识别

- 9.1 基于模型的方法
- 9.2 混合模型的估计
  - 9.2.1 非监督最大似然估计
  - 9.2.2 正态分布情况下的非监督参数估计
- 9.3 动态聚类算法
  - 9.3.1 C 均值算法
  - 9.3.2 ISODATA 方法
  - 9.3.3 基于样本与核的相似性度量的动态聚类算法
- 9.4 模糊聚类算法
  - 9.4.1 模糊集的基本知识
  - 9.4.2 模糊 C 均值算法
  - 9.4.3 改进的模糊 C 均值算法
- 9.5 分级聚类方法
- 9.6 自组织映射神经网络
  - 9.6.1 SOM 网络结构
  - 9.6.2 SOM 学习算法和自组织特性
  - 9.6.3 SOM 用于模式识别

## 10 模式识别系统的评价

- 10.1 监督模式识别方法的错误率估计
  - 10.1.1 训练错误率
  - 10.1.2 测试错误率
  - 10.1.3 交叉验证

- 10.1.4 自举法与.632 估计
- 10.2 有限样本下错误率的区间估计问题
  - 10.2.1 问题的提出
  - 10.2.2 用扰动重采样估计 SVM 错误率的置信区间
- 10.3 特征提取与选择对分类器性能估计的影响
- 10.4 从分类的显著性推断特征与类别的关系
- 10.5 非监督模式识别系统性能的评价