

## 876 《人工智能基础综合》考试大纲

注：《数据结构》占 60%，《机器学习》占 40%。

### 《数据结构》部分：

#### 一、考试题型

- 1、填空题
- 2、名词解释
- 3、简答题
- 4、分析应用题
- 5、算法设计题

#### 二、考试参考用书

《数据结构（C 语言版第 2 版）》，严蔚敏、李冬梅、吴伟民，清华大学出版社，人民邮电出版社，2022 年 1 月第 2 版

#### 三、考试内容

##### 第一章 绪论

了解：数据类型和抽象数据类型的概念；

熟悉：数据的逻辑结构、存储结构和运算之间的关系，时间复杂度和空间复杂度的概念和分析方法；

掌握：数据结构的基本概念和相关术语、数据的逻辑结构和存储结构的分类。

##### 第二章 线性表

了解：线性表的定义和基本操作、链表的概念；

熟悉：顺序表、单链表、循环链表和双向链表的概念，顺序表的应用；

掌握：线性表的顺序存储结构及链式存储结构的定义，查找、插入和删除等基本操作的实现。

##### 第三章 栈和队列

了解：递归的执行过程；

熟悉：栈和队列的顺序存储结构和链式存储结构实现，循环队列的设计，递归的概念；

掌握：栈和队列的概念、结构特性、典型应用。

#### 第四章 数组和广义表

了解：特殊矩阵和稀疏矩阵的压缩存储方法、广义表的存储结构；

熟悉：数组的概念和数组的存储结构；

掌握：多维数组元素在顺序存储结构中的存储地址的计算。

#### 第五章 树和二叉树

了解：树的定义、表示方法和基本操作；

熟悉：树的相关术语，二叉树的概念、存储结构和基本操作，二叉树遍历的定义以及先序和中序遍历的非递归算法，线索二叉树的概念，树和森林的遍历，哈夫曼树和哈夫曼编码的概念；

掌握：二叉树的定义、性质、二叉链表存储结构，满二叉树和完全二叉树的概念，二叉树遍历的递归算法，二叉树线索化的方法，树与二叉树的转换方法，哈夫曼树的构造方法。

#### 第六章 图

了解：图的定义和基本操作，无向图的连通分量和生成树的求法，拓扑排序的算法、关键路径的算法，任意一对顶点间的最短路径问题；

熟悉：拓扑排序的概念，最小生成树的概念，单源最短路径问题的算法；

掌握：图的基本概念、邻接矩阵和邻接表存储结构、深度和广度优先遍历、普里姆算法和克鲁斯卡尔算法求最小生成树的方法、拓扑排序算法的基本思想、关键路径的求法、用 Dijkstra 算法求解单源最短路径问题的过程。

#### 第七章 查找

了解：查找的基本概念、平衡二叉树的平衡化方法、B 树的概念、哈希表构建和查找的算法；

熟悉：分块查找的基本思想、哈希表的基本概念、哈希函数的构造方法和处理冲突的方法；

掌握：顺序查找和折半查找的基本思想和算法、二叉排序树的概念和基本操作、哈希表的构造方法。

## 第八章 内部排序

了解：排序的基本概念，希尔排序、二路归并排序、堆排序和基数排序的算法；

熟悉：理解各种内部排序方法的基本思想、排序过程和特点，快速排序中一次划分的算法，基数排序的思想；

掌握：堆的概念，直接插入排序、简单选择排序、冒泡排序、快速排序和二路归并排序的过程，直接插入排序、简单选择排序和冒泡排序的算法，筛选法建堆的过程，各种内部排序方法的比较。

## 《机器学习》部分：

### 一、考试题型

- 1、填空题
- 2、名词解释
- 3、简答题
- 4、计算题

### 二、考试参考用书

《机器学习》，周志华著，清华大学出版社，2016年1月第1版

### 三、考试内容

#### 第1章 绪论

熟悉：训练、测试的概念，分类与回归的概念及其区别，监督学习与非监督学习的区别及其代表方法。

掌握：训练集、测试集、训练样本、测试样本的定义，泛化能力的概念。

#### 第2章 模型评估与选择

了解：分类与回归常用的性能度量，P-R 曲线与 ROC 曲线的绘制原理及其比较方法，偏差与方差的定义及其在泛化误差分解中的角色。

熟悉：均方误差、错误率、精度、查准率、查全率、F1 的计算方法。

掌握：误差、训练误差、泛化误差的定义，过拟合与欠拟合的基本概念、模型表现及其产生原因，留出法、交叉验证法、自助法的基本过程及其优缺点。

### 第 3 章 线性模型

了解：类别不平衡问题的处理方法，多分类任务中的拆分策略及其特点；

熟悉：对数几率回归模型的表达式，几率的定义；

掌握：线性模型的基本形式，线性回归的最小二乘法的原理与参数估计过程，线性判别分析的思想、目标函数以及解的推导过程。

### 第 4 章 决策树

了解：连续属性值的二分处理方法，缺失值的处理机制；

熟悉：决策树的各结点对应什么，ID3、C4.5、CART 方法选择划分属性的依据，预剪枝与后剪枝的基本原理，剪枝处理对决策树训练时间、测试时间以及泛化性能的影响；

掌握：信息熵、信息增益、增益率、基尼指数的定义及计算方法，ID3 方法生成决策树。

### 第 5 章 神经网络

了解：全局最小与局部极小的概念及跳出局部极小的常用策略；

熟悉：神经网络中缓解过拟合的策略，激活函数的作用以及典型的激活函数；

掌握：神经元、感知机、多层前馈神经网络的结构及数学模型，误差逆传播算法的推导过程、更新公式及算法流程。

### 第 6 章 支持向量机

了解：支持向量的稀疏性，核函数解决非线性可分问题的原理，支持向量回归的基本原理；

熟悉：核函数的基本概念及常用核函数，软间隔的基本思想，软间隔支持向量机与硬间隔的区别；

掌握：支持向量与间隔的定义，支持向量机基本型及最大间隔原理，对偶问题的推导，支持向量机  $w$  和  $b$  的计算。

## 第 7 章 贝叶斯分类

了解：朴素贝叶斯分类器的属性条件独立性假设；

熟悉：先验概率、后验概率、类条件概率的基本概念，极大似然估计的原理；

掌握：贝叶斯决策论分类，极大似然估计估计参数，朴素贝叶斯分类器分类。

## 第 8 章 集成学习

了解：Adaboost 的算法流程，各种结合策略的基本原理及适用场景；

熟悉：集成学习的概念，好的集成对个体学习器的要求，随机森林算法，多样性增强的常见方法；

掌握：Boosting 算法的工作机制，Bagging 算法的工作机制，Boosting 与 Bagging 在偏差-方差分解上的区别。

## 第 9 章 聚类

了解：密度聚类的基本思想，层次聚类的基本思想；

熟悉：有序属性、无序属性的概念及距离计算中的处理方法，原型聚类的基本思想及其常见算法，高斯混合聚类的概率模型表达及 EM 算法求解过程；

掌握：聚类的概念， $k$  均值算法的目标函数以及聚类过程，学习向量量化(LVQ)的算法流程及原型向量更新规则。

## 第 10 章 降维与度量学习

了解：多维缩放 (MDS)、等度量映射、局部线性嵌入的基本思想；

熟悉：维数灾难的概念，降维的目的，主成分分析算法原理；

掌握： $k$  近邻分类，主成分分析算法流程。

## 第 11 章 特征选择与稀疏学习

了解：子集搜索的基本策略，子集评价的基本方法；

熟悉：特征选择的基本概念及其与降维的区别与联系，字典学习与稀疏编码的基本概念；

掌握：三种特征选择方法如何进行特征选择，L1 正则化的稀疏性。