

山东大学高等教育自学考试毕业及实践环节考核大纲

课程名称：人工智能和大数据（实践）

课程代码：13011

“人工智能和大数据（实践）”是计算机专业的实践类课程，是培养考生理论联系实际，综合应用所学知识解决人工智能和大数据应用场景中具体问题的重要教学环节。为确保该课程教学大纲的顺利实施及人才培养目标的实现，根据计算机（本科）教学计划，制定考核大纲：

一、课程目标

理解大数据基本概念与原理，在熟练掌握 Python 语言基本语法的基础上，能够利用 pandas 进行数据处理，利用 matplotlib 进行数据的可视化操作，理解回归，聚类分析，分类等基本算法的基本思想和原理，能够利用人工智能的第三方模块运用这些算法解决一些实际问题。

二、考核要求

第一章 数据与大数据导论

- 1 理解大数据的基本概念，采集和预处理过程，存储
- 2 了解云计算与大数据并行计算的概念
- 3 了解大数据分析的概念，方法，算法，可视化

第二章 Python 基础

- 1 熟悉 Anaconda 的安装，熟练掌握 Jupyter Notebook 的使用
- 2 掌握 Python 语言的数据类型、变量，运算符和注释
- 3 掌握列表，元组，字典的使用
- 4 了解结构化程序设计方法，重点掌握顺序结构、选择结构、循环结构等程序设计的使用方法，了解三种结构化方法的不同点
- 5 熟悉函数的使用，特别是参数的传递方式
- 6 熟悉模块导入的方式，能够区分 import 和 from ..import 的使用方式
- 7 熟悉文件的操作，了解 csv 和 json 文件的特点

第三章 Pandas 数据处理和分析

- 1 了解 Pandas 的 Series 和 DataFrame 的特点，熟练掌握其基本操作
- 2 熟悉用 Pandas 读写 CSV, EXCEL 文件
- 3 熟悉 Pandas 操作数据的各种功能：数据清洗，数据操作，索引，合并，分组，变形

第四章 Matplotlib 数据可视化

- 1 熟悉 matplotlib 绘制折线图的基本操作
- 2 熟悉进行图表装饰的方法
- 3 熟悉绘制各种图表形式：条形图，饼图，散点图，泡图，雷达图等

第五章 回归模型

- 1 了解线性和非线性回归模型的原理
- 2 掌握利用 statsmodels 库建立回归方程的方法（能建立方程即可，不用进行后续的分析）

第六章 聚类分析

- 1 了解聚类分析的基本概念，理解相似度计算方法
- 2 了解划分，层次，密度聚类的基本原理
- 3 掌握用 scikit-learn 和 matplotlib 对数据采用以下聚类算法并展示的方法：k-means，层次聚类，dbscan
- 4 熟悉利用各种指标进行聚类算法的评价分析
- 5 能灵活运用聚类算法对应用数据进行聚类分析

第七章 分类模型

- 1 了解决策树，朴素贝叶斯，支持向量机进行分类的基本原理
- 2 掌握用 scikit-learn 和 matplotlib 对数据采用以下分类算法的方法：决策树，朴素贝叶斯，支持向量机

三、考核方式

现场上机考试，根据题目要求，写出完整程序或者对给出程序进行修改，运行结果。

四、参考教材

《大数据基础与 Python 机器学习（大数据系列丛书）》，2022 年 5 月第一版，高静，申志军，姜新华，陈俊杰编，清华大学出版社