

上海市高等学校计算机等级考试（三级）

《数据科学与人工智能技术》考试大纲（试行）

（2019 年版）

一、考试性质

上海市高等学校计算机等级考试是上海市教育委员会组织的全市高校统一的教学考试，是检测和评价高校计算机基础教学水平和教学质量的重要依据之一。该项考试旨在规范和加强上海高校的计算机基础教学工作，提高学生的计算机应用能力。考试对象主要是上海市高等学校学生，每年举行一次，通常安排在当年的十月下旬、十一月上旬的星期六或星期日。凡考试成绩达到合格者或优秀者，由上海市教育委员会颁发相应的证书。

本考试由上海市教育委员会统一领导，聘请有关专家组成考试委员会，委托上海市教育考试院组织实施。

二、考试目标

上海市高等学校计算机等级考试（三级）主要考核学生的计算机综合应用能力。

“数据科学与人工智能技术”考核学生对数据科学工作流程的理解，应用探索性分析、可视化分析、建模分析等方法对数据进行处理的综合能力；考核学生对人工智能常用方法的基本原理和特点的理解，结合领域需求应用人工智能技术构建智能应用系统的综合能力。考试内容涵盖从相关理论知识到基本方法的应用实践，要求学生具有应用数据分析和人工智能方法解决实际问题、提出解决方案和决策建议的能力。

三、考试细则

1. 考试时间：150 分钟。
2. 考试方式：考试采用基于网络环境的无纸化上机考试。
3. 考试环境：
 - 1) 上海市高校计算机等级考试通用平台。
 - 2) 可选开发语言：Python、MATLAB、R。
具体软件及版本环境：
 - Anaconda 3.5.1 以上（Python 3.X）并安装 Keras
 - MATLAB 2017b 以上
 - RStudio 1.2（R3.2.5 以上）并安装 Keras

四、试卷结构

试卷总分为 100 分，由基础知识方法以及综合应用实践两部分组成。

基础知识方法：60 分，考生需完成一个可选模块和一个必选模块。可选模块一是数据分析基础模块，题型为程序填空题；可选模块二是人工智能基础模块，题型为选择题和填空题；必选模块题型为选择题。

综合应用实践：40 分，分析和解决 1~2 个实际应用问题，题型包括论述题和编程题。

考题类别		内容	题型	分值
基础知识方法	可选模块一 数据分析基础	多维数据结构与运算	程序填空题	12
		数据统计与可视化数据探索	程序填空题	38
	可选模块二 人工智能基础	人工智能概述及传统方法	选择题 填空题	30
		机器学习常用算法	选择题 填空题	20
必选模块 常用信息处理方法	文本理解、图像理解、声音识别等，AI 服务应用	选择题	10	
综合应用实践	必选模块 机器学习建模分析	根据问题目标选择合适的数据分析及人工智能方法，提出分析处理过程或智能算法；或根据给定数据集，提出分析目标和给出实现方法	论述题 (方法建模)	5
		编写程序，分析和处理数据集合，应用机器学习常用算法解决实际问题	编程题 (应用实现)	30
		对算法实现结果和应用进行分析说明	论述题 (结果分析)	5
总分				100

五、考试内容和要求

(一) 考试内容

数据分析基础（可选模块一）

1. 数据科学概述

- 数据科学的工作流程
- 大数据技术
- 数据分析工具

2. 多维数据结构与运算

1) 多维数组

- 创建多维数组对象
- 数组元素切片、筛选

2) 多维数组运算

- 基本算术运算、函数与矩阵运算
- 随机生成函数

3. 数据汇总与统计

1) 数据文件读写

- CSV、txt、Excel 文件的读写

2) 数据清洗和规整化

- 数据筛选
- 缺失、重复数据处理
- 数据合并
- 数据排序

3) 统计分析

- 统计的基本概念、常用统计量的意义
- 常用统计函数

4. 可视化数据探索

1) 常用可视化分析图形

- 散点图、柱形图、饼图、直方图、概率密度图、箱形图、折线图、半对数图

2) 绘图

- 创建图形对象、绘制图形
- 绘制子图
- 图元设置
- 图形文件保存

人工智能基础（可选模块二）

1. 人工智能概述

- 人工智能的发展历史
- 人工智能的典型任务
- 人工智能的技术体系

2. 基于知识推理的智能系统

- 专家系统
- 知识表示：产生式、语义网络
- 知识图谱

3. 基于搜索的智能系统

- 启发式搜索：启发函数、贪婪最佳优先搜索、A*搜索
- 对抗搜索：估值决策、最小最大值搜索、Alpha-Beta 剪枝搜索
- 蒙特卡洛树搜索

4. 基于仿生算法的智能系统

- 常用仿生算法
- 基因遗传算法：算法原理、算法流程
- 群智能算法：粒子群算法、蚁群算法

5. 机器学习常用算法

- 内容参见机器学习建模分析模块

机器学习建模分析（必选模块）

1. 机器学习基础知识

- 有监督学习、无监督学习
- 训练集、测试集、划分方法
- 机器学习性能评估

2. 有监督学习

- 回归分析任务，线性回归分析
- 分类分析任务，决策树、朴素贝叶斯、支持向量机

3. 聚类分析

- 聚类分析目标、聚类方法性能评估
- K-Means 聚类、层次聚类

4. 神经网络和深度学习

- 感知器、前馈神经网络结构
- 神经网络分类
- 深度学习基础知识
- 深度学习框架和工具

4. 领域应用实现

- 社会各行业和领域数据的建模分析

常用信息处理方法（必选模块）

1. 文本理解

- 自然语言处理的基本知识
- 中文文本处理步骤：分词、词性标注、特征提取、语言表示模型
- 文本分类

2. 图像理解

- 数字图像表示方法
- 图像数据存储
- 图像分类和识别

3. 语音识别

- 语音数据表示
- 语音识别的基本方法

4. 时序数据分析

- 时间序列数据的特征提取
- 时间序列数据的分析方法：ARMA 和 ARIMA

5. AI 服务应用

- 其他常用信息的处理方法及应用
- 理解和调用 AI 开放平台提供的相关服务实现智能应用，包括但不限于图像、语音、文本、知识图谱等

（二）考试能力要求

1. 掌握数据科学、人工智能的基本概念、工作流程和应用特点，具有将数据科学和人工智能方法应用于解决实际问题的意识。

2. 掌握数据科学中问题提出、数据收集、数据探索、数据建模的基本方法，具备利用数据分析方法和工具进行数据收集和探索的能力。

3. 掌握人工智能常用方法的基本原理和应用特点，具备结合领域知识选择合适方法对任务进行数据建模、分析和给出解决方案的能力。

4. 掌握一种程序设计语言和相关的数据库与人工智能方法库，具备采用编程语言实现数据科学与人工智能方法解决实际问题的能力，具有对算法的实现结果进行分析解释的能力。

5. 了解文本、图像、语音、时序数据等常用信息的处理方法和常用技术，具有提出应用场景需求，调用 AI 开放平台提供的相关服务实现有关智能应用的能力。

六、说明

1. 建议学时数：48 学时。

2. 参考教材：

1) 《数据科学技术与应用》（宋晖、刘晓强主编），电子工业出版社，2018 年。

2) 《人工智能导论》（王万良主编），高等教育出版社，2017 年。

3. 先修课程：任意一种高级程序设计语言。