**课程教学大纲-数据科学与工程算法**

|  |  |  |  |
| --- | --- | --- | --- |
| **课程代码** | **DATA0031121011.02** | **课程性质** | **专业必修** |
| **课程名称：** | **数据科学与工程算法** | | |
| **英文名称** | **Algorithms for Data Science and Engineering** | | |
| **学时/学分** | **3.0/3.0** | **其中实验/实践学时** | **1.0** |
| **开课单位** | **数据科学与工程学院** | **适用专业：** | **数据科学与大数据技术** |
| **先修课程** | **数据结构、算法基础、概率论、数理统计、数据科学与工程数学基础** | | |
| **大纲撰写人** | **王延昊** | **大纲审核人** | **高明** |
| **课程网址** | **无** | **授课语言** | **中文** |

注：课程性质选择下列类别之一：学科基础、大类平台、专业必修、专业选修、教师教育

一、课程说明

算法是数据科学与工程的重要工具之一。算法的使用、分析、设计是数据采集、数据管理、数据分析挖掘和数据可视化的基础。算法知识的学习对后续数据科学与工程相关知识的学习至关重要。本课程从数据采集、数据管理、数据分析挖掘和数据可视化整个数据的全生命周期设计开展数据科学与工程相关应用所需要的算法基础出发，从概率与统计算法、线性代数和组合优化算法等知识入手，系统讲述数据科学与工程算法设计和分析的基础理论和方法，并介绍近似算法、采样算法和随机算法等高级算法知识，培养学生使用经典算法求解问题的能力，以及算法设计、分析和优化的能力。

二、课程目标

目标1：了解数据分析处理阶段和算法设计原则（支持毕业要求1）

目标2：掌握基本的数据采样算法（支持毕业要求2）

目标3：掌握尾概率不等式及其应用（支持毕业要求2）

目标4：掌握哈希、布隆过滤器和局部敏感哈希算法（支持毕业要求2）

目标5：了解Misra Gries算法；掌握Count Sketch和CountMin Sketch算法（支持毕业要求2）

目标6：掌握马尔可夫过程和PageRank算法（支持毕业要求2）

目标7：掌握特征值计算的方法（支持毕业要求3）

目标8：掌握SVD和PCA技术（支持毕业要求3）

目标9：掌握梯度下降算法；了解矩阵分解技术（支持毕业要求3）

目标10：掌握整数规划的概念；了解整数规划算法（支持毕业要求4）

目标11：掌握子模函数的概念和爬山算法（支持毕业要求4）

目标12：了解图模型、模块度的概念和社区发现方法（支持毕业要求4）

二、课程目标与毕业要求的对应关系

|  |  |  |
| --- | --- | --- |
| **毕业要求** | **指标点** | **课程目标** |
| 了解数据科学算法的基本目标 | 了解数据分析处理阶段 | 目标1 |
| 了解算法设计原则 |
| 掌握概率统计算法及其在数据科学中的应用 | 掌握基本的数据采样算法 | 目标2，3，4，5，6 |
| 掌握尾概率不等式 |
| 掌握基本的哈希算法 |
| 掌握数据流模型与算法 |
| 掌握随机过程与随机算法 |
| 掌握线性代数与优化算法及其在数据科学中的应用 | 掌握矩阵的特征值与特征向量计算方法 | 目标7，8，9 |
| 掌握SVD和PCA技术 |
| 掌握梯度下降算法 |
| 了解矩阵因式分解 |
| 掌握组合优化算法及其在数据科学中的应用 | 掌握整数规划及其算法 | 目标10，11，12 |
| 掌握子模函数定义与优化 |
| 了解图模型 |
| 了解图模块度概念与社区发现算法 |

三、教学内容与学时安排

（重难点请标注🞛）

第一章 绪论（支撑课程目标1）

学时：4

1.数据分析处理阶段

2.算法设计原则

要求学生：完成课程目标1

第二章 数据采样算法（支撑课程目标2）

学时：4

1.数据抽样基本概念

2.系统抽样

3.分层抽样

4.水库抽样

要求学生：完成课程目标2；完成习题2

第三章 尾概率不等式及其应用🞛（支撑课程目标3）

学时：4

1.Markov不等式

2.Chebyshev不等式

3.Chernoff不等式

4.尾概率不等式的应用 - Morris算法

要求学生：完成课程目标3；完成习题3

第四章 哈希技术🞛（支撑课程目标4）

学时：4+8

1.哈希的基本概念

2.布隆过滤器

3.局部敏感哈希

要求学生：完成课程目标4；完成习题4；完成课程项目1《基于局部敏感哈希的图节点相似性查询系统》

第五章 数据流模型及频繁项挖掘🞛（支撑课程目标5）

学时：4

1.数据流模型

2.频繁项挖掘

3.Misra Gries算法

4.Count Sketch算法

5.CountMin Sketch算法

要求学生：完成课程目标5；完成习题5

第六章 随机游走及其应用（支撑课程目标6）

学时：4

1.随机过程

2.PageRank算法

要求学生：完成课程目标6；完成习题6

第七章 矩阵特征值计算🞛（支撑课程目标7）

学时：4

1.方阵的特征值与特征向量

2.幂法

3.反幂法

4.瑞利商迭代法

要求学生：完成课程目标7；完成习题7

第八章 奇异值分解与主成分分析🞛（支撑课程目标8）

学时：4+8

1.对角化与奇异值分解

2.奇异值分解

3.主成分分析

要求学生：完成课程目标8；完成习题8；完成课程项目2《基于主成分分析的图像压缩技术实现》

第九章 矩阵分解（支撑课程目标9）

学时：4+8

1.梯度下降算法

2.基于梯度下降的矩阵分解

3.非负矩阵分解

要求学生：完成课程目标9；完成习题9；完成课程项目3《基于矩阵分解的推荐系统实现》

第十章 整数规划（支撑课程目标10）

学时：4

1.整数规划的定义

2.分支定界法

3.割平面法

要求学生：完成课程目标10；完成习题10

第十一章 子模函数及其应用（支撑课程目标11）

学时：4

1.子模函数及其性质

2.集合覆盖

3.爬山算法

要求学生：完成课程目标11；完成习题11

第十二章 模块度与社区发现（支撑课程目标12）

学时：4

1.图模型

2.模块度

3.社区发现

要求学生：完成课程目标12；完成习题12

四、教学方法

1、理论课教学：本课程每周2课时理论课教学，对每个教学目标所对应的课程知识进行讲授。

2、习题课：本课程共开设4次习题课，对课后习题进行讲解，巩固课程知识，培养学生的知识运用能力。

3、课程项目：本课程共开设3个课程项目，共24课时。课程项目是课程知识的延伸，培养学生在实践中运用课程知识解决实际问题的能力。

五、考核方式

本课程的考核方式如下：

1. **考勤（5分）：**随堂点名，综合整个学期表现赋分；
2. **作业（5分）：**学生上交课后作业后由助教批改，综合整个学期表现赋分；
3. **项目1（10分）：**学生完成课程项目1并提交项目报告和实现代码由助教评价并赋分；
4. **项目2（10分）：**学生完成课程项目2并提交项目报告和实现代码由助教评价并赋分；
5. **项目3（10分）：**学生完成课程项目3并提交项目报告和实现代码由助教评价并赋分；
6. **期中考试（10分）：**在第10周随堂考试，考察内容为课程目标1-6，由任课老师出卷、阅卷并赋分；
7. **期末考试（50分）：**由学校统一安排在考试周，考察内容为课程目标1-12，，由任课老师按照学校统一安排出卷、阅卷并赋分。

表-课程目标与考核方式对应关系

| **考核方式**  **课程目标** | **考勤** | **作业** | **项目1** | **项目2** | **项目3** | **期中考试** | **期末考试** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **课程目标1** | √ | √ |  |  |  | √ | √ |
| **课程目标2** | √ | √ | √ |  |  | √ | √ |
| **课程目标3** | √ | √ |  | √ | √ |  | √ |
| **课程目标4** | √ | √ |  |  |  |  | √ |

六、推荐教材和参考资料

**1.推荐教材：**《数据科学与工程算法基础》，高明、胡卉芪，高等教育出版社，2021年，第1版，ISBN 978-7-04-055336-9

## 八、评分标准【请按照本门课程采用的课程考核方式选择下表之一填写】（具体分段可以根据实际情况调整）

表1：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **课程目标** | **评分标准** | | | | |
| **90-100** | **80-89** | **70-79** | **60-69** | **0-59** |
| **完成毕业要求1-4，拥有数据科学与工程算法知识与实践能力** | **优异（优秀完成毕业要求并主动扩展实践课程知识）** | **优秀（优秀完成毕业要求）** | **良好（较好完成毕业要求）** | **合格（能够完成毕业要求）** | **不合格（未能完成毕业要求）** |

表2：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **课程目标** | **评分标准** | | | | |
| **优（A）** | **良（B+）** | **中（B-）** | **及格（C）** | **不合格（F）** |
|  |  |  |  |  |  |

表3：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **课程目标** |  | | **评分标准** | |
| **优** | **合格** | | **不合格** |
|  |  |  | |  |