

期中总结

王延昊

Email: yhwang@dase.ecnu.edu.cn

2021.10.22

第一章 绪论

- 数据分析处理基本阶段（了解）
- 算法设计原则（了解）
- 算法评价指标
 - 效率指标（了解）
 - 分类问题精度指标（*掌握）
 - 回归问题精度指标（*掌握）
 - 排序问题精度指标（了解）

第一章 绪论

- 基本题型
 - 根据分类算法结果混淆矩阵，计算分类精度指标
 - 根据给定样本和回归函数，计算回归精度指标

第二章 抽样算法

- 抽样问题的基本概念（了解）
- 系统抽样（*掌握）
- 分层抽样
 - 等额分配 / 等比例分配（*掌握）
 - 奈曼分配 / 经济分配（了解）
- 水库抽样（*掌握）

第二章 抽样算法

- 基本题型
 - 根据给定抽样条件，计算被抽样样本
 - 根据给定抽样条件，设计抽样方案

第三章 尾概率不等式

- Markov 不等式 (*掌握)
- Chebyshev 不等式 (*掌握)
- Chernoff 不等式 (*掌握)
- Morris, Morris+ 和 Morris++ 算法 (*掌握)

第三章 尾概率不等式

- 基本题型
 - 根据题目给定条件，使用概率不等式计算概率上界
 - 理解概率不等式的证明，并将其运用到概率不等式的扩展形式
 - 理解 Morris, Morris+ 和 Morris++ 算法复杂度的证明，并将其运用到相关问题

第四章 哈希技术

- 哈希技术的基本概念 (*掌握)
- 布隆过滤器 (*掌握)
- 局部敏感哈希
 - 集合 Jaccard 相似度和距离 (*掌握)
 - 最小哈希 (*掌握)
 - 基于最小哈希的局部敏感哈希 (*掌握)

第四章 哈希技术

- 基本题型
 - 理解布隆过滤器的基本原理和误判率分析，对给定的布隆过滤器或扩展数据结构，分析其误判率
 - 理解 Jaccard 相似度和最小哈希的概念，对给定的集合和哈希函数数组，计算其 Jaccard 相似度和最小哈希签名

第五章 频繁项挖掘

- 数据流模型（了解）
- Misra Gries 算法（*掌握）
- Sketch 算法
 - Count Sketch 算法（*掌握）
 - Count-Min Sketch 算法（*掌握）

第五章 频繁项挖掘

- 基本题型
 - 理解 Misra Gries 算法流程，对给定的数据流，描述其执行过程并计算输出结果
 - 理解 Count Sketch 算法和 Count-Min Sketch 算法，对给定的数据流和哈希函数组，描述其执行过程并计算输出结果
 - 理解 Count Sketch 算法和 Count-Min Sketch 算法的复杂度和误差分析，将其运用到相关问题

第六章 EM算法

- 似然函数和最大似然估计 (*掌握)
- EM 算法及其应用 (了解)

第六章 EM算法

- 基本题型
 - 对给定的分布概率分布函数，计算其样本的似然函数和参数的最大似然估计

第七章 随机游走

- 马尔可夫链的概念和性质 (*掌握)
- 马尔可夫链平稳分布的存在性和唯一性条件 (*掌握)
- PageRank 算法 (*掌握)

第七章 随机游走

- 基本题型
 - 对给定的转移概率矩阵，判断其对应的马尔可夫链平稳分布是否存在，并求解其平稳分布
 - 对给定的转移概率图，判断其对应的马尔可夫链的可约性，计算其周期，并判断其是否存在唯一的平稳分布

期中考试

- 考试时间：2021年11月1日（星期一）上午 8:00 至 9:30
- 考试基本要求
 - 闭卷考试，不得携带课本，笔记和计算器，手机关机
 - 间隔就坐，任意两人至少间隔一个座位
 - 试题和答题纸都写上学号和名字，考试结束后统一交回
- 考试内容
 - 题目难度不大，以课后练习题题型为主（没有原题）
 - 计算量可能比较大，注意把握考试时间