

# 2024-2025 学年春季学期《数据科学与大数据技术的数学基础》

## 期末复习大纲

2025.6

注意：考题全是计算/简答/证明题，没有选择/填空题。建议参考复习大纲充分理解相关知识点。

### 第一讲：一致性哈希

关于是否存在机器占据区间比例过大或过小的证明过程

### 第二讲：布隆过滤器

当使用一个或多个哈希函数用于映射时，误报概率的推导及分析过程

### 第三讲：CM Sketch 方法

马尔可夫不等式的内容与使用。

### 第四讲：最小哈希

(1) 多个独立同分布随机变量的平均值的期望与方差的计算；

(2) 切比雪夫不等式的内容与使用。

### 第六讲：局部敏感哈希

(1) Jaccard 相似度的概念及与最小哈希之间的关系；

(2) 推导在局部敏感哈希方法下两个文件需要比对的概率。

### 第七讲、第八讲：主成分分析

(1) 特征值、特征向量的定义；

(2) 特征值分解的含义，小规模矩阵特征值分解的手动计算步骤；

(3) 正交矩阵的基本性质；

(4) 利用主成分分析对原始数据进行近似的近似误差推导过程；

(5) 利用幂迭代法求解各个主成分的思想及步骤。

### 第九讲、第十讲：奇异值分解

(1) 给定小规模矩阵，手动计算矩阵的完整奇异值分解；

(2) 根据矩阵的奇异值分解，解决低秩矩阵近似问题；

(3) 理解奇异值分解的定义与等价形式之间的关系。

### 第十二讲、第十三讲：压缩感知

(1) 理解用于信号恢复的  $L_1$  最小化问题；

(2) 理解用于信号恢复的  $L_2$  最小化问题；

(3) 利用正交匹配追踪算法求解  $L_0$  最小化问题的思想及步骤。