

大数据算法课程 (2021年春)

丁虎

huding@ustc.edu.cn

中国科学技术大学，计算机科学与技术学院

基本内容 (10-12周)

1. 常用概率不等式
2. 降维方法: PCA/近似PCA, JL-Transform
3. JL-Transform的应用: 矩阵乘法、压缩感知、SVD、线性回归
4. 最近点查询: kd-tree, Quad-tree, R-tree; LSH; product quantization
5. 机器学习基础: VC dimension/ PAC learning, AdaBoost, multiplicative weights update method
6. 聚类及近似算法: k-means, k-means++, hierarchical clustering, DBSCAN
7. 支持向量机: Gilbert algorithm, nu-SVM, soft margin SVM, SMO, Core-SVM, Kernel method
8. 核心集技术, 次线性算法, 流算法
9. 分布式计算, 通信复杂度

考核标准

- 出勤5%
- 作业三次（期中前2次，期中后1次） $5\%*3$
- 期中+期末考试 $40\%*2$

助教

- 秦睿哲 red46@mail.ustc.edu.cn
- 陈潭 ct1997@mail.ustc.edu.cn
- 黄若民 hrm@mail.ustc.edu.cn

参考教材

- **Mathematic foundation for data analysis, Jeff Phillips**
<http://www.cs.utah.edu/~jeffp/M4D/M4D.html>
- **Foundations of Data Science, Avrim Blum, John Hopcroft, and Ravindran Kannan** <https://www.cs.cornell.edu/jeh/book.pdf>