

“大数据分析中的算法”教学大纲

◆ 基本目的:

大数据给数据分析和处理带来了前所未有的机遇和挑战。本课程介绍大数据分析中一些算法：数据的稀疏和低秩表达，稀疏和低秩矩阵优化，社交网络计算中的图与网络流问题，机器学习和数据挖掘的最优化算法，随机优化算法，并行计算等等

Big data has brought unprecedented opportunities and challenges to data analysis and processing. This course introduces some basic concepts of algorithms for big data analysis: sparse and low-rank data representation, sparse and low-rank matrix optimization, graph and network flow problems in social network computation, optimization algorithms for machine learning and data mining, stochastic optimization algorithms, parallel computing, etc.

◆ 课程对象:

高年级本科生和研究生。

◆ 参考书:

- (1) Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, Introduction to Algorithms, The MIT Press, <http://mitpress.mit.edu/books/introduction-algorithms>
- (2) Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, Cambridge University Press, <http://www.mmids.org/>
- (3) Stephen Boyd and Lieven Vandenberghe, Convex optimization, Cambridge University Press, 2004, <http://stanford.edu/~boyd/cvxbook/>
- (4) Jorge Nocedal and Stephen Wright, Numerical Optimization, Springer, 2006, <http://www.ece.northwestern.edu/~nocedal/book/>
- (5) 袁亚湘, 孙文瑜, 最优化理论与方法, 科学出版社, 2003

◆ 内容摘要和学时分配:

1. 课程简介, 3学时
课程简介, 大数据分析中的最优化理论与算法介绍
2. 线性规划, 半定规划, 6学时
线性规划, 单纯形方法, 半定规划, 对偶理论
4. 稀疏优化与低秩矩阵恢复, 9学时
压缩感知和稀疏优化基本理论和算法
低秩矩阵恢复的基本理论和算法
PCA, robust PCA (matrix separation), sparse PCA
5. 社交网络计算中的图和网络流问题, 9学时

the network simplex problem
the shortest path problem
the maximum flow problem
the minimum spanning tree problem

6. 机器学习和数据挖掘, 9学时

聚类分析: clustering

高维数据降维: eigenvalue, SVD

链接分析: page rank

推荐系统: matrix completion,

大规模机器学习: support vector machine

7. 现代医学成像与高维图像分析, 3学时

相位恢复以及低温电子显微镜和三维重构中的若干反问题

8. 大数据分析的随机优化算法, 3学时

9. 大数据分析的并行计算、分布式计算、分散式计算, 6学时

OpenMP, MPI, 稀疏优化的并行计算, 分散式计算

10. 课程项目报告, 6学时

学生分组做小课题报告

◆ **教学方式:**

课堂讲授: 80%

讨论: 20%

◆ **成绩评定办法:**

成绩评定:

(1) 4-5次大作业, 包括习题和程序: 50%

(2) 课程项目: 50%

要求: 作业和课程项目必须按时提交, 迟交不算成绩, 抄袭不算成绩