

Sketching and Sampling for Big Data Analysis

Overview

This is a course aimed at introducing various dimensionality reduction techniques, such as sampling and sketching, to speed up commonly occurring optimization problems, in particular those occurring in numerical linear algebra.

Today, there is a tremendous interest in processing “Big Data” due to the ubiquity of large data sets being generated in production systems. This has led to an interest in designing algorithms and systems for problems with very large inputs. The problems range from matrix problems and numerical linear algebra, optimization problems and graph problems.

This course will highlight the recent advances in algorithms for numerical linear algebra and optimization that have come from the technique of linear sketching, whereby given a matrix, one first compresses it to a much smaller matrix by pre-multiplying it by a (usually) random matrix with certain properties. Much of the expensive computation can then be performed on the smaller matrix, thereby accelerating the solution for the original problem. This technique has led to the fastest known algorithms for fundamental problems in this area. We will consider least squares as well as ℓ_1 -regression problems, low rank approximation, and many variants of these problems, such as those in distributed environments. We will also discuss connections of these methods with and using graph sparsifiers. Some of this work is partially covered in the speaker's monograph: “Sketching as a Tool for Numerical Linear Algebra”, Foundations and Trends in Theoretical Computer Science, vol 10, issue 1-2, pp. 1-157, 2014, publisher: NOW Publishing.

This is a theoretical course (with lectures), and it is assumed that participants will have background in Linear Algebra and Probability and are familiar with mathematical proofs.

Course participants will learn these topics through lectures and hands-on experiments. Also case studies and assignments will be shared to stimulate research motivation of participants.

| | |
|--------------------------------|---|
| Modules | Sketching and Sampling for Big Data Analysis : Jan 2- Jan 6, 2017 Number of participants for the course will be limited to twenty. |
| You Should Attend If... | <ul style="list-style-type: none">▪ you are a computer scientist, or work with Big data and are interested in designing and learning about the state of the art in dimensionality reduction and efficient computation of numerical linear algebra.▪ you work in Big data and use regression and other standard tools for data analysis.▪ you are a student or faculty from academic institution interested in learning about state of the art research in the area of streaming, sketching. |
| Fees | The participation fees for taking the course is as follows: Participants from abroad : NIL Industry/ Research Organizations: NIL Academic Institutions: NIL A copy of the instructional material will be provided. The participants will be provided with accommodation on payment basis. |

The Faculty



Prof. David Woodruff is a faculty member at Carnegie Mellon University. His research interests are in sketching, data streaming algorithms and lower bounds, communication complexity, numerical linear algebra, machine learning and sparse recovery.

Course Co-ordinator

Prof. Sumit Ganguly
Phone: 0512-2597597
E-mail: sganguly@cse.iitk.ac.in

<http://www.gian.iitkgp.ac.in/>