Spring:

Mathematical Foundations for Data Science

Instructor: Thomas Strohmer
Where: Mathematical Sciences Building 2112
When: Tuesday and Thursdays at 2:10-3:30PM

Course Objective:
The rapidly increasing availability of data and tools for its analysis have led to an explosion of new insights that are transforming our understanding of everything from human behavior to the structure of the universe. Experiments, observations, and numerical simulations in many areas of science nowadays generate massive amounts of data. This rapid growth heralds an era of "data-centric science," which requires new paradigms addressing how data are acquired, processed, distributed, and analyzed. This course will cover mathematical models and concepts for developing algorithms that can deal with some of the challenges posed by Big Data and Artificial Intelligence.

Prerequisite:
Linear algebra and a basic background in probability as well as basic experience in programming (preferably Matlab) will be required. Some basic knowledge in optimization is recommended.

List of topics: (subject to minor changes)

- Principal Component Analysis, Singular Value Decomposition.
- Curses and blessings of dimensionality.
- Probability in high dimensions. Concentration of measure, matrix concentration inequalities.
- Data clustering, community detection.
- Dimension reduction. Johnson-Lindenstrauss, sketching, random projections.
- Compressive sensing. Efficient acquisition of data, sparsity, low-rank matrix recovery.
- Randomized numerical linear algebra.
- Diffusion maps, manifold learning, intrinsic geometry of massive data sets.
- Some basics on Deep Learning