

Introduction to Statistical Learning, Data Analysis and Signal Processing

Spring 2017

Mauro Maggioni, www.math.jhu.edu/~mauro



Synopsis of course content

Introduction to high dimensional data sets: key problems in statistical and machine learning. **Geometric aspects.** Principal component analysis, linear dimension reduction, random projections. Concentration phenomena: examples and basic inequalities. Metric spaces and embeddings thereof. Kernel methods. Nonlinear dimension reduction, manifold models.

Regression. Vector spaces of functions, linear operators, projections. Orthonormal bases; Fourier and wavelet bases, and their use in signal processing and time series analysis. Basic approximation theory. Linear models, least squares. Bias and variance tradeoffs, regularization. Sparsity and compressed sensing. Multiscale methods.

Graphs and networks. Random walks on graphs, diffusions, page rank. Block models. Spectral clustering, classification, semi-supervised learning.

Algorithmic and computational aspects of the above will be consistently in focus, as will be computational experiments on synthetic and real data.

Prerequisites

Linear algebra will be used throughout the course, as will multivariable calculus and basic probability (discrete random variables). Basic experience in programming in C or MATLAB or R or Octave. Recommended: More than basic programming experience in Matlab or R; some more advanced probability (e.g. continuous random variables), some signal processing (e.g. Fourier transform, discrete and continuous).

Grading

Grade to be based on weekly assignments, midterm and final exams, and a *final team project*. The final team project includes a report and the presentation of a poster on the topic of the project (typically involving studying a small number of papers and summarizing them and/or working on a data set). Weekly problem sets will focus on computational projects and theory.

Students from all areas of science, mathematics and applied mathematics, engineering, computer science, statistics, quantitative biology, economics that need advanced level skills in solving problems related to the analysis of data, signal processing, or statistical modeling are encouraged to enroll.