High-Dimensional Approximation, Probability, and Statistical Learning

Instructor: Mauro Maggioni

Web page: https://mauromaggioni.duckdns.org/teaching/

Synopsis

The course covers fundamental mathematical ideas for approximation and statistical learning problems in high dimensions. We start with studying high-dimensional phenomena, through both the lenses of probability (concentration inequalities) and geometry (concentration of measure phenomena). We then consider a variety of techniques and problems in high-dimensional statistics and machine learning, ranging from dimension reduction (random projections, embeddings of metric spaces, manifold learning) to classification and regression (with connections to approximation theory, Fourier analysis and wavelets, Reproducing Kernel Hilbert Spaces, tree-based methods, multiscale methods), estimation of probability measures in high dimensions (density estimation, but also estimation of singular measures, and connections with optimal transport). We then consider graphs and networks, Markov chains, random walks, spectral graph theory, models of random graphs, and applications to clustering. Finally, we discuss problems at the intersection of statistical estimation, machine learning, and dynamical/physical systems, in particular Markov state space models, Hidden Markov Models, and interacting particle/agent systems, as well as model reduction for stochastic dynamical systems. Computational aspects will be discussed in all topics above (including, e.g., randomized linear algebra, fast nearest neighbors methods, basic optimization).

Detailed Topics

- High-dimensional phenomena, probability, and concentration of measure phenomena. Limit laws (large numbers, central limit theorem); concentration inequalities and geometric concentration; random matrices and their spectra, relationships with their continuous limits. Applications of random matrices to covariance matrix estimation, Singular Value Decomposition (SVD), linear dimension reduction, random projections, Johnson-Lindenstrauss Lemma, data and matrix compression, randomized linear algebra.

- Basic problems in statistical learning: regression, density estimation. The interplay between randomness and approximation theory. Curse of dimensionality: geometric aspects, and manifestations in approximation theory. Uniform central limit theorems. Regression: problem statement, examples. Curse of dimensionality in regression: lower bounds. Linear regression: least squares. Regularization. Reproducing Kernel Hilbert Spaces. Nonparametric and multiscale methods. Regression of Lipschitz and Hölder functions. Fourier and multiscale methods. Tree-based methods. Regression on manifolds. Adaptivity. Estimating probability measures and densities. Spaces of probability measures. Optimal transportation distances. Estimation of singular measures. Multiscale methods.

- Basic approximation theory in low-dimensions: linear and nonlinear approximation by Fourier and wavelets in classical smoothness spaces, Sobolev spaces, and Besov spaces. Notions of

complexity of function spaces.
Applications in image processing, inverse problems and PDEs.

- Introduction to metric spaces. Mappings between metric spaces, Lipschitz maps, distortion; embedding metric spaces into Euclidean space (positive and negative results), approximation of metric spaces by trees. Applications to algorithms and statistics. Manifold learning.

- Clustering problems: K-means, spectral clusterings, connections with graph problems, random matrix problems, optimal transport.

- Graphs and networks: random walks, diffusion. Basic spectral graph theory. Pagerank. Random graphs, Erdös-Rényi graphs, stochastic block models. Clustering: K-means, K-flats, vector quantization, spectral clustering. Semi-supervised learning.
  Diffusion processes on graphs and their applications; diffusion geometry.
  Signal processing on graphs; wavelets on graphs.

- Machine Learning and Physics/dynamical systems: Markov models, Markov state models, hidden Markov models; model reduction for dynamical systems.
  Interacting particle systems and learning of interaction kernels.
  Homogenization, averaging and model reduction fo stochastic dynamics systems.

References

Foundations of Data Science, D. Blum, J. Hopcroft & R. Kannan.
High Dimensional Probability, An Introduction with Applications in Data Science, R. Vershynin.
High Dimensional Statistics lecture notes, P. Rigolet.
Universal Algorithms for Learning Theory Part I: Piecewise Constant Functions, Binev et al.
Estimation in high dimensions: a geometric perspective, R. Vershynin.
Measure Concentration lecture notes, A. Barvinok.
Nonlinear Approximation, R. A. DeVore
Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science, A. Bandeira
Introduction to nonparametric estimation, A. Tsybakov
A distribution-free theory of nonparametric regression, L. Gyorfi, M. Kohler, A. Krzyzak, H Walk
Lectures on Spectral Graph Theory, F.R.K. Chung.

Grading
Grade to be based on assignments (30%), one midterm (30%) and a final project (40%).
Weekly problem sets will include theory, analysis and computational projects.

Prerequisites
Analysis I & II; basic probability.

Recommended but not required: probability beyond the basics, or introduction to statistics; functional analysis. Programming skills in Matlab or C or R.