High-Dimensional Approximation, Probability, and Statistical Learning

Instructor: Mauro Maggioni — Office: Whitehead 302D

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 phenomena). We then consider a variety of techniques and problems in high-dimensional statistics and machine learning, ranging from dimension reduction (random projects, 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. 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: limit laws (large numbers, central limit theorem); concentration inequalities and geometric concentration; random projections, Johnson-Lindenstrauss Lemma.

- Singular value decomposition (SVD), low-rank matrices. First applications of SVD: dimension reduction, variances and covariances. Data and matrix compression; applications to computation.

- Concentration of measure phenomena. Basic concentration inequalities. Random matrices and their spectra, relationships with their continuous limits. Applications of random matrices to covariance matrix estimation, signal processing, compressed sensing.

- 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. Introduction to applications in imaging, inverse problems and PDE's.

- 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. Tree-based methods. Regression of Lipschitz and Hölder functions. Fourier and multiscale methods. Regression on manifolds. Adaptivity. Estimating probability measures and densities. Density Estimation. Spaces of probability measures. Optimal transportation distances. Estimation of singular measures. Multiscale methods.

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

- Machine Learning and Dynamics: Markov models, Markov state models, hidden Markov models; model reduction for dynamical systems. Interacting particle systems and learning of interaction kernels.

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.

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.