Syllabus

Instructor: Xiuyuan Cheng Email: xiuyuan.cheng@duke.edu
Time: TuTh, 11.45am-1.00pm
Classroom: Gross Hall 318

Synopsis:

This course gives an introduction to certain theoretical and computational tools in data analysis and machine learning from a mathematical point of view. It is designed for students who are interested in research in the field. Topics involve, but are not limited to, spectral methods, manifold learning, neural networks as well as general issues with high dimensional data. The course will be accessible to first-year graduate students as well as senior undergraduate students with backgrounds in applied mathematics, electrical engineering, statistics and related fields of study.

Below are a list of tentative topics to be covered:

- Principal component analysis (PCA) in high dimension
  - Marcenko-Pastur law, random matrix theory and universality
  - Spiking model in covariance estimation
  - Eigen-shrinkage and matrix de-noising
- Clustering on graphs
  - Review of spectral clustering and limitations
  - SDP relaxation: theory and computational issues
- Graph-based dimension reduction
  - Stochastic neighbor embedding (SNE), comparison with manifold learners (Isomap, LLE, diffusion map, etc) and multidimensional scaling (MDS)
  - Heat kernel on manifold, convergence of graph laplacian
- Estimation on graphs
  - De-noising functions by non-local means
  - Synchronization of group elements on graphs
  - Application: rotation registration in Cryo-electron microscopy (CryoEM)
- Concentration of measure
  - Review of concentration inequalities
  - Application: the spectra of random graphs
  - Application: randomized fast low-rank matrix approximation
- Representation learning by neural networks
  - Classifiers and “auto-encoders”
  - Representation by intermediate layers in a neural network
  - Convolutional neural networks and the instability in input
- High dimensional two samples
  - Distance between two samples in one dimension
  - Kernel density estimation and limitations in high dimension
  - Application: evaluating deep generative neural networks
Course Material:
   Lecture notes will be distributed as the class goes. List of bibliography and suggested readings will be provided.

Prerequisites:
   Analysis (multivariate calculus), Linear algebra, Probability (introduction-level, not measure theory). Experience with coding (e.g. Matlab, python etc.) is helpful, but not required.

Assignment and Grading:
   Homework will be assigned without mandatory hand-in. Evaluation will be based on (i) class participation (40%) and (ii) final written report: one lecture scribing, or review of one topic, or research paper (60%). Audition welcome.