Stat 991: Multivariate Analysis, Dimensionality Reduction, and Spectral Methods
Syllabus
Instructor: Sham Kakade

Topics

Modern statistical approaches on large datasets must directly analyze and manipulate data in either matrix or vector formats. This course will focus on the statistical theory and practice of manipulating such data. The topics covered will be: multivariate analysis, dimensionality reduction, convexity issues of working with matrices, and spectral methods.

With regards to dimensionality reduction, we will cover PCA, CCA, and random projections (e.g. Johnson-Lindenstrauss) and examine potential applications. With regards to convexity issues, the course will examine the rudimentary question of how accurate is an SVD of a random matrix (we will examine a generalization of the Chernoff method to matrices). Other potential topics may include matrix completion (filling in the entries of a matrix with missing entries), subspace identification (e.g. learning time series models like Kalman filters based on a multivariate, covariance analysis), locality sensitive hashing (randomly projecting data for efficient storage and recall), matrix based regularization methods (and related convexity issues), and kernel methods/Gaussian process regression.

Some of the major topics discussed in the course will include the following:

- Dimensionality reduction, including SVD and random projections.

- Accuracy of SVD: How accurate is the subspace found? We will cover a recent concentration result for random matrices.

- Implications for learning: We will examine projecting the data to low dimensional spaces and then learning on these lower dim spaces. We will cover both regression and clustering here.

- CCA, Subspace ID and Time Series: how to (probably learn) state space models, including Kalman filters and HMMs.

- Matrix Completion: how to fill in missing entries of a matrix?

Prerequisites

The course is appropriate for a graute student with some background in statistics and machine learning. The course will assume a basic level of mathematical maturity, so please contact the instructor if you have concerns.

Requirements

The course will consist of homework and possibly a project. TBD.