Workshop on Geometry, Random Matrices, and Statistical Inference

January 16 – 19, 2007

SPEAKER ABSTRACTS

Sanjoy Dasgupta
University of California-San Diego
Department of Computer Science
dasgupta@cs.ucsd.edu

"Projection Pursuit, Gaussian Scale Mixtures, and the EM Algorithm"

The EM algorithm for fitting Gaussian mixture models is one of the most widely-used clustering methods. Yet, surprisingly little is known about its behavior. For instance, there are many different ways to initialize EM, and to merge/remove intermediate clusters, and the effects of these different strategies are not understood in a principled way.

I'll describe a new probabilistic analysis of EM. First of all, it will emerge that many common methods of initializing and running EM produce hopelessly suboptimal results even in ideal clustering scenarios. On the other hand, a particular variant of EM will provably recover a near-optimal clustering, provided that the clusters are adequately separated and that their distributions are of a certain fairly general form.

The type of cluster distributions allowed is motivated by a new result in projection pursuit, along the lines of the folklore that "all projections of high-dimensional data look Gaussian". Specifically, I'll show that for any D-dimensional distribution with finite second moments, there is a precise sense in which almost all of its linear projections into $d < D$ dimensions look like a scale-mixture of spherical Gaussians (concentric spherical Gaussians with the same center). The extent of this effect depends upon the ratio of d to D, and upon the "eccentricity" of the high-dimensional distribution.

Steven Damelin
Georgia Southern University
Department of Mathematical Sciences
damelin@georgiasouthern.edu

"On Kernels, Energy and Metrics"

In this talk, we link together important ideas of discrepancy, energy and metrics for large classes of symmetric, positive definite kernels defined on compact subsets of Euclidean space.

The kernels in question, for example logarithmic repulsive kernels, arise for example in computer vision, statistical inference and in the study of fluctuations of eigenvalues of random matrices.

This is based, partly on joint work of the author with Grabner (Graz), Levesley (Leicester), and Hickernell (IIT).

Vladimir Koltchinskii
Georgia Institute of Technology
School of Mathematics
vlad@math.gatech.edu

"Sparsity in High Dimensional Learning Problems"

Learning problems (such as regression or classification) can be often formulated as penalized empirical risk minimization over a linear span of a very large dictionary of functions with a properly chosen complexity penalty. One of the choices of penalty that has been very popular in the recent years is ℓ_1-penalty. This penalty function is convex and, if the loss function is also convex (which is the case in L_2- or in L_1-regression as well as in large margin classification methods such as boosting or SVM), the penalized empirical risk minimization becomes a convex optimization problem. We will consider ℓ_1-penalization as well as more general ℓ_p-penalization with $p>1,$ but close enough to $1.$ A number of inequalities will be considered that relate the degree of "sparsity" of the empirical solution to the degree of "sparsity" of the true solution. Oracle inequalities showing the impact of "sparsity" on the excess risk of the empirical solution will be also discussed.

Guy Lebanon
Purdue University
Department of Statistics
lebanon@purdue.edu

"Random Geometry and Statistical Translation in Text Analysis"

High dimensional structured data such as text and images is often poorly understood and misrepresented in statistical modeling. The standard histogram representation suffers from high variance and performs poorly in general. We explore novel connections between statistical translation, heat kernels on manifolds and graphs, and random geometries. These connections show a surprising link between the statistical ideas of regularization and biased estimation and common practices in text analysis.

Liza Levina
University of Michigan
Department of Statistics
elevina@umich.edu

“Regularized Estimation of Large Covariance Matrices”

Estimation of covariance matrices has a number of important applications, which include principal component analysis, classification by linear or quadratic discriminant analysis, and inferring independence and conditional independence between variables. This talk will summarize recent results for regularizing covariance matrices when there is a natural ordering of the variables, and discuss extensions to cases when no such ordering is available. I will also discuss some questions (no answers!) on connections between covariance estimation and nonlinear dimension reduction.

Yuriy Mileyko
Duke University
Department of Computer Science
yury@cs.duke.edu

“Computing Homology of High-dimensional Point Clouds”

In this talk I will consider the problem of recovering homology groups of a subspace of a high-dimensional Euclidean space given only a finite number of points lying on or near the subspace. We approach this problem by constructing a specific family of nested simplicial complexes, witness complexes, and computing their persistent homology groups. I shall give an overview of witness complexes and persistent homology, and present several results and open questions.

Partha Niyogi
University of Chicago
Department of Computer Science
niyogi@cs.uchicago.edu

“A Geometric Perspective on Learning”

Increasingly, we face machine learning problems in very high dimensional spaces. We proceed with the intuition that although natural data lives in very high dimensions, they have relatively few degrees of freedom. One way to formalize this intuition is to model the data as lying on or near a low dimensional manifold embedded in the high dimensional space. This point of view leads to a new class of algorithms that are “manifold motivated” and a new set of theoretical questions that surround their analysis. A central construction in these algorithms is a graph or simplicial complex that is data-derived and we will relate the geometry of these to the geometry of the underlying manifold. In particular, we will see the role of the Laplace Beltrami operator in many of these developments. Applications to embedding, clustering, classification, and semi-supervised learning will be considered.

Guilherme Rocha
University of California-Berkeley
Department of Statistics
gvrocha@stat.berkeley.edu

"Grouped and Hierarchical Selection through Composite Absolute Penalties (CAPs)"

For datasets with many predictors and few samples, side information often must be added to fitting. We introduce Composite Absolute Penalties (CAP) to blend predefined grouping and hierarchical information among the predictors into regression and classification. Special cases include Zou & Hastie(2005)'s elastic net, Kim et. al(2005)'s Blockwise Sparse Regression and Yuan & Lin(2006)'s GLASSO. CAPs are built by combining norm penalties at the across and within group levels. For disjoint groups, a Bayesian interpretation lays bare the role of the norms used to construct CAP. Hierarchical selection is reached by defining nested groups. For general CAPs, we use the BLASSO and cross-validation to compute CAP estimates. For CAPs built from L_1 and L_\infty norms, we give efficient algorithms and regularization selection criteria. The enhnaced prediction performance of CAP estimates is shown through simulated experiments.

Armin Schwartzman
Harvard School of Public Health
Department of Biostatistics
armins@hsph.harvard.edu

"Distributions for Random Positive Definite Matrices"

In classical multivariate statistics, the most common probability model for positive definite (PD) matrices is the Wishart distribution, which arises as the distribution of the sample covariance matrix of a multivariate normal sample. Randomness in general PD matrix data, however, can be of a very different nature. In Diffusion Tensor Imaging, for instance, PD matrix data is obtained directly from measurements of brain anatomy. In this talk I explore alternative ways to model random PD matrices. This is done keeping in mind three objectives: 1) Arbitrary covariance between matrix elements; 2) Data-analysis friendliness; 3) Match between the MLE for the mean parameter and the already known means for positive definite matrices, i.e. arithmetic, geometric and intrinsic. The new probability distributions obtained are the log normal, the Riemannian log normal and the geodesic normal, all based on the normal distribution for symmetric matrices followed by an appropriate matrix log transformation. The construction gives insights into the geometry of the PD manifold and ways to do statistics on more general manifolds.

Tao Shi
Ohio State University
Department of Statistics
taoshi@stat.ohio-state.edu

"Multi-resolution Covariance Modelling in Spatial Random Effect Model"

In spatial statistics, estimating and inverting the spatial covariance matrix is difficult when sample size n is large. The computation complexity of kriging is n^3. In this paper, we propose a Spatial Mixed Effects (SME) statistical model to predict the missing values, denoise the observed values, and quantify the spatial-prediction uncertainties.
The computations associated with the SME model are linear scalable to the number of data points, which makes it feasible to process massive global satellite data. We apply our proposed methodology, which we call Fixed Rank Kriging (FRK), to the level-3 Aerosol Optical Depth dataset collected by NASA's Multi-angle Imaging SpectroRadiometor (MISR) instrument flying on the Terra satellite. Overall, our results were superior to those from nonstatistical methods and, importantly, FRK has an uncertainty measure associated with it.

This is a joint work with Prof. Noel Cressie.

Amit Singer
Yale University
Department of Mathematics
amit.singer@yale.edu

"The 3D Random Computed Tomography Structuring of Proteins and the Specral Non-Linear ICA Algorithm"

In this talk we present a reconstruction algorithm for the 3D atomic structure of randomly oriented proteins using modified graph laplacians. The 2003 Nobel Prize in Chemistry was co-awarded to R. MacKinnon who was the first to structure a protein channel (the potassium channel) in 1998 by crystallizing the protein and then using the classical x-ray computed tomography (CT). However, most membrane proteins cannot be crystallized, and the classical CT cannot be used. In our experimental setup, we are given real noisy 2D projections (electron microscope images) in random directions, because the proteins are randomly oriented rather than being aligned as in MacKinnon's setup.

Still, we show that the 3D reconstruction is made possible by a certain modification of the images' graph laplacian combined with the 3D Fourier slice theorem. The reconstruction is a particular case of a more general spectral non-linear independent component analysis (ICA) algorithm that combines local PCA with the graph laplacian.

This is a joint work with Ronald Coifman, Yoel Shkolnisky and Fred Sigworth (Yale University).