BIOSTATISTICS SEMINAR

David B. Dunson, PhD
Arts and Sciences Distinguished Professor of Statistical Science,
Duke University

Learning & exploiting low-dimensional structure in high-dimensional data

This talk will focus on the problem of learning low-dimensional geometric structure in high-dimensional data. We allow the lower-dimensional subspace to be non-linear. There are a variety of algorithms available for "manifold learning" and non-linear dimensionality reduction, mostly relying on locally linear approximations and not providing a likelihood-based approach for inferences. We propose a new class of simple geometric dictionaries for characterizing the subspace, along with a simple optimization algorithm and a model-based approach to inference. We provide strong theory support, in term of tight bounds on covering numbers, showing advantages of our approach relative to local linear dictionaries. These advantages are shown to carry over to practical performance in a variety of settings including manifold learning, manifold de-noising, data visualization (providing a competitor to the popular tSNE), classification (providing a competitor to deep neural networks that requires fewer training examples), and geodesic distance estimation. We additionally provide a Bayesian nonparametric methodology for inference, using a new class of kernels, which is shown to outperform current methods, such as mixtures of multivariate Gaussians.

Thursday January 30, 2020 **3:30 pm - 4:30 pm**
Blue Cross and Blue Shield of North Carolina Foundation Auditorium

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH