Speaker: Tyrus Berry

Title: The crucial role of statistics in manifold learning

Abstract: A powerful assumption when analyzing high-dimensional data sets is that the data is restricted to lie on or near a low-dimensional manifold. Recently, kernel based methods have been developed which give consistent estimators of important geometric operators, such as the Laplace-Beltrami operator. In this talk we first overview the goals and methods of manifold learning before turning to recent results and their implications for the field.

The theory of kernel density estimation (KDE) has been extended to kernel operator estimation (KOE). We introduce notions of pointwise and spectral consistency for operator estimators and explore the bias-variance tradeoff. The variance of these estimators reveals that fixed-bandwidth kernel methods do not converge for non-compact manifolds (unlike KDE). Using variable-bandwidth kernels we obtain convergence on a large class of non-compact manifolds. These statistical results can be illustrated with simple graph constructions on small data sets, but have far-reaching consequences. Moreover, the bias-variance tradeoff reveals the source of the curse-of-dimensionality and indicates a possible direction towards analyzing high-dimensional data. Finally, for manifolds with boundary we introduce a boundary distance estimator which enables consistent density estimation.