Statistics Seminar

Department of Mathematical Sciences

DATE:	Friday, April 27, 2015
TIME:	1:15pm to 2:15pm
LOCATION:	WH 100E
SPEAKER:	Yifan Xu (Binghamton University)
TITLE:	Fast clustering using adaptive density peak detection

Abstract

Common limitations of clustering methods in bioinformatics include the slow computation and algorithm convergence, the prespecification of a number of different intrinsic parameters, and the lack of robustness. The recent clustering approach by Rodriguez and Laio proposed a fast search algorithm of cluster centers based on their local densities. However, the selection of the key intrinsic parameters in the algorithm was not systematically investigated. It is relatively difficult to estimate the “optimal” parameters since the original definition of the local density in the algorithm is based on a truncated counting measure.

We propose a clustering procedure with adaptive density peak detection, where the local density is estimated through the nonparametric multivariate kernel estimation. We also develop an automatic cluster centroid selection method through maximizing an average silhouette index. The advantage and flexibility of the proposed method are demonstrated through simulation studies and the analysis of a few benchmark gene expression data sets. The method only needs to perform in one single step without any iteration, and can be parallelized, thus is fast and has a great potential to apply on big data analysis.We also made an R-package “ADPclust” that is available on CRAN soon.

This is a joint work with Xiao-Feng Wang from Department of Quantitative Health Sciences at Cleveland Clinic Foundation.