

Statistics Seminar
Department of Mathematical Sciences

DATE:	Thursday, October 27, 2016
TIME:	1:15p-2:40p
LOCATION:	WH 100E
SPEAKER:	Haomiao Meng, Binghamton University
TITLE:	Regression in Heterogeneous Problems

Abstract

We develop a new framework for modeling the impact of sub-cluster structure of data on regression. The proposed framework is specifically designed for handling situations where the sample is not homogeneous in the sense that the response variables in different regions of covariate space are generated through different mechanisms. In such situation, the sample can be viewed as a composition of multiple data sets each of which is homogeneous. The traditional linear and general nonlinear methods may not work very well because it is hard to find one single unifying model to fit multiple data sets simultaneously. The proposed method is flexible to ensure that the data generated from different regions can be modeled using different functions. The key step of our method is to incorporate the k-means clustering idea into the traditional regression framework such that the regression and clustering tasks can be performed simultaneously. The k-means clustering algorithm is extended to solve the optimization problem in our model which groups the samples with similar response-covariate relationship together. General conditions under which the estimation of the model parameters is consistent are investigated. By adding appropriate penalty terms, the proposed model can conduct variable selection to eliminate the uninformative variables. The conditions under which the proposed model can achieve asymptotic selection consistency are also studied. The effectiveness of proposed method is demonstrated through applications to simulated and real data.

From:

<https://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:

<https://www2.math.binghamton.edu/p/seminars/stat/161027>

Last update: **2016/10/26 20:04**

